



---

## Hierarchic Modeling of Salmon Harvest and Migration

Author(s): Ken B. Newman

Source: *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 5, No. 4 (Dec., 2000), pp. 430-455

Published by: International Biometric Society

Stable URL: <https://www.jstor.org/stable/1400659>

Accessed: 20-04-2019 14:00 UTC

## REFERENCES

Linked references are available on JSTOR for this article:

[https://www.jstor.org/stable/1400659?seq=1&cid=pdf-reference#references\\_tab\\_contents](https://www.jstor.org/stable/1400659?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

*International Biometric Society* is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Agricultural, Biological, and Environmental Statistics*

# Hierarchical Modeling of Salmon Harvest and Migration

KEN B. NEWMAN

Using fishery recoveries from a tagged cohort of coho salmon, the ocean spatial-temporal abundance of the cohort is predicted using a state-space model. The model parameters, which reflect spatial distribution, mortality, and movement, vary considerably between different cohorts. To evaluate the effect of proposed management plans on a future cohort, uncertainty in the cohort-specific parameters is accounted for by a hierarchic model. As an application, release–recovery and fishing effort data from several cohorts of a hatchery-reared coho salmon stock originating from Washington state are used to calculate maximum likelihood estimates of the hyperparameters. Markov chain Monte Carlo is used to approximate the likelihood for the hyperparameters. The Markov chain simulates the sampling distribution of the state-space model parameters conditional on the data and the estimated hyperparameters and provides empirical Bayes estimates as a by-product. Given the estimated hyperparameters and the hierarchic model, fishery managers can simulate the variation in cohort-specific parameters and variation in the migration and harvest processes to more realistically describe uncertainty in the results of any proposed management plan.

**Key Words:** Empirical Bayes; Metropolis–Hastings algorithm; Spatial-temporal.

## 1. INTRODUCTION

Biologists from fishery management agencies on the west coast of North America annually develop and evaluate preseason management plans for the harvest of stocks of Pacific Ocean salmon. The objectives of these management plans include protecting stock health by ensuring that specified numbers of fish per stock return to spawn (the escape-ment), providing fish for commercial fisheries, enabling recreational fisheries to operate for extended periods of time, and achieving some level of politically defined parity in catch allocation between competing fisheries. Actions specified by management plans to achieve these objectives include specifying the opening and closing of fisheries on a time and area basis, setting catch quotas or catch ceilings, and enacting gear restrictions and size limits.

To assist in the evaluation of plans, mathematical models of varying complexity and realism have been developed. A serious limitation of the currently used models is the lack of

---

Ken B. Newman is Associate Professor, Division of Statistics, University of Idaho, Moscow, ID 83844-1104 (E-mail: [newman@uidaho.edu](mailto:newman@uidaho.edu)).

©2000 American Statistical Association and the International Biometric Society  
*Journal of Agricultural, Biological, and Environmental Statistics*, Volume 5, Number 4, Pages 430–455

an explicit migration mechanism. Some models implicitly assume that fish not harvested in one time and area are equally available to fisheries in any area in the subsequent time period. Other models constrain the availability of fish previously in one area to a few nearby areas in the current time period, but the constraints are formulated in a very subjective manner, making quantitative assessment of their reasonableness difficult. This latter issue is an example of a much more pervasive problem. Most of these models have been developed in an ad hoc manner with historical data being utilized in nonsystematic ways to arrive at parameter estimates with unknown precision and accuracy. Assumptions behind the estimation procedures and model processes are difficult, at best, to elucidate.

As an alternative to existing models, Newman (1998) formulated a state-space model (SSM) for the abundance of coho salmon (*Oncorhynchus kisutch*) stocks on a time and area basis. State-space models are gaining increased recognition and usage as an appropriate tool for modeling sequential fisheries data (Mendelsohn 1988; Sullivan 1992; Speed 1993; Schnute 1994; Reed and Simons 1996). SSMs provide a formal structure for linking observable time series, such as catch, to unobservable time series, such as salmon abundance. Algorithms such as the Kalman filter (Kalman 1960) provide a method for calculating likelihood functions and subsequently maximum likelihood estimates of unknown parameters and associated standard errors. The Kalman algorithms also provide estimates of the unobservable process, conditioning on the observable time series and the parameter estimates. The SSM formulated for coho salmon explicitly includes a migration mechanism and systematically utilizes information provided by fishery recoveries of tagged fish and by measures of fishing effort to calculate maximum likelihood estimates of unknown parameters. Conditional on the parameter estimates, the unobservable abundance of the salmon stock by time and area can be estimated.

The parameters of the salmon SSM reflect survival between time of ocean entry and time of harvest, the spatial distribution of the fish at the time harvest begins, the relationship between fishing effort and harvest, and migratory behavior. All these factors are functions of freshwater environment, ocean environment, and the nature of the ocean fisheries, among other things. Due to interannual variation in these conditions, the SSM parameters vary between cohorts of the same stock. When evaluating the consequences of implementing a proposed salmon fishery management plan, the uncertainty in the spatial-temporal abundance of a stock and the relationship between fishing effort and harvest rate, among other things, needs to be considered. When using the SSM as a predictive tool, this uncertainty includes variation in parameters, initial abundances, input data, and so-called process noise.

The focus of this paper is on the application of a hierarchic structure to model the interannual variation in SSM parameter values for a single coho salmon stock. The top level of the hierarchy is a probability distribution, a hyperdistribution, for the state-space model parameters. The bottom level is the state-space model, which is a combination of the unobservable abundance and the observable catch recoveries of tagged salmon.

The remainder of the paper is organized as follows. The available data are first discussed. The SSM for the abundance and catch recoveries of a coho salmon cohort is described in Section 3. In Section 4, the procedure for calculating maximum likelihood estimates of

the parameters for the hyperdistribution is discussed. An application to six consecutive cohorts of tagged coho salmon released from the same hatchery is presented in Section 5. A demonstration of evaluating a proposed management plan, quantified most simply in terms of proposed fishing effort levels, is also given. The demonstration includes variation in SSM parameters as well as the other sources of variation mentioned above. The final section is a discussion of results along with areas of current research.

## 2. DATA GENERATION

### 2.1 CODED-WIRE TAG RELEASES AND RECOVERIES

For over 20 years, salmon hatcheries in western North America raising and releasing juvenile coho salmon have tagged and externally marked subsamples of the releases. The most common tag is a coded-wire tag, a small binary coded metal pin, inserted in the snout of the juvenile salmon. The binary coding is used to identify a fish by release year and location of release. These tags cannot be decoded without killing the fish and the presence of a coded-wire tag has been made evident by a fin clip whereby the small adipose fin is removed at the time of tagging.

After release (usually in the spring), some of these fish are caught 12–18 months later in marine and freshwater fisheries taking place at different times along the coast of western North America (from the waters off central California to southeastern Alaska). Other fish escape the fisheries and return to the hatchery of origin (or stray to natural spawning grounds), generally during the fall of the second year after release. Fish caught or escaping during the first year of release are labeled age 2 fish, while those caught or escaping the following year (January–December) are age 3 fish. The vast majority of recoveries are of age 3 fish.

Fishery catches and escapement to natural spawning areas are sampled for coded-wire tags, while hatchery returns are often completely enumerated. Recoveries of tagged fish that have had their tags decoded are called observed recoveries. Estimates of the total number of tagged fish caught, naturally escaping, or returning to a hatchery are called expanded recoveries. A publicly available coastwide database, maintained by the Pacific States Marine Fisheries Commission (<http://www.psmfc.org/>), contains a listing of nearly all the coded-wire tag releases of salmon, the number of fish tagged with a particular code, and the number of observed and expanded recoveries of these fish. For each observed recovery, the date of recovery is recorded, as is the type of fishery. The exact location of the recovery is generally not available because catch samples are usually taken at ports and the location of the catches is at best defined as a somewhat broad geographic region (referred to as a catch and/or statistical area). Thus, the temporal resolution of recovery information is moderately fine (exceptions being catches from freezer boats, which can stay out a week or longer), while the spatial resolution is moderately coarse (exceptions being hatchery recoveries).

## 2.2 FISHING EFFORT

Information about fishing effort is not centrally located, nor have standards for quantifying fishing effort been established between states and the province. Individual state or provincial fishery agencies do, however, record the effort in their waters and generally use either boat landings or boat days as a measure of marine commercial fishing effort. The temporal resolution is usually weekly (or biweekly), and the spatial resolution generally matches that used for tag recoveries. Information on recreational fishing effort is often temporally coarser (monthly) and less precise.

## 2.3 DATA USED IN THE APPLICATION

Evaluation of annual management plans for ocean fisheries includes examining the effects on the catch and escapement of over 50 different coho salmon stock groupings, dispersed from northern California to southeastern Alaska. Just one stock in these stock groupings is analyzed in this paper, however. The stock is released annually from the Washington Department of Fisheries and Wildlife Humptulips hatchery located on the west coast of the Olympic Peninsula ( $47^{\circ}\text{N}$ ,  $124^{\circ}\text{W}$ ).

This stock was chosen specifically because the majority of its fishery recoveries are from marine commercial troll fisheries located near the coasts of Oregon, Washington, and British Columbia. Less than 5% of all recoveries come from recreational fisheries and “inside” marine fisheries, primarily Puget Sound and Georgia Straits, which simplifies modeling of the spatial distribution. Thus, just the effort expended by the commercial troll fleet, available on a weekly basis in boat days, was used.

Because of variation in sampling rates between different recovery points, rather than using observed recoveries, expanded recoveries (of age 3 fish) were used.

## 3. A SALMON HARVEST AND MIGRATION SSM

A state-space model is a model for two parallel time series, an unobservable series, the state process,  $\mathbf{n}_t$ ,  $t = 0, 1, \dots, T$ , and an observable series, the observation process,  $\mathbf{c}_t$ ,  $t = 1, 2, \dots, T$  (West and Harrison 1997). It is usually assumed that, given the past state,  $\mathbf{n}_{t-1}$ ,  $\mathbf{n}_t$  is independent of all other previous states (the state process is first-order Markovian) and, given  $\mathbf{n}_t$ ,  $\mathbf{c}_t$  is independent of all other states.  $\mathbf{n}_0$  is the initial value for the state process, generating  $\mathbf{n}_t$ , which subsequently generates the first realization of the observation process,  $\mathbf{c}_1$ .

Newman (1998) used a normal dynamic linear model version of an SSM for modeling the weekly survival and movement of a single cohort of coho salmon. Frequent interaction with salmon biologists and experimenting with new data sets has led to a slightly different formulation for some of the SSM subcomponents. This newer formulation is used in the application and is what is described here.

The spatial framework is a line along the west coast of North America from southern Oregon to northern British Columbia. The line is partitioned into  $A$  nonoverlapping

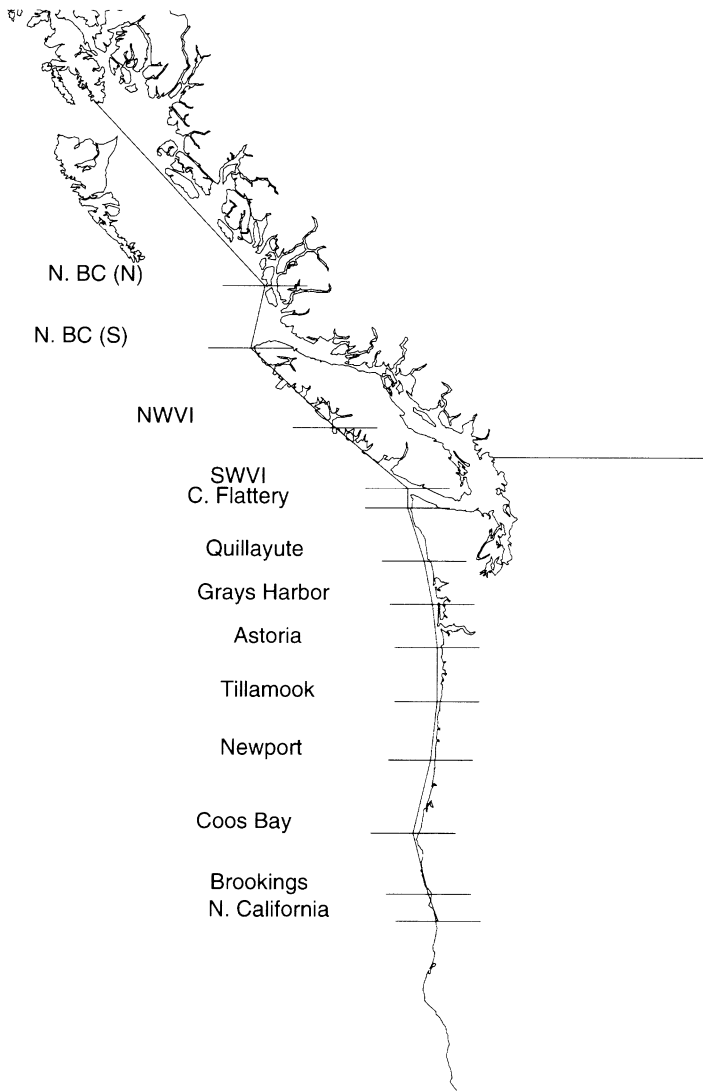


Figure 1. Fishery Management Areas Along the Pacific Northwest Coast With Line Segments Marking the Approximate Northern and Southern Boundaries of Each Area. The 12 catch areas modeled were those from Brookings to Northern British Columbia (N), inclusive. The hatchery stocks analyzed in the application are located in the Gray's Harbor area.

segments, a simplification of the two dimensional fishery management areas used to designate tag recovery locations (see Figure 1). The time resolution is weekly, consisting of  $T$  weeks during the second summer spent in the ocean, usually beginning sometime in June, when harvest begins, and ending sometime in October, when most fish have returned to freshwater. In the application,  $A = 12$  and  $T = 16$ .

The observation process,  $c_t$ , is a length- $A$  vector of expanded recoveries of age 3 tagged fish in each fishery management area. There is no distinction between the types of fishery

recoveries in an area, i.e., recreational fishery and commercial fishery recoveries are combined. Recoveries at the hatchery are treated as recoveries in the fishery management area adjacent to the natal stream.

The state process,  $\mathbf{n}_t$ , is a corresponding vector of unobservable abundances of the tagged cohort in each of the  $A$  areas during the weeks  $t = 0, 1, \dots, T$ .  $\mathbf{n}_0$  is the initial abundance of tagged fish the week prior to fishing.

The particular form of the SSM is

$$\mathbf{n}_t \mid \mathbf{n}_{t-1} = M_t S_t \mathbf{n}_{t-1} + \mathbf{w}_t, \quad t = 1, \dots, T, \quad (3.1)$$

$$\mathbf{c}_t \mid \mathbf{n}_t = H_t \mathbf{n}_t + \mathbf{v}_t, \quad t = 1, \dots, T. \quad (3.2)$$

$S_t$ ,  $M_t$ , and  $H_t$  are  $A \times A$  survival, movement, and harvest matrices. Both  $S_t$  and  $H_t$  are diagonal matrices—survival and harvest in one area has no effect on survival and harvest in another area during the same time period. The state process error component,  $\mathbf{w}_t$ , is an  $A \times 1$  vector following a multivariate normal distribution with mean zero and covariance matrix  $\Sigma_{w,t}$ . Similarly, the observation process error component,  $\mathbf{v}_t$ , is  $A \times 1$  and multivariate normal with mean zero and covariance matrix  $\Sigma_{v,t}$ . The covariance matrices  $\Sigma_{w,t}$  and  $\Sigma_{v,t}$  are functions of the proportions in  $S_t$ ,  $M_t$ ,  $H_t$  as well as the abundance and are based on approximations to the variances and covariances for sums of independent multinomial random variables (details in the appendix of Newman (1988)).

The structure of the SSM can be conceptually partitioned into three modules—the initial distribution of the abundance ( $\mathbf{n}_0$ ), survival and mortality, and movement. This perspective has led to considerable experimentation in formulations for each module and has generated suggestions by biologists to evaluate different theories.

### 3.1 INITIAL DISTRIBUTION

To begin the process, the initial state vector,  $\mathbf{n}_0$ , is found by first calculating the expected number of fish alive. A known number,  $R$ , of tagged salmon leave the freshwater natal area (e.g., a hatchery) and, assuming independence between the fish, the expected number surviving to the beginning of the harvest season is  $R\theta_s$ , where  $\theta_s$  is the unknown survival rate. The survivors are assumed to distribute themselves along the line segment according to a (scaled) beta( $\theta_{i1}, \theta_{i2}$ ) distribution (the subscript  $i$  is for initial). The expected numbers per area are used to construct  $\mathbf{n}_0$ .

### 3.2 SURVIVAL AND MORTALITY

For the survival matrix  $S_t$  (and, relatedly, the harvest matrix  $H_t$ ), there are three parameters— $\theta_{q_{U.S.}}$ ,  $\theta_{q_{Canada}}$ , and  $\theta_n$ . The first two correspond to catchability coefficients found in the Baranov catch equations (Ricker 1975) and the third reflects natural mortality during the harvest season. The diagonal elements of  $S_t$  are the probabilities of surviving both fishing-related and natural forces of mortality. In particular, for probabilities of survival from time  $t$  to  $t + 1$  for an area  $a$  in U.S. waters and an area  $b$  in Canadian waters, the

corresponding diagonal elements of  $S_t$  are

$$\begin{aligned} S_{t+1}[a, a](\text{U.S.}) &= \exp[-M - F_{a,t}(\text{U.S.})] \\ S_{t+1}[b, b](\text{Canada}) &= \exp[-M - F_{b,t}(\text{Canada})], \end{aligned}$$

where

$$\begin{aligned} F_{a,t}(\text{U.S.}) &= 0.00002\theta_{q_{\text{U.S.}}} \frac{\text{effort}_{a,t}}{(a_R - a_L)} \\ F_{b,t}(\text{Canada}) &= 0.00002\theta_{q_{\text{Canada}}} \frac{\text{effort}_{b,t}}{(b_R - b_L)} \\ M &= 0.001\theta_n. \end{aligned}$$

The terms 0.00002 and 0.001 are scaling constants that keep the parameter estimates of approximately the same magnitude and thus improve the optimization accuracy (Gill, Murray, and Wright 1981). The terms  $a_R - a_L$  and  $b_R - b_L$  are the lengths of fishing areas  $a$  and  $b$ , thus scaling the effort measure to effort per unit length; e.g. 500 units of effort in a short segment should have greater impact than 500 units spread across a longer segment. The parameter  $\theta_n$  is a natural mortality parameter (scaled by 0.001), assumed constant between time periods. In the absence of fishing,  $\exp(-0.001\theta_n)$  is the survival rate in a single time period. For the first time period,  $t = 1$ , 100% survival is assumed ( $S_1 = I$ , an identity matrix).

In the final time period, the harvest rate in all areas is fixed at 100%, and all hatchery recoveries made during later weeks (after time  $T$ ) and those made in the final few weeks (just prior to  $T$ ) are aggregated into this time period. There are several reasons for this artificial manipulation of harvest and recoveries. There is no or minimal ocean fishing taking place at this time and thus there is no information about abundance in the ocean. One reason for the lack of ocean fishing for coho salmon in the later time periods is that many of the returning cohort may be in the natal river or just milling in the vicinity of the stream mouth (Sandercock 1991). Nearly all recoveries made in the final time periods are made at the hatchery (occasional exceptions being freshwater fishery recoveries). The dates of hatchery recoveries can extend over several months, but the majority are made during an interval that usually begins the last quarter of the time series (mid-September) and ends sometime in October. For the modeling of ocean distribution and migration, however, once the fish are in the natal area, their further behavior is not relevant. Because the likelihood function includes a term of squared differences between observed and estimated recoveries, a final period harvest rate of 100% penalizes model parameters that result in very many fish left at sea and, in a sense, forces remaining fish into the natal area by the final time period.

### 3.3 MOVEMENT

The columns of the movement matrix at time  $t$ ,  $M_t$ , are vectors of probabilities for movement from one area to any other area. Movement beyond the natal area, however, whether north or south of the natal area, is not allowed, assuming the freshwater influence



will attract the fish back to spawn. Nor is movement outside the line segment (to the far north or south) allowed; thus, the system is closed and each column totals to one.

The probability for moving from one area to another area is based on an individual fish movement model. For any location at any time, a probability distribution for location in the next time step is defined. The probability of moving from one area to another is found by a double integration, the outer integral being over all the locations within the current area (any location within a given area assumed equally probable) and the inner integral being over the line segment defining the other area.

For a fish at location  $p$  at time  $t$ ,  $p_t$ , the probability distribution for its location at time  $t + 1$ , is  $\text{beta}(\alpha, \beta)$ , with parameters that are functions of location and time. The  $\alpha$  parameter is a function of time and location, while the  $\beta$  parameter depends on time alone. The formulation is made in terms of the expected value,  $\mu(p_t, t)$ , and the parameter  $\beta(t)$ , such that

$$\mu(p_t, t) = (p_t + \epsilon) \left[ \frac{\exp(4.6 - 0.01\theta_{m1}t)}{1.0 + \exp(4.6 - 0.01\theta_{m1}t)} \right] \quad (3.3)$$

$$\beta(t) = 1.0 + 200.0\theta_{m2} \frac{\exp(4.6 - t^2/T)}{1 + \exp(4.6 - t^2/T)}, \quad (3.4)$$

where  $\theta_{m1}$  and  $\theta_{m2}$  are unknown parameters (with subscript  $m$  denoting movement). In the remainder of this section,  $\mu(p_t, t)$  and  $\beta(t)$  will simply be denoted as  $\mu$  and  $\beta$ . The parameter  $\alpha$  of the beta distribution is found by

$$\alpha = \mu\beta/(1 - \mu),$$

where the dependence of  $\mu$  and  $\beta$  on location and time is implicit.

The expected value of the next location,  $\mu$ , is a product of the current location (and a slight offset,  $\epsilon$ , to avoid numerical integration problems with  $p_t = 0$ ) and a multiplier  $\leq 1.0$  (a logit function) that, early in the time period ( $t$  near zero), is nearly 1.0. As time increases, the multiplier shrinks toward zero, which is the location on the line of the natal area.

The constant 200.0 in Equation (3.4) is a scaling constant for  $\theta_{m2}$  (to improve optimization accuracy). The logit function multiplier shrinks toward zero as time increases. The constant 1.0 is used to bound  $\beta$  below by 1.0, thus preventing bathtub-shaped distributions, where probability is concentrated at the extreme end of the distribution away from the natal area.

### 3.4 PARAMETER ESTIMATION

The total number of unknown parameters is eight; i.e.,  $\Theta = (\theta_s, \theta_{i1}, \theta_{i2}, \theta_{qU.S.}, \theta_{qCanada}, \theta_n, \theta_{m1}, \theta_{m2})$ , with the first three being related to the initial abundance and spatial distribution, the next three being related to survival during the fishing season, and the last two being related to movement. The Kalman filter is used to evaluate the likelihood with respect to  $\Theta$ , and maximum likelihood estimates can be calculated. Conditional on maximum likelihood estimates of  $\Theta$  and the observation process, smoothed estimates of

abundance,  $E(\mathbf{n}_t \mid \hat{\Theta}, \mathbf{c}_1, \dots, \mathbf{c}_T)$ , may be calculated using the recursive Kalman smoothing algorithm (Shumway 1988).

In the application, six of the eight parameters were estimated and two were fixed, namely  $\theta_{i2} = 2.0$  and  $\theta_n = 0.0$ . The initial distribution parameter  $\theta_{i2}$  can be estimated, but the estimates are often quite variable. Given the coarse partitioning of the coastline into large management areas, many different combinations of the initial distribution parameters are nearly equivalent in terms of goodness of fit. Setting  $\theta_{i2}$  to 2.0 still allows considerable flexibility in the initial distribution. The natural mortality parameter,  $\theta_n$ , is at times highly correlated with the initial survival parameter,  $\theta_s$ , and can be difficult to estimate. With  $T = 16$  weeks, the natural mortality is assumed to be relatively slight for these maturing adult fish. With  $\theta_n = 0$ , the survival matrix is simply the previous period harvest matrix,  $S_{t+1} = I - H_t$ .

#### 4. HIERARCHIC MODEL

To account for between-cohort variation in the SSM parameter values, a hierarchic model was formulated. The top level of the hierarchy was a hyperdistribution that generates the cohort-specific values of the SSM parameters,  $\Theta$ . The hierarchy for a generic SSM can be written as follows:

$$\begin{aligned} \text{hyperdistribution: } & h(\Theta \mid \Omega), \\ \text{state process: } & g(\mathbf{n}_t \mid \mathbf{n}_{t-1}, \Theta) \quad t = 1, \dots, T, \\ \text{observation process: } & f(\mathbf{c}_t \mid \mathbf{n}_t, \Theta) \quad t = 1, \dots, T, \end{aligned}$$

where  $h$ ,  $g$ , and  $f$  are probability density functions and  $\Omega$  is a vector of hyperparameters. For a normal dynamic linear model,  $g$  and  $f$  are normal density functions, as implied by Equations (3.1) and (3.2).

##### 4.1 ESTIMATING $\Omega$

Let  $C_k$  be all the observation data for cohort  $k$ ,  $C_k = \mathbf{c}_{1,k}, \dots, \mathbf{c}_{T,k}$ , where each component in the time series is a vector of tag recoveries from cohort  $k$  in the  $A$  management regions. Denote the set of data from  $K$  cohorts by  $C^K = (C_1, C_2, \dots, C_K)$  with an associated set of cohort-specific parameters,  $\Theta^K = (\Theta_1, \Theta_2, \dots, \Theta_K)$ . To find the maximum likelihood estimate of  $\Omega$ , the marginal log likelihood,  $\log(L(\Omega \mid C^K)) = \log(f(C^K \mid \Omega))$ , is needed. Evaluating the likelihood requires integrating over the SSM parameter space for each cohort, i.e.,

$$f(C^K \mid \Omega) = \int f(C^K \mid \Theta^K) h(\Theta^K \mid \Omega) d\Theta^K. \quad (4.1)$$

The argument  $f(C^K \mid \Theta^K)$  is the likelihood for all  $K$  individual cohort state-space models, with the implicit assumption that, conditional on  $\Theta^K$ , the observations are independent of  $\Omega$ . A further assumption is that, conditional on  $\Theta_1, \Theta_2, \dots, \Theta_K$ , the individual cohorts are

independent. Thus,

$$f(C^K | \Theta^K) = \prod_{k=1}^K f(C_k | \Theta_k),$$

where each  $f(C_k | \Theta_k)$  can be found by the Kalman filter.

The integration in Equation (4.1) is usually impossible analytically, and the proposed solution is to use Monte Carlo methods. For example, for a given  $\Omega$ , generate a sample from  $h(\Theta^K | \Omega, C^K)$ , say  $\Theta_1^{*,1}, \Theta_2^{*,1}, \dots, \Theta_K^{*,1}, \Theta_1^{*,2}, \dots, \Theta_K^{*,N}$ , and estimate  $f(C^K | \Theta^K)$  with

$$\hat{f}(C^K | \Omega) = \frac{1}{N} \sum_{i=1}^N \left[ \prod_{k=1}^K f(C_k | \Theta_k^{*,i}) \right], \quad (4.2)$$

which will converge in probability to  $E[f(C^K | \Theta^K, \Omega)]$ , with the expectation taken with respect to the distribution of  $\Theta^K | \Omega, C^K$ .

A practical problem with the above Monte Carlo solution is that  $\hat{f}(C^K | \Omega)$  is particular to the value of  $\Omega$  used. When searching for the maximum likelihood estimate for  $\Omega$ , a prohibitive number of simulations is required, one set of simulations per candidate value of  $\Omega$ . A more efficient approach suggested by Geyer (1996) is to use importance sampling.

The idea of importance sampling as used here is to generate one sample from the wrong distribution, known as the importance distribution, and then make adjustments to the sum used for Monte Carlo integration so that convergence is to the desired integral. For example, suppose the desired integral is  $\int \phi(x)f(x)dx$ . Sample  $n$  times from  $g(x)$  and estimate the integral with  $n^{-1} \sum_{i=1}^n \phi(x_i)f(x_i)/g(x_i)$  (Ripley 1987). So long as  $g$  dominates  $f$  for all  $x$  in the range being integrated over, namely  $g(x) = 0 \Rightarrow f(x) = 0$ , the sum converges to  $\int \phi(x)f(x)dx$ .

For this application, the importance distribution is denoted  $h(\Theta^K | \Psi, C^K)$ , where  $\Psi$  is a fixed vector of hyperparameter values such that  $h$  dominates  $h(\Theta^K | \Omega, C^K)$  for all possible values of  $\Omega$  and the observed  $C^K$ . The advantage is that a single set of simulations from the importance distribution can be used repeatedly for different values of  $\Omega$ . Geyer recommends the following objective function:

$$l(\Omega) = \log \left[ \frac{f(C^K | \Omega)}{f(C^K | \Psi)} \right]. \quad (4.3)$$

To see how importance sampling is used, observe that Equation (4.3) can be reexpressed as

$$l(\Omega) = \log \left[ E_{\Psi} \left( \frac{f(C^K, \Theta^K | \Omega)}{f(C^K, \Theta^K | \Psi)} \right) \right], \quad (4.4)$$

where the expectation is taken with respect to the probability function  $h(\Theta^K | \Psi, C^K)$ . Note

$$E_{\Psi} \left( \frac{f(C^K, \Theta^K | \Omega)}{f(C^K, \Theta^K | \Psi)} \right) = \int \frac{f(C^K, \Theta^K | \Omega)}{f(C^K, \Theta^K | \Psi)} h(\Theta^K | \Psi, C^K) d\Theta^K$$

$$\begin{aligned}
&= \int \frac{f(C^K, \Theta^K | \Omega)}{f(C^K, \Theta^K | \Psi)} \frac{f(C^K, \Theta^K | \Psi)}{f(C^K | \Psi)} d\Theta^K \\
&= \frac{1}{f(C^K | \Psi)} \int f(C^K, \Theta^K | \Omega) d\Theta^K \\
&= \frac{f(C^K | \Omega)}{f(C^K | \Psi)}.
\end{aligned}$$

Thus, given a sample from  $h(\Theta^K | \Psi, C^K)$ , (4.4) can be estimated by

$$\begin{aligned}
\hat{l}(\Omega) &= \log \left[ \frac{1}{N} \sum_{i=1}^N \frac{\prod_{k=1}^K f(C_k | \Theta_k^{*,i}) h(\Theta_k^{*,i} | \Omega)}{\prod_{k=1}^K f(C_k | \Theta_k^{*,i}) h(\Theta_k^{*,i} | \Psi)} \right] \\
&= \log \left[ \frac{1}{N} \sum_{i=1}^N \frac{\prod_{k=1}^K h(\Theta_k^{*,i} | \Omega)}{\prod_{k=1}^K h(\Theta_k^{*,i} | \Psi)} \right]. \tag{4.5}
\end{aligned}$$

Interestingly, the data,  $C^K$ , disappear from the objective function [Eq. (4.5)], although the data are implicitly present in the generated  $\Theta^K$ .

The choice of the parameter for the importance sampling distribution,  $\Psi$ , is critical to the efficiency and accuracy of the estimated objective function [Eq. (4.5)]. Convenient starting values for  $\Psi$  are maximum likelihood estimates, or method of moments estimates, of  $\Omega$  based on the hyperdistribution for  $\Omega$  with maximum likelihood estimates of  $\Theta$  treated as data. For example, let  $\Psi$  be the value maximizing  $\prod_{k=1}^K h(\hat{\Theta}_k | \Psi)$ , where  $\hat{\Theta}_k$  is the maximum likelihood estimate of  $\Theta_k$  for cohort  $k$ , conditioning on  $C_k$  alone.

To further improve efficiency and accuracy, Geyer (1996) recommends an iterative optimization—after finding the maximum likelihood estimate (MLE) for  $\Omega$  in the  $i$ th iteration,  $\hat{\Omega}^i$ , based on  $\Psi^i$ , set  $\Psi^{i+1}$  equal to  $\hat{\Omega}^i$  and reestimate  $\Omega$ . The iteration ends once the values of  $\Psi$  and  $\hat{\Omega}$  nearly coincide.

## 4.2 MARKOV CHAIN MONTE CARLO GENERATION OF $\Theta^K | C^K, \Psi$

To estimate the objective function [Eq. (4.5)], a sample from  $h(\Theta^K | \Psi, C^K)$  is needed. The density function,

$$\begin{aligned}
h(\Theta^K | \Psi, C^K) &= \frac{f(C^K, \Theta^K | \Psi)}{f(C^K | \Psi)} \\
&= \frac{f(C^K | \Theta^K) h(\Theta^K | \Psi)}{f(C^K | \Psi)},
\end{aligned}$$

will typically be difficult to find in closed form because the denominator,  $f(C^K | \Psi)$ , is unknown and is in fact the function of interest [see Eq. (4.3)]. The Metropolis–Hastings algorithm (Hastings 1970) provides a means of generating a sample from any probability distribution known up to at least a constant of proportionality. In this case, the unknown constant is the denominator  $f(C^K | \Psi)$ .

The Metropolis–Hastings algorithm was applied to a single cohort at a time in the application. The steps of the algorithm for this SSM setting are as follows (Gilks,

Richardson, and Spiegelhalter 1996): To begin the simulation, specify initial starting values  $\Theta_1^0, \dots, \Theta_K^0$ . For the  $i$ th simulation of the  $k$ th cohort,

- (1) generate a candidate value  $\Theta_k^{*,i}$  from a proposal distribution  $q(\Theta_k)$ ;
- (2) calculate

$$\alpha = \min \left( 1, \frac{f(C_k | \Theta_k^*) h(\Theta_k^* | \Psi)}{f(C_k | \Theta_k^{i-1}) h(\Theta_k^{i-1} | \Psi)} \frac{q(\Theta_k^{i-1})}{q(\Theta_k^*)} \right);$$

- (3) generate a uniform(0, 1) random variable  $U$ ;
- (4) if  $U < \alpha$ , set  $\Theta_k^i = \Theta_k^*$ ; else  $\Theta_k^i = \Theta_k^{i-1}$ .

Assuming the proposal distributions  $q(\Theta_k)$  dominate  $h(\Theta_k | \Psi)f(C_k | \Theta_k)$ , then  $\Theta^i \rightarrow \Theta | C^K, \Psi$  in distribution (Tierney 1996). Thus, an infinite variety of proposal distributions will work in theory. In practice, the choice of the proposal distribution affects the acceptance rate, how often  $U < \alpha$ , the rate of convergence to the desired distribution, and the degree of mixing or rapidity of movement around the sample space (Roberts 1996).

Several iterations of the chain must be made (a burn-in period of length  $B$ ) before simulated values ostensibly converge to the desired distribution. The (estimated) objective function to maximize is

$$\hat{l}(\Omega) = \log \left[ \frac{1}{N-B} \sum_{i=1}^{N-B} \frac{\prod_{k=1}^K h(\Theta_k^{*,i} | \Omega)}{\prod_{k=1}^K h(\Theta_k^{*,i} | \Psi)} \right]. \quad (4.6)$$

Practical implementation of Markov chain Monte Carlo (MCMC) requires making the following three choices:

- proposal distributions  $q(\Theta_k)$ ,
- $B$ , the length of the burn-in period,
- $N$ , the length of the chain.

The proposal distributions were handcrafted for the particular application to coho salmon and are discussed in Section 5. To determine  $B$  and  $N$ , the method of Raftery and Lewis (1996) was used. Raftery and Lewis analyze the output from the realization of a single chain to determine the length of the chain  $N$  and the burn-in period  $B$  necessary to achieve a specified precision for estimating the upper and lower 0.025 quantiles of the sampling distribution of each parameter. Several short runs were also made using different tuning parameter values for the proposal distribution to examine the degree of consistency in recommended chain lengths and burn-in periods.

### 4.3 VARIANCE ESTIMATES FOR $\hat{\Omega}$

There are several approaches, in principle, to estimating the variance of  $\hat{\Omega}$ . One is to first use Monte Carlo methods to approximate the Hessian of the negative log likelihood, the observed information matrix, and use the inverse of the Hessian as an estimate of the covariance matrix for  $\hat{\Omega}$  (Geyer 1996). Differentiating the log likelihood even once, however, is analytically intractable given how  $\Theta^K$  enters  $f(C^K | \Theta^K)$ . Even in the nonhierarchic

setting, Newman (1998) used numerical approximations to the Hessian to estimate the variances of  $\hat{\Theta}_k$ .

The approach chosen here is the jackknife, where one MCMC sample is needed per cohort. First, the data for cohort  $k$  is removed and the MLE for  $\Omega$  is estimated by

- (1) generating an MCMC sample from  $f(\Theta_{-k}^K | C_{-k}^K, \hat{\Omega})$ ,
- (2) maximizing Equation (4.6) with  $\hat{\Omega}$  used in place of  $\Psi$ .

The jackknife estimate of the variance–covariance matrix is

$$\hat{\Sigma}_{\mathcal{J}} \{(\hat{\omega}_i, \hat{\omega}_j)\} = \frac{K-1}{K} \sum_{k=1}^K (\hat{\omega}_{i,-k} - \bar{\omega}_{i,-})(\hat{\omega}_{j,-k} - \bar{\omega}_{j,-}), \quad (4.7)$$

where  $\hat{\omega}_{i,-k}$  is the estimate for parameter  $i$  when the  $k$ th cohort is omitted and  $\bar{\omega}_{i,-}$  is the average of the  $K$   $\hat{\omega}_{i,-k}$ .

A third, extremely computer-intensive approach is a parametric bootstrap. It requires  $\mathcal{B}$  MCMC samples, however, where  $\mathcal{B}$  is suitably large, say 200 (Efron and Tibshirani 1993). (For additional discussion of this issue, see Carlin and Louis (1996).)

#### 4.4 INFERENCE FOR $\Theta_k$

Conditional on maximum likelihood estimates of the hyperparameters  $\Omega$ , Markov chain Monte Carlo can again be used to generate samples from  $h(\Theta^K | \hat{\Omega}, C^K)$ . An empirical estimate of the distribution can be constructed from the resulting samples. Empirical Bayes estimates of the  $\theta$ 's are an easy by-product of the chain,

$$\hat{E}[\Theta_k | \hat{\Omega}, C^K] = \frac{1}{N-B} \sum_{i=B+1}^N \Theta_k^i,$$

where the sum asymptotically converges to  $E[\Theta_k | \hat{\Omega}, C^K]$  (Tierney 1996).

The (conditional) standard error of this estimate can be estimated from the standard deviation of the simulated  $\Theta_k$ . These estimates, however, condition on  $\hat{\Omega}$ , when in fact one would like to be conditioning on the true value,  $\Omega$ . The correct variance, conditioning on the true  $\Omega$  and data, is

$$\text{var}(\Theta_k | C^K, \Omega) = \text{var}\left[E(\Theta | C^K, \Omega, \hat{\Omega})\right] + E\left[\text{var}(\Theta | C^K, \Omega, \hat{\Omega})\right].$$

Thus, the estimated variance is an underestimate. Potential approaches for accounting for this uncertainty are mentioned later in the paper. In the application, however, the data dominates the conditional estimates of  $\Theta$  relative to the effect of the hyperparameters, and the severity of underestimation is likely minimal.

## 5. APPLICATION

The hierarchic state-space model was fit to 6 years of coded-wire tag recoveries from six cohorts of coho salmon released as juvenile fish from the Humptulips hatchery. The fish were released during the spring months of 1984 through 1989 and later recovered in

ocean fisheries or at the hatchery as age 3 fish during the summer and early fall months of 1986 through 1991, respectively. The number released in 1984 through 1989 were 150,498, 63,360, 76,714, 78,760, 70,584, and 80,082, and the corresponding expanded numbers of recoveries in 1986 through 1991 were 7,179, 849, 4,668, 2,928, 1,527, and 4,234. Most recoveries were made during the months of July–October. The percentage of recoveries made at the hatchery (or somewhere in the natal area) ranged from 40 to 68%. The total commercial troll fishing effort ranged from 97,000 to 127,000 boat days, with roughly 66–84% of the total effort coming from the Canadian fishery. The number of expanded recoveries for each cohort by week and catch area, the associated commercial troll effort, and the boundaries for the catch areas are available at the URL <http://www.uidaho.edu/~newman/ssm.html>.

### 5.1 HYPERDISTRIBUTION

The (multivariate) hyperdistribution  $h(\Theta \mid \Omega)$  was formed by combining six independent univariate distributions, one for each of the six free parameters of  $\Theta$ . For the initial survival parameter, expressed in percent, a beta distribution scaled to the interval 0–10 was used. For the remaining parameters, gamma distributions were used. The beta and gamma distributions provide considerable flexibility in the shape of the univariate distributions and have suitable domains for the possible values of  $\Theta$ ; e.g., the harvest parameters  $\theta_{q_{U.S.}}$  and  $\theta_{q_{Canada}}$  need to be positive. The resulting hyperdistribution included both fixed and free (hyper)parameters and is

$$\begin{aligned}\theta_s &\sim \text{beta}(\omega_1, 2.3) \\ \theta_{i1} &\sim \text{gamma}(\omega_2, 0.1) \\ \theta_{q_{U.S.}} &\sim \text{gamma}(\omega_3, 5.0) \\ \theta_{q_{Canada}} &\sim \text{gamma}(\omega_4, 2.0) \\ \theta_{m1} &\sim \text{gamma}(\omega_5, 1.1) \\ \theta_{m2} &\sim \text{gamma}(\omega_6, 30.0).\end{aligned}$$

The fixed hyperparameter values were chosen based on the observed variation in the maximum likelihood estimates such that the hyperdistributions could generate variances about 20% greater than those observed. For example, the sample average and variance for the 6 years of maximum likelihood estimates of  $\theta_{q_{U.S.}}$  were  $\bar{x} = 10.66$  and  $s^2 = 43.16$ . The expected value for  $\theta_{q_{U.S.}}$  is  $\mu = \alpha\beta$  and the variance is  $\sigma^2 = \alpha\beta^2$ , equivalently,  $\beta = \sigma^2/\mu$ . The second shape parameter,  $\beta$ , was set equal to  $1.2 \times s^2/\bar{x} = 4.86$ , which was rounded to 5.00. The same approach was used, with minor deviations, to specify each of the fixed values.

### 5.2 MCMC

To simulate realizations of  $\Theta \mid C^K, \Psi$  using MCMC, six separate proposal distributions were used for each of the six  $\Theta$ 's and each parameter was generated one at a time. The beta

distribution, scaled to  $(0, 10)$ , was used as a proposal distribution for  $\theta_s$  with the parameters chosen such that the expected value equaled the previous value,  $\Delta_1$  say, in the chain. Log-normal distributions were used for the remaining  $\theta$ 's; again, previous values were used to center the proposal distributions. The six proposal distributions are

$$\begin{aligned} q(\theta_s \mid \Delta_1) &= \text{beta} \left( \frac{\sigma_1 \Delta_1 / 10.0}{1 - \Delta_1 / 10.0}, \sigma_1 \right) \\ q(\theta_{i1} \mid \Delta_2) &= \text{lognormal}(\log(\Delta_2), \sigma_2) \\ q(\theta_{qU.S.} \mid \Delta_3) &= \text{lognormal}(\log(\Delta_3), \sigma_3) \\ q(\theta_{qCanada} \mid \Delta_4) &= \text{lognormal}(\log(\Delta_4), \sigma_4) \\ q(\theta_{m1} \mid \Delta_5) &= \text{lognormal}(\log(\Delta_5), \sigma_5) \\ q(\theta_{m2} \mid \Delta_6) &= \text{lognormal}(\log(\Delta_6), \sigma_6). \end{aligned}$$

The proposal distributions were chosen largely because of the simplicity of simulating values from them and the flexibility of their shapes, which could be readily manipulated to increase the rate of mixing (coverage of the sample space) and rate of convergence. The parameters,  $\sigma_1, \dots, \sigma_6$ , are tuning parameters chosen to achieve both these objectives. They were selected partially by trial and error on the basis of the acceptance rate and by examining the estimated standard errors of the maximum likelihood estimates of the  $\theta$ . If acceptance rates were much higher than 80% or lower than 20% (see Roberts (1996) for a discussion of choice of proposal distributions and desirable acceptance rates), the tuning parameters were changed. Another factor in the decision was the estimated rate of convergence using the method of Raftery and Lewis (1996).

The initial values for  $\Psi$ , the parameters of the importance sampling distribution, for the MCMC simulations were approximately method of moments estimates of  $\Omega$ , which were calculated using the sample averages of the six cohorts' estimates of the  $\theta$  and the values of the fixed hyperparameters. For example, the sample average of  $\hat{\theta}_{i1}$  was 2.40. According to the hyperdistribution, the expected value of  $\theta_{i1}$  is  $\omega_2 \times 0.1$ ; thus, an estimate of  $\omega_2$  is  $2.40/0.10 = 24$ . This value was modified to 30 to allow for the inflation in the variance of the hyperdistribution and was used as an initial value for the corresponding  $\Psi$  value. Additional simulations were started with different values for  $\Psi$  to assess sensitivity of results.

The initial values for the simulation of the 36  $\theta$ 's (six cohorts and six parameters) were the average values (or nearly so) for the 6 years of maximum likelihood estimates (based on the nonhierarchical SSM). The chain was run 1,000 times and the simulated  $\Theta$ 's were input to Raftery and Lewis's *gibbsit* (1996) routine to estimate chain length and burn-in time. The Raftery and Lewis procedure is directed at yielding sufficiently precise estimates of the lower and upper quantiles for the sampling distribution of  $\Theta \mid C^K, \Psi$ ; this is not necessarily the required chain length for estimating the likelihood function with sufficient precision, but it appeared adequate.

Based on *gibbsit*, the median necessary chain length for the 36 parameters was 3,800 with a maximum burn-in of 57 (the worst case was  $\theta_{m2}$ , for the 1990 cohort, requiring 10,000 iterations). The likelihood [Eq. (4.6)] was evaluated using the initial chain of 1,000



with a burn-in of 100, and the maximum likelihood estimate for  $\Omega$  was calculated using the optimization program NPSOL (Gill, Murray, Saunders, and Wright 1986). The MLE for  $\Omega$  then replaced the initial  $\Psi$  and another chain of length 5,000 was generated. There was no significant change in the estimate of  $\Omega$  using the second chain (with a burn-in period of 100), and these estimates are the ones shown here. Stability in the parameter estimates, however, was observed for relatively short chains. Jackknife estimates of the variance-covariance matrix for  $\hat{\Omega}$  were calculated from the final chain by deleting the generated values for the  $i$ th year and estimating  $\Omega_i$ . This same chain was also used to estimate the conditional distributions of the  $\Theta$ 's. The resulting maximum likelihood estimates for  $\Omega$  and the sampling distribution of the  $\Theta$ 's are discussed next.

### 5.3 ESTIMATES OF HYPERPARAMETERS, $\Omega$

The MLEs for  $\Omega$  are shown in Table 1 along with the jackknife estimates of the standard errors. The least precisely estimated parameters, as measured by coefficient of variation, are  $\omega_2$  (for initial location) and  $\omega_6$  (the second movement parameter).

Six years of data is not a lot of information, and inferences about the hyperdistributions for the SSM parameters should be viewed with caution. However, a demonstration of the types of interpretation that can be done is useful. Estimates of the expected values and standard deviations for the  $\Theta$  parameters of the SSMs can be made conditional on  $\hat{\Omega}$  (Table 1). For example, the average survival rate between time of release to the beginning of the final year in the ocean is estimated to be 4.48%, within the range of other independently made estimates of coho survival (Bilton, Alderdice, and Schnute 1982).

The estimated initial spatial distribution is on average  $\text{beta}(2.43, 2.0)$ , which places the mode of the distribution on the northwest coast of Vancouver Island.

On average, the harvest rate per unit of troll effort (a boat day) is estimated to be higher for U.S. fisheries ( $E(\theta_{q_{\text{U.S.}}} | \hat{\Omega}, C^K) = 11.25$ ) than for Canadian fisheries ( $E(\theta_{q_{\text{Canada}}} | \hat{\Omega}, C^K) = 5.18$ ), though these parameters are highly variable and interpretation is confounded by the exclusion of recreational fishery effort and errors in the measurement of effort.

Table 1. MLEs for Hyperparameters  $\Omega$  and Jackknife Estimates of Standard Errors. The conditional expected values and standard deviations for the corresponding SSM parameters are denoted  $E(\theta | \hat{\omega})$  and  $SD(\theta | \hat{\omega})$ . The observed averages and standard deviations for the (cohort-specific) maximum likelihood estimates of the SSM parameters are denoted  $\bar{\theta}$  and  $s(\hat{\theta})$ .

Hyperparameter	$\hat{\omega}$	$\widehat{SE}(\hat{\omega})$	$E(\theta   \hat{\omega})$	$SD(\theta   \hat{\omega})$	$\bar{\theta}$	$s(\hat{\theta})$
$\omega_1$ (initial survival)	1.87	0.64	4.48	2.19	4.32	2.02
$\omega_2$ (initial distribution 1)	24.30	1.44	2.43	0.38	2.40	0.37
$\omega_3$ (U.S. fishing mortality)	2.25	0.52	11.25	7.16	10.66	6.57
$\omega_4$ (Canadian fishing mortality)	2.59	0.36	5.18	2.91	4.66	2.75
$\omega_5$ (Movement 1)	31.42	1.70	34.56	4.83	33.78	4.61
$\omega_6$ (Movement 2)	0.53	0.17	15.90	8.12	54.88	99.03

Last, to interpret the movement parameters  $(\theta_{m1}, \theta_{m2})$ , the expected movement transition matrix can be examined as a function of location and time. Early in the fishing season, weeks 1 to 4, the fish are remaining in their current area with high probability, ranging from 80 to 97%, with somewhat higher probability of moving further away from the natal area than toward it. This milling tendency gradually weakens over time and the probability of movement toward the natal area increases. For example, fish located in Cape Flattery, two regions north of the natal area, stay with probability 0.87 and move south with probability 0.09 during week 6; but by week 8, they stay with probability 0.65 and move south with probability 0.35; and by week 11, the probabilities are 0.12 and 0.88. The behavior in the final time periods for the outside regions is possibly not reasonable. Variance in the next location distribution increases such that the probabilities of moving to the extreme regions, away from the origin, can increase relative to previous time periods. The increase is slight, however, and given that a tiny fraction of the abundance is in the outer regions at later times, the effect on predicted abundance is negligible.

#### 5.4 A SIMPLER ALTERNATIVE TO THE HYPERDISTRIBUTION

A less formal approach to describing the variability in the SSM parameters between cohorts is simply to average the maximum likelihood estimates for the  $\theta$ 's and calculate the standard deviations. The results are shown in Table 1. In general, the expected averages and the observed averages are fairly close, with  $\theta_{m2}$  being an exception. The consistency between the expected and observed averages is partially due to the way the fixed hyperparameters were chosen. Given more years of data, perhaps, more or all the hyperparameters could be estimated.

Relatedly, a simple way to simulate future values of the SSM parameters (e.g., to evaluate fishery management plans) is to resample from the MLEs for the  $\theta$ 's. Just 6 years of estimates were available, however, so the resulting sample space would be relatively small. Another advantage of the hierarchic model is the structure it provides for incorporating explanatory variables, such as ocean conditions, for modeling the SSM parameters. Both points will be further discussed later in the paper.

#### 5.5 POSTERIOR DISTRIBUTIONS FOR $\Theta_k$

The MCMC output, once convergence is reached, is a sample from the distribution of the  $\theta$ 's conditional on both the data ( $C^K$ ) and the hyperdistribution as specified by the maximum likelihood estimate of  $\Omega$ . Thus, uncertainty about the value of  $\Theta$  for a given cohort, say  $\Theta_k$ , can be communicated in a richer manner than simply reporting point estimates and standard errors and perhaps more accurately than relying on asymptotic normality. The simulated sampling distributions for each  $\theta$  (Figure 2) did in fact indicate that the distributions were fairly symmetric, with the exception of the parameter  $\theta_{m2}$ .

The averages of the simulated values are empirical Bayes estimates. Table 2 compares those estimates with the MLEs based on individual cohort data. The standard deviations of the simulated  $\theta$ 's are also contrasted with the usual asymptotic standard errors for MLEs.

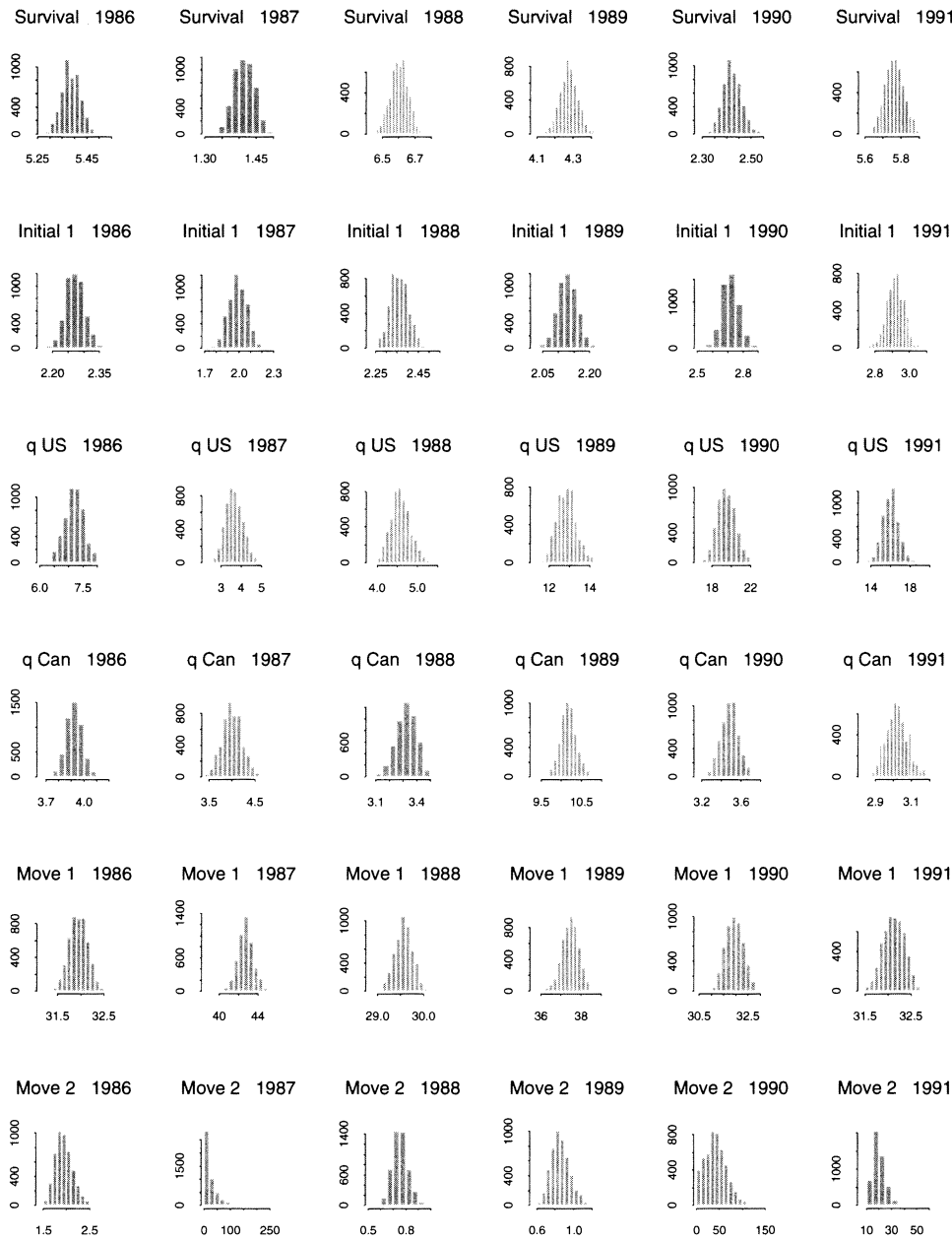


Figure 2. Histograms of Simulated  $\Theta \mid \hat{\Omega}, C^K$  for Six Cohorts of Coho Salmon Based on MCMC of  $N = 5,000$  and  $B = 100$ .

Recall that the former measure should be an underestimate due to uncertainty in  $\hat{\Omega}$ . The standard deviations of the posterior distribution of  $\Theta \mid \hat{\Omega}, C_k$  were calculated from the last  $N - B$  simulated MCMC  $\theta$ 's, with every other observation deleted (thinned) to minimize the potential dependencies. On the basis of output from Raftery and Lewis's *gibbsit* (1996)

Table 2. MLEs and Empirical Bayes (EB) Estimates of SSM Parameters for Six Cohorts of Coho Salmon. The estimated standard errors for the MLEs and the standard deviations of the posterior distributions of  $\Theta \mid \hat{\Omega}, C_k$  are shown below in parentheses. NA denotes missing values.

Year	$\theta_s$		$\theta_{i1}$		$\theta_{qUS}$		$\theta_{qCanada}$		$\theta_{m1}$		$\theta_{m2}$	
	MLE	EB	MLE	EB	MLE	EB	MLE	EB	MLE	EB	MLE	EB
1986	5.39 (0.04)	5.39 (0.04)	2.28 (0.03)	2.26 (0.03)	7.20 (0.27)	7.20 (0.33)	3.92 (0.06)	3.93 (0.07)	31.92 (0.20)	31.95 (0.20)	1.96 (0.19)	1.95 (0.20)
1987	1.42 (0.03)	1.42 (0.03)	1.97 (0.08)	1.99 (0.08)	3.52 (0.41)	3.67 (0.43)	4.00 (0.23)	4.01 (0.22)	42.65 (0.26)	42.69 (0.83)	252.6 (NA)	25.83 (29.62)
1988	6.61 (0.05)	6.61 (0.05)	2.36 (0.04)	2.36 (0.04)	4.58 (0.24)	4.58 (0.25)	3.32 (0.07)	3.33 (0.07)	29.56 (0.20)	29.57 (0.19)	0.76 (0.07)	0.76 (0.06)
1989	4.28 (0.05)	4.28 (0.05)	2.13 (0.03)	2.13 (0.03)	12.84 (0.47)	12.84 (0.48)	10.21 (0.21)	10.18 (0.20)	34.47 (0.41)	37.45 (0.42)	0.84 (0.10)	0.86 (0.10)
1990	2.43 (0.04)	2.43 (0.040)	2.73 (0.06)	2.72 (0.06)	19.59 (0.99)	19.51 (0.95)	3.48 (0.09)	3.49 (0.09)	31.98 (0.31)	31.94 (0.36)	54.16 (NA)	42.46 (23.53)
1991	5.74 (0.05)	5.77 (0.05)	2.94 (0.05)	2.92 (0.05)	16.77 (0.78)	16.01 (0.82)	2.96 (0.06)	3.02 (0.06)	31.07 (0.23)	32.13 (0.23)	22.95 (4.73)	19.91 (5.08)

routine, to create independent samples, thinning every second realization was deemed necessary for only three parameters,  $\theta_{m2}$  for 1987, 1990, and 1991. This estimate of the standard deviation is simple but wasteful in that more sophisticated estimators utilizing every observation could be used (Geyer 1992). To estimate the standard errors of the MLEs, the inverse of the Hessian matrix for the negative log likelihood was numerically approximated.

Differences between the individual cohort-based MLEs and the empirical Bayes estimates were extremely minor in almost all cases. The estimates of  $\theta_{m2}$  were the least precisely estimated from both approaches, as was most evident with the 1987 and 1990 data. The likelihood was relatively flat in the dimension of  $\theta_{m2}$ , widely different values yielded approximately the same likelihood value, and the numerical routine for calculating the Hessian had difficulty calculating derivatives in the region of the maximum likelihood estimate; hence, the missing values. The high degree of similarity between the maximum likelihood estimates and the empirical Bayes estimates suggests that the data for a given cohort is informative enough that any “borrowing of strength” from other cohorts is minimal, i.e., that, given  $C_k$ ,  $\hat{\Omega}$  has little effect on the estimate of  $\Theta_k$ .

5.6 EVALUATING PRESEASON MANAGEMENT PLANS

The merit of the hierarchic modeling is most evident when one aims to predict the abundance and catches for a future cohort under a proposed management plan. With only 6 years of information, simply sampling from the empirical distribution of MLEs for the  $\Theta$  could lead to a quite discontinuous array of abundance and catch patterns. Sampling from the continuous hyperdistributions will smooth the distribution of predicted values.

As a demonstration, three different preseason management plans are compared. For simplicity, a management plan is a matrix of proposed time and area fishing effort levels, e.g., number of commercial troll boat days. This is not entirely realistic because salmon management plans are more complex, involving things like season openings and closings

that do not translate directly into fishing effort. But the complexities of plan evaluation and uncertainties regarding the consequences of implementing a plan can still be conveyed using the effort matrix.

For a given cohort, natural variation arises at each level of the hierarchic model due to random variation in the

- (a) SSM parameters:  $\Theta$  generated from  $f(\Theta \mid \Omega)$ ,
- (b) state process:  $\mathbf{n}_t \mid \mathbf{n}_{t-1}, \Theta$ ,
- (c) observation process:  $\mathbf{c}_t \mid \mathbf{n}_t, \Theta$ .

It is more accurate to include uncertainty in the values of the hyperparameters, too. For example, sample from a multivariate normal distribution with mean vector  $\hat{\Omega}$  and the covariance matrix estimated by jackknifing. Other sources of uncertainty are the variation in the number of fish released and the actual fishing effort.

To evaluate a management plan and demonstrate the uncertainty of the outcomes, some of the above uncertainties were simulated. Given that a particular management plan has been translated into an effort matrix, the steps were to

- (1) generate the SSM parameters from  $f(\Theta \mid \hat{\Omega}^*)$ ;
- (2) simulate the number of fish released using a normal distribution centered at a target value;
- (3) randomly perturb the proposed fishing effort using a Poisson distribution per time and area cell with mean value equal to the proposed effort level;
- (4) simulate the SSM processes (abundance and catch) using multivariate normal distributions, with expectations given by Equations (3.1) and (3.2) and the covariance matrices described earlier.

The output from each simulation was a catch matrix, or a matrix of tag recoveries by time and area. To evaluate the ability of a particular management proposal to achieve desired objectives, the relevant differences between the observed matrix and what a desired matrix would look like were input to a loss function. In the following example, loss was defined as a function of the difference between predicted escapement and a target escapement, with a more severe penalty for underescapement than for overescapement. Given multiple stakeholders in the fishing process, multiple loss functions may need to be defined that are either evaluated separately or combined using an agreed-upon weighting.

Three management plans (effort matrices) were compared using the  $\hat{\Omega}$  for the Washington coho salmon stock. Plan 1 assumed 1986 effort levels, plan 2 decreased the 1986 Canadian effort levels to one third of their original values and tripled the 1986 U.S. effort levels, and plan 3 was a reversal of plan 2, i.e., Canadian effort levels were tripled and U.S. effort levels were reduced by two thirds.

The targeted number of tagged fish released was 70,000 with a coefficient of variation of 0.5%, reflecting uncertainty in the actual number of tagged fish leaving the hatchery. The loss function was defined in terms of deviation from a target escapement of 2,000 fish, with a quadratic penalty for underescapement and a linear penalty for overescapement. For plan

$a_i$  and simulation  $j$ ,

$$l(Esc_{i,j}, a_i) = \frac{(Esc_{i,j} - 2,000)^2}{2,000}, \quad Esc_{i,j} \leq 2,000$$

$$l(Esc_{i,j}, a_i) = \frac{Esc_{i,j} - 2,000}{2,000}, \quad Esc_{i,j} > 2,000.$$

Each plan was simulated  $N = 1,000$  times (simulations for each plan beginning with the same random number seed). The risk,  $r$ , for each plan was estimated by averaging the  $N$  evaluations of the loss function,

$$\hat{r}(a_i) = \frac{1}{N} \sum_{j=1}^N l(Esc_{i,j}, a_i).$$

Figure 3 shows the simulated escapements under each plan as well as the associated losses. The estimated risks were 374, 271, and 528 for plans 1, 2, and 3, respectively, and the corresponding standard deviation of losses were 151, 18, and 409, indicating that plan 2 has a lower expected loss and has much less variability in its losses. However, given that the median escapements for all three plans are less than the target escapement, a fourth management plan would likely be proposed.

## 6. DISCUSSION

### 6.1 ADVANTAGES OF THE HIERARCHIC STRUCTURE AND AREAS NEEDING IMPROVEMENT

As has been argued elsewhere (Schnute 1994; Newman 1998), the state-space model structure is often appropriate for sequential fisheries data. When the SSM is used for annual harvest management planning, superimposing the additional structure of a hyperdistribution on the SSM's survival and migration parameters is a reasonable method for incorporating year-to-year variation in cohort abundance and behavior. For comparison of different preseason fishery management proposals, a more realistic assessment of the uncertainty of total harvest, harvest rates, and escapement is possible. The framework is also conducive to formalizing the decision making process by specifying loss functions and then estimating the risk, i.e., expected loss, under different management scenarios.

Another potential advantage of the hierarchic structure is that empirical Bayes estimates and related posterior measures provide a means of cohorts borrowing strength from other cohorts. For example, for a cohort with a relatively small number of tag recoveries due to a reduced harvest year, the information about the SSM parameters is diminished relative to other cohorts in other years. If analyzed in isolation, the SSM parameter estimates for that cohort may be quite imprecise. When combined with information from other years, the empirical Bayes estimates may be more stable. Furthermore, the entire sampling distribution for each parameter can be estimated without making assumptions of asymptotic normality. The gain in terms of better understanding of the  $\theta$ 's is admittedly slight with the particular data set studied here (e.g., the asymmetry in the sampling distribution for  $\theta_{m2}$  and the

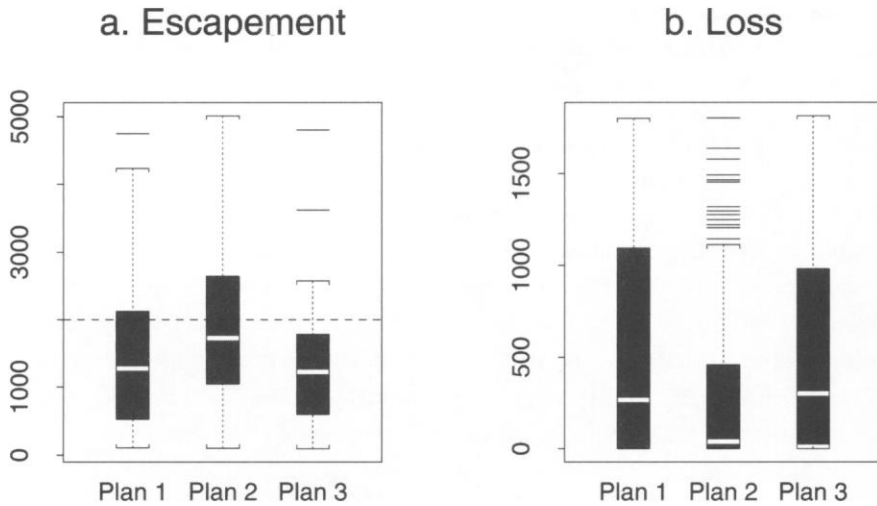


Figure 3. Escapement and Losses for Three Management Plans Based on 1,000 Simulations. Loss function is quadratic for underescapement and linear for overescapement. The target escapement was 2,000 fish.

ability to estimate the spread of the distribution when the numerically calculated Hessian failed). Differences between empirical Bayes estimates and maximum likelihood estimates based on each cohort alone were generally minor, apparently due to the large number of tag recoveries available.

One shortcoming of the formulation was the fixing of 6 of the 12 hyperparameters. Given more years of data, it should be possible to estimate more hyperparameters.

A more serious problem may be the assumption of independence between the SSM parameters. There is some degree of dependence between some of these parameters within a given cohort. For example, if the ocean conditions are poor at the time the juvenile salmon enter saltwater, initial survival and initial distribution are likely affected simultaneously. The two harvest-related parameters could be related as well; changes in ocean conditions or abundance of hatchery stocks could cause similar changes in the way both U.S. and Canadian fisheries operate and in their harvesting efficiency. One solution is to formulate a multivariate distribution with dependencies between some of the  $\theta$ 's, such as a bivariate log-normal distribution for the two harvest-related parameters. Determining a good covariance structure may be quite difficult, however.

A scientifically more satisfying alternative is to model  $\Omega$  as a function of covariates thought to influence the  $\theta$ 's. For example, suppose sea surface temperature,  $SST$ , influences initial survival, initial spatial distribution, and migration. For cohort  $k$ , the following models could be assumed:

$$\begin{aligned}\omega_{s,k} &= \exp [\beta_{0,s} + \beta_{1,s}SST_k] \\ \omega_{i1,k} &= \exp [\beta_{0,i1} + \beta_{1,i1}SST_k] \\ \omega_{m1,k} &= \exp [\beta_{0,m1} + \beta_{1,m1}SST_k] .\end{aligned}$$

If the ocean covariates could be measured before harvesting begins, this formulation would have the additional advantage that preseason assessment could be improved by narrowing the distribution for  $\Theta$ . The likelihood (4.1) would be rewritten as

$$f(C^K | X^K, \beta, \Omega) = \int f(C^K | \Theta^K) h(\Theta^K | X^K, \beta, \Omega) d\Theta^K,$$

where  $\beta = \{\beta_{0,s}, \dots, \beta_{1,m1}\}$ ,  $X^K$  are the cohort-specific covariates, and  $\Omega$  is now the remaining subset of hyperparameters not modeled as a function of  $X^K$ .

Another issue is treatment of the error in  $\hat{\Omega}$ , which has been ignored in both the calculation of the empirical Bayes estimates of  $\Theta$  and in the simulation of management plans. To account for this uncertainty, one might sample from the joint distribution  $f(\Theta_k, \hat{\Omega} | C^K) = h(\hat{\Omega} | C^K) f(\Theta_k | C^K, \hat{\Omega})$ . An approach mentioned previously is to assume multivariate normality for  $\hat{\Omega}$  using the point estimates as the mean vector and the jackknife covariance matrix (4.7), although a potential practical problem would be negative values for some  $\omega$ . Then a sample is generated from  $f(\Theta_k | C^K, \hat{\Omega}^*)$ , which can be done by MCMC, an extremely computer-intensive solution. As an alternative, the sample from  $f(\Theta_k | C^K, \hat{\Omega})$  could perhaps be reused by means of rejection sampling (Ripley 1987); i.e., for  $i = B + 1$  to  $N$  and for each cohort  $k$ , keep  $\Theta_i^{k,*}$  (from the MCMC sample) with probability

$$\frac{f(C_k, \Theta_k^{*,i} | \hat{\Omega}^*)}{f(C_k, \Theta_k^{*,i} | \hat{\Omega})} \frac{1}{M},$$

where  $M$  is an upper bound on the ratio of the two probability densities.

An alternative to the above is a fully Bayesian approach, Bayes empirical Bayes. One defines a (multivariate) prior distribution for  $\Omega$  (or, if using covariates, for the  $\beta$ 's),  $\pi(\Omega | \eta)$ , where  $\eta$  would be fixed and known. The marginal posterior distributions for  $\Theta$  and  $\Omega$  would be analytically intractable but could, in principle, be sampled from using MCMC. In other words,

$$f(\Theta^K, \Omega | C^K) \propto \prod_{k=1}^K f(C_k | \Theta_k) h(\Theta_k | \Omega) \pi(\Omega | \eta).$$

To evaluate a proposed management plan, one could sample from an enlarged joint probability distribution with  $C_{\text{new}}$  and  $\Theta_{\text{new}}$  being generated,

$$f(C_{\text{new}}, \Theta_{\text{new}}, \Theta^K, \Omega | C^K) \propto f(C_{\text{new}} | \Theta_{\text{new}}) h(\Theta_{\text{new}} | \Omega) \prod_{k=1}^K f(C_k | \Theta_k) h(\Theta_k | \Omega) \pi(\Omega | \eta).$$

An example of Bayes empirical Bayes applied to fisheries data is given by Liermann and Hilborn (1997). (See also Dominici, Parmigiani, Reckhow, and Wolpert (1997) for a Bayes empirical Bayes example using MCMC.)

The state-space model itself could be improved upon and extended in several ways (Newman 1998), including (1) multiple, competing gear types (e.g., recreational and commercial), (2) nonnormal and nonlinear state and observation distributions (likely



requiring another layer of MCMC simulations for parameter estimation), and (3) more realistic, but complex, spatial frameworks (incorporating, e.g., the interior marine waters of Puget Sound).

## 6.2 COMPUTATIONAL ISSUES

Two principles underlying this work, (1) using a state-space model to characterize an observable time series of salmon counts or observations and (2) using a hierarchic structure to model differences between time series for different cohorts or stocks, are generalizable to other species of fish and some other animals. The formulation in any particular case will, however, be idiosyncratic, and it may be helpful to discuss some of the computational issues involved in implementing a new model.

The computational tasks can be put into two categories: calculating the objective function [e.g., Eq. (4.6)] and then finding the maximum likelihood estimates for the hyperparameters.

Calculating the objective function was done using MCMC methods and importance sampling. Using the same family of hyperdistributions for the importance distribution, where the parameters  $\Psi$  of the importance distribution were chosen on the basis of the individual MLEs for  $\theta$  from each cohort, worked well. The number of iterations of exchanging the most recent estimate of  $\Omega$  for  $\Psi$ , as recommended by Geyer (1996), was few, two or three at most. The choice of the proposal distributions for the MCMC was probably less important than the choice of the tuning parameters. Ease of simulation may be the primary criterion for selection of the proposal distribution. More critical to convergence rates and mixing were the tuning parameters. By approximately centering the proposal distribution on the previously generated values and manipulating the tuning parameters in a trial and error manner, the acceptance rate could be manipulated fairly well. For 28 of the 36 parameters, the acceptance rates were between 20 and 50%, considered a good range by some (Roberts 1996). The recommended burn-in period, based on Raftery and Lewis's *gibbsit* (1996) program, was quite short,  $B$  less than 100.

The time required to generate 5,000 simulations from the Markov chain was around 50 hours run on an HP 9000/831 Unix workstation. With 36 SSM parameters to generate and an average acceptance rate of 36.33%, nearly half a million random variables were simulated.

Given a means of calculating the objective function, the maximum likelihood estimates were found using a nonlinear optimization routine, NPSOL (Gill et al. 1986). NPSOL will perform numerical calculations of the gradient, which could not be analytically determined. Alternative optimization programs that use automatic differentiation would perhaps perform faster, although NPSOL only required a few minutes to find the MLEs for the hyperparameters, and there were no convergence problems. On the other hand, finding MLEs for the SSM parameters,  $\theta$ , on a per cohort basis took between 5 and 10 minutes. There were sometimes convergence problems with  $\theta_{m2}$ , the problematic movement parameter, but not for the other parameters. The use of scaling constants for most of the SSM parameters

to make the relative magnitude of all parameter estimates similar is advised by Gill et al. (1981) and that may have minimized convergence problems.

## ACKNOWLEDGMENTS

This work was financially supported by the National Marine Fisheries Service and the University of Washington's Columbia Basin Research Center. Helpful suggestions have been provided by Din Chen, Charlie Geyer, Allan Hicks, Robert Kope, Gary Morishima, Jim Norris, two anonymous referees, and an associate editor. Cara Campbell and Jim Scott oversaw the difficult process of retrieving the appropriate tag recovery data (from the Pacific States Fisheries Management Commission) and fishing effort data (from California Department of Fish and Game, Oregon Department of Fish and Wildlife, Washington Department of Fish and Wildlife, Canadian Department of Fisheries and Oceans, and Alaska Department of Fish and Game). Allan Hicks also provided computational assistance.

[Received October 1998. Accepted April 2000.]

## REFERENCES

- Bilton, H. T., Alderdice, D. F., and Schnute, J. T. (1982), "Influence of Time and Size at Release of Juvenile Coho Salmon (*Oncorhynchus kisutch*) on Returns at Maturity," *Canadian Journal of Fisheries and Aquatic Science*, 39, 426–447.
- Carlin, B. P., and Lewis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, New York: Chapman and Hall.
- Dominici, F., Parmigiani, G., Reckhow, K. H., and Wolpert, R. L. (1997), "Combining Information From Related Regressions," *Journal of Agriculture, Biological, and Environmental Statistics*, 2, 313–332.
- Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.
- Geyer, C. J. (1992), "Practical Markov Chain Monte Carlo (With Discussion)," *Statistical Science*, 7, 657–699.
- (1996), "Estimation and Optimization of Functions," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, New York: Chapman and Hall, pp. 241–258.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), "Introducing Markov Chain Monte Carlo," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, New York: Chapman and Hall, pp. 1–19.
- Gill, P. E., Murray, W., Saunders, M. A., and Wright, M. H. (1986), "User's Guide for NPSOL (Version 4.0): A Fortran Package for Nonlinear Programming," Technical Report SOL 86-2, Systems Optimization Laboratory, Department of Operations Research, Stanford, CA.
- Gill, P. E., Murray, W., and Wright, M. H. (1981), *Practical Optimization*, New York: Academic.
- Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.
- Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Transactions ASME Journal of Basic Engineering*, 82, 35–45.
- Liermann, M., and Hilborn, R. (1997), "Depensation in Fish Stocks: A Hierarchic Bayesian Meta-Analysis," *Canadian Journal of Fisheries and Aquatic Sciences*, 54, 1976–1984.
- Mendelssohn, R. (1988), "Some Problems in Estimating Population Sizes From Catch-at-Age Data," *Fishery Bulletin*, 86, 617–630.
- Newman, K. B. (1998), "State-Space Modeling of Animal Movement and Mortality With Application to Salmon," *Biometrics*, 54, 274–297.
- Raftery, A. E., and Lewis, S. M. (1996), "Implementing MCMC," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, New York: Chapman and Hall, pp. 115–130.
- Reed, W. J., and Simonis, C. M. (1996), "Analyzing Catch-Effort Data by Means of the Kalman Filter," *Canadian Journal of Fisheries and Aquatic Sciences*, 53, 2157–2166.

- Ricker, W. E. (1975), "Computation and Interpretation of Biological Statistics of Fish Populations," Bulletin 191, Fisheries Research Board of Canada, Department of Fisheries and Oceans, Ottawa.
- Ripley, B. D. (1987), *Stochastic Simulation*, New York: Wiley.
- Roberts, G. O. (1996), "Markov Chain Concepts Related to Sampling Algorithms," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, New York: Chapman and Hall, pp. 45–57.
- Sandercock, F. K. (1991), "Life History of Coho Salmon," in *Pacific Salmon Life Histories*, eds. C. Groot and L. Margolis, Vancouver: UBC Press, pp. 396–445.
- Schnute, J. (1994), "A General Framework for Developing Sequential Fisheries Models," *Canadian Journal of Fisheries and Aquatic Sciences*, 51, 1676–1688.
- Shumway, R. H. (1988), *Applied Statistical Time Series Analysis*, New Jersey: Prentice Hall.
- Speed, T. (1993), "Modelling and Managing a Salmon Population," in *Statistics for the Environment*, eds. V. Barnett and K. F. Turkman, New York: Wiley, pp. 267–292.
- Sullivan, P. (1992), "A Kalman Filter Approach to Catch-at-Length Analysis," *Biometrics*, 48, 237–257.
- Tierney, L. (1996), "Introduction to General State-Space Markov Chain Theory," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, New York: Chapman and Hall, pp. 59–74.
- West, M., and Harrison, J. (1990), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag.