

A hierarchical Bayesian model for estimating historical salmon escapement and escapement timing

Zhenming Su, Milo D. Adkison, and Benjamin W. Van Alen

Abstract: In this paper, we present an improved methodology for estimating salmon escapements from stream count data. The new method uses a hierarchical Bayesian model that improves estimates in years when data are sparse by “borrowing strength” from counts in other years. We present a model of escapement and of count data, a hierarchical Bayesian statistical framework, a Gibbs sampling approach for evaluation of the posterior distributions of the quantities of interest, and criteria for determining when the model and inference are adequate. We then apply the hierarchical Bayesian model to estimating historical escapement and escapement timing for pink salmon (*Oncorhynchus gorbuscha*) returns to Kadashan Creek in Southeast Alaska.

Résumé : On trouvera ici une méthodologie améliorée pour estimer les échappées de saumon à partir des données de dénombrement dans un cours d'eau. La nouvelle méthode utilise un modèle hiérarchique bayésien (HBM) qui améliore les estimations pour les années où les données sont rares en “empruntant de la puissance” aux dénombrements des autres années. Nous présentons un modèle pour les données d'échappées et de dénombrement, un cadre statistique hiérarchique bayésien, une stratégie d'échantillonnage de type Gibbs pour l'évaluation des distributions a posteriori des valeurs recherchées et des critères pour déterminer quand le modèle et les inférences sont adéquats. Nous appliquons le modèle HBM à l'estimation des échappées des années antérieures et du moment des retours des Saumons roses à Kadashan Creek dans le sud-est de l'Alaska.

[Traduit par la Rédaction]

Introduction

Escapement data are essential for salmon population conservation and management (Quinn and Deriso 1999). In North America, most escapement data are obtained by aerial or foot escapement surveys (Cousens et al. 1982) in which fish observed in a stream are enumerated at several occasions to obtain a series of counts. Each count is just an estimate of the actual number of spawners in a stream at the time of the survey, not the total escapement (some fish will have died before counting and some will not yet have entered). Thus, some method must be used to convert these counts to escapement. Usually, escapement is estimated with the “area-under-the-curve” method (English et al. 1992; Hilborn et al. 1999), a somewhat ad hoc calculation that uses the counts in conjunction with an estimate of the average longevity of spawners (stream life) (Perrin and Irvine 1990; Fukushima and Smoker 1997). More recently, statistical models have been developed to estimate escapement from escapement counts. Quinn and Gates (1997) developed a bi-

ologically detailed stream escapement model to estimate salmon escapement dynamics. Hilborn et al. (1999) provided a maximum likelihood method to estimate stream escapement using a simpler model. In addition to providing an estimate of the total escapement, these statistical methods can estimate escapement timing as well. Escapement timing information is an important prerequisite for run reconstruction techniques (Starr and Hilborn 1988).

For these modeling approaches, separately estimating the escapement in each year may be adequate for years with informative data (Quinn and Gates 1997; Hilborn et al. 1999). However, in some years, counts may be sparse and insufficient to provide reliable estimates. Even with a large number of survey counts, estimation can easily fail due to a lack of data for the last half of the run (Hilborn et al. 1999). Area-under-the-curve estimates are even more affected by such holes in the data, as part of the curve is missing. Such deficiencies are common in historical data; weather, human resource shortages, or budget difficulties may limit counts, and managers often discontinue counts once escapement goals are met.

Hilborn et al. (1999) proposed using prior information on run timing to constrain the escapement model when data for a particular year were uninformative. Such information is most logically derived from the escapement timing of the stock in other years. The timing of salmon returning to a particular stream is remarkably consistent from year to year (Heard 1991), so historical data may contain strong prior information on escapement timing in the current year. One could specify prior distributions for escapement timing parameters from this historical information and perform a

Received January 4, 2001. Accepted May 24, 2001. Published on the NRC Research Press Web site at <http://cjfas.nrc.ca> on July 18, 2001.
J16155

Z. Su¹ and M.D. Adkison. Juneau Center, School of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Juneau, AK 99801, U.S.A.

B.W. Van Alen. Alaska Department of Fish and Game, Douglas, AK 99824, U.S.A.

¹Corresponding author (e-mail: ftzs@uaf.edu).

Bayesian or even a maximum likelihood analysis separately for each year of data (Hilborn et al. 1999). A more formal statistical methodology, hierarchical modeling (Lindley and Smith 1972; Gelman et al. 1995), exists that would allow estimation for all years simultaneously, “borrowing strength” from years with informative data to improve estimates for years with uninformative data.

Hierarchical modeling has many applications in biology, medicine, economics, etc. (Davidian and Giltinan 1995; Gelman et al. 1995). It is also naturally applied in meta-analysis where studies have to be pooled together (Mosteller and Chalmers 1992; for a fisheries example, see Liermann and Hilborn 1997). Hierarchical models assume dependence among parameters, with parameters in lower levels regarded as samples drawn from higher-level population distributions.

Compared with nonhierarchical approaches, one advantage of hierarchical models is that they make full use of all of the data available and can estimate parameters at all levels simultaneously. Another advantage of hierarchical modeling is that it saves degrees of freedom by structuring some dependence among the parameters, reducing the effective number of parameters and thereby avoiding problems of overfitting that are often suffered by nonhierarchical models (Gelman et al. 1995). Bayesian (Gelman et al. 1995; Carlin and Louis 1996) and traditional maximum likelihood procedures (Davidian and Giltinan 1995) can both be applied for the fitting of hierarchical models, although the Bayesian approach provides more accurate assessments for uncertainty in parameters (Browne and Draper 2000).

In this paper, we develop a statistical escapement estimation methodology using hierarchical Bayesian modeling (HBM). The hierarchical component is the assumption that the timing of escapement in each year is a sample drawn from a stock-specific distribution of possible timings. This assumption enables us to use all of the historical survey data for a stream to improve escapement estimates for years with uninformative data as well as to make inferences about typical escapement characteristics of the stream.

We also incorporate an additional improvement over past escapement estimation practices. Traditional methods, such as the area-under-the-curve method and Hilborn et al.’s (1999) model, used an average stream life to convert counts to escapement. However, it is well known that longevity of salmon on the spawning ground declines as the season progresses (Dangel and Jones 1988; Perrin and Irvine 1990; Fukushima and Smoker 1997), so using an average stream life may bias the estimation. Our spawning abundance model incorporates a stream life that declines as the season progresses.

The article is organized as follows. First, we describe a salmon spawning abundance model. We then develop an HBM that combines counts for all of the years together. A hybrid Metropolis–Gibbs algorithm is used for fitting the HBM. We derive the full conditional distributions needed for this approach. We demonstrate a whole set of Bayesian methodologies for high-dimensional Bayesian model fitting through Markov chain Monte Carlo (MCMC) techniques (Gilks et al. 1996), detailed convergence diagnostics of MCMC (Appendix B), and Bayesian model checking and comparison (Appendix C). Finally, we apply the HBM ap-

proach to escapement estimation for a pink salmon (*Oncorhynchus gorbuscha*) stock in Southeast Alaska.

Models and methods

A salmon spawning abundance model

There are two processes determining the dynamics of salmon spawning abundance in a stream during the spawning season: entry and mortality. Let η_t denote the number of spawners at the beginning of day t , x_t the number of spawners entering the stream during day t , and d_t the number of fish that die in the same period; then the spawning abundance is

$$(1) \quad \eta_{t+1} = \eta_t + x_t - d_t = \sum_{k=1}^t x_k - \sum_{k=1}^t d_k$$

with an initial condition $\eta_{T_1} = 0$ at the starting day T_1 of the stream entry.

Weir studies show that salmon escapement timing can be asymmetric, with pulses of stream entry. With infrequent stream counts, it is usually intractable to estimate these irregular features of stream entry. To some extent, these irregularities in entry are smoothed in taking stream counts by the overlap of fish that arrived on multiple days. In our model, the daily entry of salmon into the stream is approximated by a normal curve:

$$(2) \quad x_t = E e^{-(t-M)^2/(2S^2)} / \Psi$$

where E denotes the total escapement, M denotes the mean date of arrival, and S denotes the standard deviation of the dates of arrival (i.e., the duration of the run). The value

$$\Psi = \sum_{t=T_1}^{T_2} e^{-(t-M)^2/(2S^2)}$$

is a normalizing constant ensuring that

$$\sum_{t=T_1}^{T_2} (e^{-(t-M)^2/(2S^2)} / \Psi) = 1$$

so that daily entry takes certain percentage of the total escapement. The value T_2 represents the last day of the stream entry.

We add one level of biological realism not usually incorporated in escapement models. Let $l(t)$ be the stream life of the fish entering in day t . It is known that $l(t)$ is a declining function of date within the spawning season (Dangel and Jones 1988; Perrin and Irvine 1990; Fukushima and Smoker 1997). Here, we use an exponential equation to express the stream life by date of entry:

$$(3) \quad l(t) = \phi_1 e^{-\phi_2(t-M)}$$

where ϕ_1 and ϕ_2 are two shape parameters estimated separately from independent stream life surveys and M is the mean date of arrival. The uncertainty in these two parameters is briefly considered in a sensitive analysis (see An application, Stream life estimates section below).

In our model, we assume that fish entering the stream on day ξ die $l(\xi)$ days later:

$$(4.1) \quad \delta_{\xi+l(\xi)} = x_\xi = E e^{-(\xi-M)^2/(2S^2)} / \Psi$$

where $\delta_{\xi+l(\xi)}$ denotes the number of those fish that enter the stream in day ξ and die afterward on day $\xi + l(\xi)$. Due to the declining nature of our stream life function, it is possible that on some day, fish from several different entry days might die (this is not the case for a constant stream life where each entry day is mapped to a unique

mortality day). We express the actual number of fish that died on day t as the sum over stream entry days:

$$(4.2) \quad d_t = \sum_{\{\xi\}} \delta_{\xi+l(\xi)} \quad \text{with } \xi \in \{\xi: \xi + l(\xi) = t\}$$

Thus, the spawning abundance can be expressed as

$$(5) \quad \eta_t = \sum_{k=1}^t x_k - \sum_{k=1}^t d_k \\ = \sum_{k=1}^t x_k - \sum_{k=1}^t \sum_{\{\xi\}} \delta_{\xi+l(\xi)} \\ = E \left(\sum_{k=1}^t e^{-(k-M)^2/(2S^2)} - \sum_{k=1}^t \sum_{\{\xi\}} e^{-(\xi-M)^2/(2S^2)} \right) / \Psi \\ = E \Phi_t$$

where

$$\Phi_t = \left(\sum_{k=1}^t e^{-(k-M)^2/(2S^2)} - \sum_{k=1}^t \sum_{\{\xi\}} e^{-(\xi-M)^2/(2S^2)} \right) / \Psi$$

the proportion of the total escapement alive in the stream on day t .

Let c_t denote the observed stream count on day t . We assume a normally distributed measurement error in the counts. Thus, we have

$$(6) \quad c_t = \eta_t + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2)$. Whenever necessary, the counts are transformed to meet the previous normal model assumptions (see A hierarchical methodology, The hierarchical model section below).

Separate estimation for each year

Given particular values of M , S , and E , the model equations above can be used to predict the number of fish in the stream on any date (η_t). Where we employed separate estimation below, maximum likelihood parameter estimates (MLEs) for a particular year were estimated using a nonlinear search algorithm that selected candidate values of M , S , and E to minimize SSQ, where SSQ was the following:

$$(7) \quad \text{SSQ} = \sum_{\{t\}} (c_t - \eta_t(M, S, E))^2$$

The estimate of σ^2 was calculated as $\text{SSQ}/(\text{no. of counts} - \text{no. of parameters})$.

Confidence intervals for each parameter were obtained by a bootstrap approach. In each bootstrap trial, residuals from the MLE fit were randomly sampled with replacement and added to the predicted counts to create a pseudo-data set. MLEs for these pseudo-data were then calculated. This bootstrap was repeated 3000 times (the percentiles became stable after about 500 iterations), and then the resulting estimates were ordered by size. Finally, the 2.5th and 97.5th percentiles were taken as the lower and upper bounds of the 95% confidence interval.

A hierarchical Bayesian methodology

The hierarchical model

Suppose that we have Y years of survey data for a stream. For year y , there are n_y counts. Let $c_{y,t}$ denote the observed count on survey day t for year y , M_y , S_y , and E_y the three unknown escapement parameters for year y , and σ^2 the data error variance. As in eq. 6, we have a model for the (possibly transformed) $c_{y,t}$ as

$$(8) \quad c_{y,t} | \theta_y, \sigma^2 \stackrel{\text{indep}}{\sim} N(\eta_{y,t}, \sigma^2), y = 1, \dots, Y$$

where $\theta_y = (M_y, S_y, E_y)$. In this model, we allow for the possibility of a different timing and escapement in each year but assume a homogeneous measurement error for all years.

The hierarchical component of the model is the assumption that the timing of escapement is similar from year to year (Heard 1991). We express both the similarity and heterogeneity of escapement timing by assuming that the M_y and $\ln(S_y)$ are drawn from the following distributions:

$$(9) \quad M_y | \mu_m, \tau_m^2 \stackrel{\text{iid}}{\sim} N(\mu_m, \tau_m^2) \\ \ln(S_y) | \mu_s, \tau_s^2 \stackrel{\text{iid}}{\sim} N(\mu_s, \tau_s^2)$$

where μ_m and μ_s are population mean parameters and τ_m^2 and τ_s^2 are population variance parameters. Preliminary analysis of stream escapement data in our study area leads us to assume that M_y and $\ln(S_y)$ are independent of each other. Because of the high variability in abundance from year to year, we do not impose a hierarchical structure on the E_y s. Equations 8 and (or) 9 form a two-stage hierarchical model. There are in total $3Y + 5$ unknown parameters for this model, which are $\theta = (\{\theta_y\}, \sigma^2)$ (here, θ_y is redefined as the set $(M_y, \ln(S_y), E_y)$ and $y = 1, \dots, Y$) and four hyperparameters $\lambda = (\mu_m, \tau_m^2, \mu_s, \tau_s^2)$.

For the purpose of simplicity, the error in counts is assumed to have the same variance (σ^2) for all years. However, annual escapements often fluctuate by many orders of magnitude. To achieve common measurement error variances, we could scale the observed counts for each year by dividing by the peak count $c_{y,p}$ (labeled "scaling"). This is equivalent to setting $\sigma_y = \sigma c_{y,p}$ but keeps our model in a simple form. Square root transformation (SQRT) (Quinn and Gates 1997) of the counts was also considered, and its performance was compared with scaling (we also explored a logarithm transformation (Hilborn et al. 1999)).

We denote the previous model the M_y - S_y model. Two more parsimonious models can be derived from the M_y - S_y model: (i) the M_y - S model with $\ln(S_y) = \ln(S)$ and parameters $\theta = (\{M_y, E_y\}, \ln(S), \sigma^2)$ and $\lambda = (\mu_m, \tau_m^2, \mu_s, \tau_s^2)$ and (ii) the M - S_y model with $M_y = M$ and parameters $\theta = (M, \{\ln(S_y), E_y\}, \sigma^2)$ and $\lambda = (\mu_s, \tau_s^2)$. We consider both the M_y - S and M - S_y models as alternatives, with scaling and SQRT as two data transformation options, but continue to focus on the more general M_y - S_y model in the narrative.

Prior specification

A full Bayesian analysis requires specifying prior distributions for the hyperparameters $\lambda = (\mu_m, \tau_m^2, \mu_s, \tau_s^2)$ and the data-level parameters $\{E_y\}$ and σ^2 . We assume prior independence for all of these parameters. The prior specifications form the third stage of the HBM.

For μ_m and μ_s , we set their priors as $\mu_m \sim N(a_m, V_m)$ and $\mu_s \sim N(a_s, V_s)$, where a_m , V_m , a_s , and V_s are constants to be specified by the modelers. Choosing large V_m and V_s will make these two prior distributions relatively uninformative but still proper (Spiegelhalter et al. 1997) (a proper prior is a prior density, say, $p(\vartheta)$, that can integrate to 1 (or any positive finite value), i.e., the integral $\int p(\vartheta | y) dy$ is a constant).

The specification of the priors of the population variance parameters (e.g., τ_m^2 and τ_s^2) needs special treatment in a HBM because an automatic use of the standard "uninformative" improper priors (e.g., $p(\sigma^2) \propto 1/\sigma^2$) for these parameters might lead to improper posteriors (Gelman et al. 1995; Hobert and Casella 1996) (for an improper posterior density, say $p(\vartheta | y)$, the integral $\int p(\vartheta | y) dy$ is not finite for all y). The safest way to get around this problem is by specifying proper prior distributions for these parameters. We set the priors for τ_m^2 and τ_s^2 as inverse gamma distributions (IG) (see Gelman et al. 1995): $\tau_m^2 \sim \text{IG}(c_m, d_m)$ and $\tau_s^2 \sim \text{IG}(c_s, d_s)$, respectively. To obtain uninformative priors for these hypervariance pa-

rameters, we set $c = d = \varepsilon$, where ε is a small positive value (Spiegelhalter et al. 1997).

For convenience, uninformative priors are specified for the remaining parameters: $E_y \sim U(0, \infty)$ and $\sigma^2 \sim \text{IG}(0.001, 0.001)$. Similar priors are specified for the M_y - S and M - S_y models.

Combining the priors, population distributions, and likelihood (see Hilborn et al. 1999), we obtain the joint posterior distribution for all the parameters in the M_y - S_y model as

$$(10) \quad p(\theta, \lambda | D) \propto p(\lambda) p(\sigma^2) \prod_{y=1}^Y N(M_y | \mu_m, \tau_m^2) \\ \times \prod_{y=1}^Y N(\ln(S_y) | \mu_s, \tau_s^2) \prod_{y=1}^Y p(E_y) \prod_{t=1}^{n_y} N(c_{y,t} | \eta_{y,t}, \sigma^2)$$

where $p(\lambda) = p(\mu_m) p(\tau_m^2) p(\mu_s) p(\tau_s^2)$ and D denotes observed data.

For the M_y - S_y model, we have

$$(11) \quad p(\theta, \lambda | D) \propto p(\lambda) p(\ln(S)) p(\sigma^2) \\ \times \prod_{y=1}^Y N(M_y | \mu_m, \tau_m^2) \prod_{y=1}^Y p(E_y) \prod_{t=1}^{n_y} N(c_{y,t} | \eta_{y,t}, \sigma^2)$$

where $p(\lambda) = p(\mu_m) p(\tau_m^2)$.

For the M - S_y model, the joint posterior distribution is

$$(12) \quad p(\theta, \lambda | D) \propto p(\lambda) p(M) p(\sigma^2) \\ \times \prod_{y=1}^Y N(\ln(S_y) | \mu_s, \tau_s^2) \prod_{y=1}^Y p(E_y) \prod_{t=1}^{n_y} N(c_{y,t} | \eta_{y,t}, \sigma^2)$$

Where $p(\lambda) = p(\mu_s) p(\tau_s^2)$.

MCMC: an adaptive hybrid Metropolis–Gibbs sampler

We choose a hybrid MCMC algorithm, the Metropolis–Gibbs sampler (Tierney 1994), to evaluate our complex joint posterior distributions. The hierarchical model structure and our specification of appropriate conjugate priors enabled us to derive full conditional distributions with familiar distributional forms (e.g., normals and inverse gammas) for most of the parameters in our HBM. This allowed use of Gibbs sampling (Geman and Geman 1984; Gelfand and Smith 1990), a means of generating sample values from the joint posterior distribution by sequential draws from the full conditional distributions. Nonstandard forms for the distribution of the escapement timing parameters required embedding Metropolis steps within the Gibbs procedure (Tierney 1994; Carlin and Louis 1996).

The Gibbs sampler and the full conditional distributions

The Gibbs sampler generates random draws for all the parameters by sampling successively from the full posterior conditional distribution of each parameter $p(\phi_i | \{\phi_j\}_{j \neq i}, D)$, $i = 1, \dots, p$, where $\{\phi_j\}_{j \neq i}$ is the set of all of the parameters of a model except for ϕ_i . Under very broad conditions (Geman and Geman 1984), the Gibbs sampler produces a Markov chain with the joint posterior distribution as its equilibrium distribution. Consequently, the sampled values of ϕ_i from the Gibbs sampler after convergence can be regarded as random draws obtained from its marginal posterior distribution. Thus, simple summary statistics and kernel density estimates based on these draws can be used to summarize the marginal posterior distribution for each parameter.

The full conditional distribution of a parameter is obtained by extracting the terms in the posterior distribution only involving that parameter (its prior and likelihood) and treating all other terms as

constants. For example, the terms involving μ_m in eq. 10 are its prior term $p(\mu_m) = N(\mu_m | a_m, V_m)$ and its likelihood term.

$$\prod_{y=1}^Y N(M_y | \mu_m, \tau_m^2)$$

The likelihood term, treated as a function of μ_m , can be shown to be $N(\mu_m | \hat{\mu}_m, \tau_m^2/Y)$, where

$$\hat{\mu}_m = \left(\frac{1}{Y} \right) \sum_{y=1}^Y M_y$$

Multiplying the likelihood by the prior, we obtain

$$(13) \quad \mu_m | \{M_y\}, \tau_m^2 \sim N \left[\frac{a_m \tau_m^2 + Y \hat{\mu}_m V_m}{\tau_m^2 + Y V_m}, \frac{V_m \tau_m^2}{\tau_m^2 + Y V_m} \right]$$

Similarly, we derived the rest of the full conditionals for the M_y - S_y , M_y - S and M - S_y models, and these are listed as follows:

$$(14) \quad \tau_m^2 | \{M_y\}, \mu_m \sim \text{IG}(c_m + Y/2, d_m + Y \hat{\tau}_m^2/2)$$

where

$$\hat{\tau}_m^2 = \frac{1}{Y} \sum_{y=1}^Y (M_y - \mu_m)^2$$

$$(15) \quad \mu_s | \{\ln(S_y)\}, \tau_s^2 \sim N \left[\frac{a_s \tau_s^2 + Y \hat{\mu}_s V_s}{\tau_s^2 + Y V_s}, \frac{V_s \tau_s^2}{\tau_s^2 + Y V_s} \right]$$

with

$$\hat{\mu}_s = \frac{1}{Y} \sum_{y=1}^Y \ln(S_y)$$

$$(16) \quad \tau_s^2 | \{\ln(S_y)\}, \mu_s \sim \text{IG}(c_s + Y/2, d_s + Y \hat{\tau}_s^2/2)$$

where

$$\hat{\tau}_s^2 = \frac{1}{Y} \sum_{y=1}^Y (\ln(S_y) - \mu_s)^2$$

$$(17) \quad \sigma^2 | \{M_y, \ln(S_y), E_y\} \sim \text{IG}(c_1 + n/2, d_1 + n \hat{\sigma}^2/2)$$

where

$$\hat{\sigma}_2 = \frac{1}{n} \sum_{y=1}^Y \sum_{t=1}^{n_y} (c_{y,t} - \eta_{y,t})^2$$

and

$$n = \sum_{y=1}^Y n_y$$

$$(18) \quad E_y | M_y, \ln(S_y), \sigma^2 \sim N(\hat{\mu}_{E_y}, V_{E_y})$$

where

$$\hat{\mu}_{E_y} = \sum_{t=1}^{n_y} (\Phi_{y,t} c_{y,t}) / \sum_{t=1}^{n_y} (\Phi_{y,t})^2$$

and

$$V_{E_y} = \sigma^2 \left/ \sum_{t=1}^{n_y} (\Phi_{y,t})^2 \right.$$

$$(19) \quad p(M_y | \mu_m, \tau_m^2, \ln(S_y), E_y, \sigma^2) \\ \propto N(\mu_m, \tau_m^2) \exp \left(-\frac{1}{2\sigma^2} \sum_{t=1}^{n_y} (c_{y,t} - \eta_{y,t})^2 \right)$$

$$(20) \quad p(\ln(S_y) | \mu_s, \tau_s^2, M_y, E_y, \sigma^2) \\ \propto N(\mu_s, \tau_s^2) \exp \left(-\frac{1}{2\sigma^2} \sum_{t=1}^{n_y} (c_{y,t} - \eta_{y,t})^2 \right)$$

For the M_y - S model, eqs. 15, 16, and 20 are replaced by

$$(21) \quad p(\ln(S) | \{M_y, E_y\}, \sigma^2) \\ \propto N(\ln(S) | a_s, V_s) \exp \left(-\frac{1}{2\sigma^2} \sum_{y=1}^Y \sum_{t=1}^{n_y} (c_{y,t} - \eta_{y,t})^2 \right)$$

For the M - S_y model, eqs. 13, 14, and 19 are replaced by

$$(22) \quad p(M | \{\ln(S_y), E_y\}, \sigma^2) \\ \propto N(M | a_m, V_m) \exp \left(-\frac{1}{2\sigma^2} \sum_{y=1}^Y \sum_{t=1}^{n_y} (c_{y,t} - \eta_{y,t})^2 \right)$$

The Metropolis steps

The full conditional distributions for the parameters M_y , $\ln(S_y)$, $\ln(S)$, and M (eqs. 19–22) are not in closed form due to the non-linearity of the spawning abundance model. It is not straightforward to generate random values from these distributions. We apply Metropolis steps to sample from each of these conditionals (Carlin and Louis 1996). For more details of the implementation of Metropolis steps, see Appendix A.

In Appendices B and C, we also provide further details for the implementation of the MCMC convergence diagnostics and model comparison and adequacy checking.

An application

Kadashan Creek pink salmon

We applied the HBM to estimate historical pink salmon escapements for Kadashan Creek in Southeast Alaska. Escapement counts for this stream were obtained from the Alaska Department of Fish and Game Integrated Database (Van Alen 2000). Most of these counts are collected through aerial escapement surveys.

In this paper, we illustrate the results for Kadashan Creek over the period 1974–1998. Escapement counts from this stream vary considerably among years in both frequency and duration (Fig. 1). The scale of the counts also varies by severalfold across years, necessitating transformation of the counts. However, scatter plots of the counts suggest similar run timing among years. Thus, the hierarchical approach seems reasonable for this stock.

Stream life estimates

The relationship between stream life and entry date was estimated outside of the HBM framework. Dangel and Jones (1988) studied stream life of pink salmon spawners in several streams in Southeast Alaska: the Kadashan River in 1986 and Pleasant Bay Creek, Black Bear Creek, and Sashin Creek in 1986 and 1987. We used these data to estimate the decline in stream life with date as follows. First, we centered the stream life values for each stream in

each year around the mean date of the corresponding stream entry (obtained from the weir data of Dangel and Jones 1988). We then combined the data from all streams (Fig. 2) and fitted them to the stream life equation, eq. 3. The estimated parameter values were $\hat{\phi}_1 = 14.16$ and $\hat{\phi}_2 = 0.0284$.

The period when Kadashan Creek was surveyed in 1986 was from July 31 to August 10, much shorter than the period of the stream entry (July 18 to August 28). Nevertheless, we fit eq. 3 to the Kadashan data and obtained $\hat{\phi}_1 = 11.8$ and $\hat{\phi}_2 = 0.0078$ (a small declining rate) for this creek. These parameters imply that stream life in Kadashan Creek was approximately constant. We looked at the effect of assuming a constant stream life (set to 11.8 ($\hat{\phi}_1$)) and compared the results using this assumption (referred to as CnSL) with those obtained assuming a declining stream life.

Details of estimation

We estimated escapement, the date of peak escapement, and the standard deviation in date of escapements for historical Kadashan Creek data using the M_y - S_y , M_y - S , and M - S_y models and separate estimation of each year. A Pascal program running under the Borland Delphi environment was developed to perform all the calculations. We used both the scaling and SQRT transformations of count data from 1974 to 1998. Because of convergence problems, some model comparisons were made using only data from 1977 to 1998 (see Results below).

The data appeared to contain ample information to estimate μ_m , μ_s , the E_y s, and σ^2 (see Results below). We specified uninformative priors for these parameters: $\mu_m \sim N(210, 10^5)$, $\mu_s \sim N(2.5, 10^3)$, $E_y \sim U(0, \infty)$, and $\sigma^2 \sim \text{IG}(0.001, 0.001)$.

We set $\tau_m^2 \sim \text{IG}(0.001, 0.001)$. The data seemed less informative for the population variance parameters τ_m^2 and τ_s^2 in the M_y - S_y model, especially for the latter (see Results below). Accordingly, we checked the sensitivity of the posterior estimates to the specification of the prior of τ_s^2 . Under the scaling transformation, we compared the priors $\text{IG}(0.01, 0.01)$ and $\text{IG}(0.1, 0.1)$. Under the SQRT transformation, we compared the priors $\text{IG}(0.001, 0.001)$ and $\text{IG}(0.01, 0.01)$.

For the M_y - S model, we set $p(\ln(S)) \propto N(2.5, 10^3)$. For the M - S_y model, we set $p(M) \propto N(210, 10^5)$ and $\tau_s^2 \sim \text{IG}(0.001, 0.001)$. Other priors were kept the same as for the M_y - S_y model.

Results

Convergence diagnostics

We first describe in detail the steps taken to assess convergence (Appendix B) for the M_y - S_y model with the scaling transformation of the counts. Similar processes were repeated for other models and data transformations.

The MCMC did not converge if we used the M_y - S_y model with the complete data set. A fundamental ambiguity existed between τ_m^2 and τ_s^2 . The MCMC would spend long periods with either τ_m^2 or τ_s^2 at a value of zero, effectively reducing the M_y - S_y model to one of its simpler forms. Using these simpler forms, the M_y - S and M - S_y models, the MCMC did converge. The M_y - S_y also converged when the first three years of the data series, which were very uninformative, were discarded.

We ran the Gibbs sampler an initial 3000 iterations to tune the standard deviations of the proposal distributions used in the Metropolis steps to draw values of M_y and S_y so as to obtain an acceptance rate in the range of 40–60% for each of these parameters (Appendix A). We then ran the Gibbs sampler for an additional 5000 iterations. The results from these 5000 iterations showed that the hyperparameters and the

Fig. 1. Count data (circles) and model fits for Kadashan Creek, 1974–1998. The solid line is the fit of the M_y -S model using scaled counts and a nonconstant stream life. The broken line is the fit obtained by separate estimation of each year.

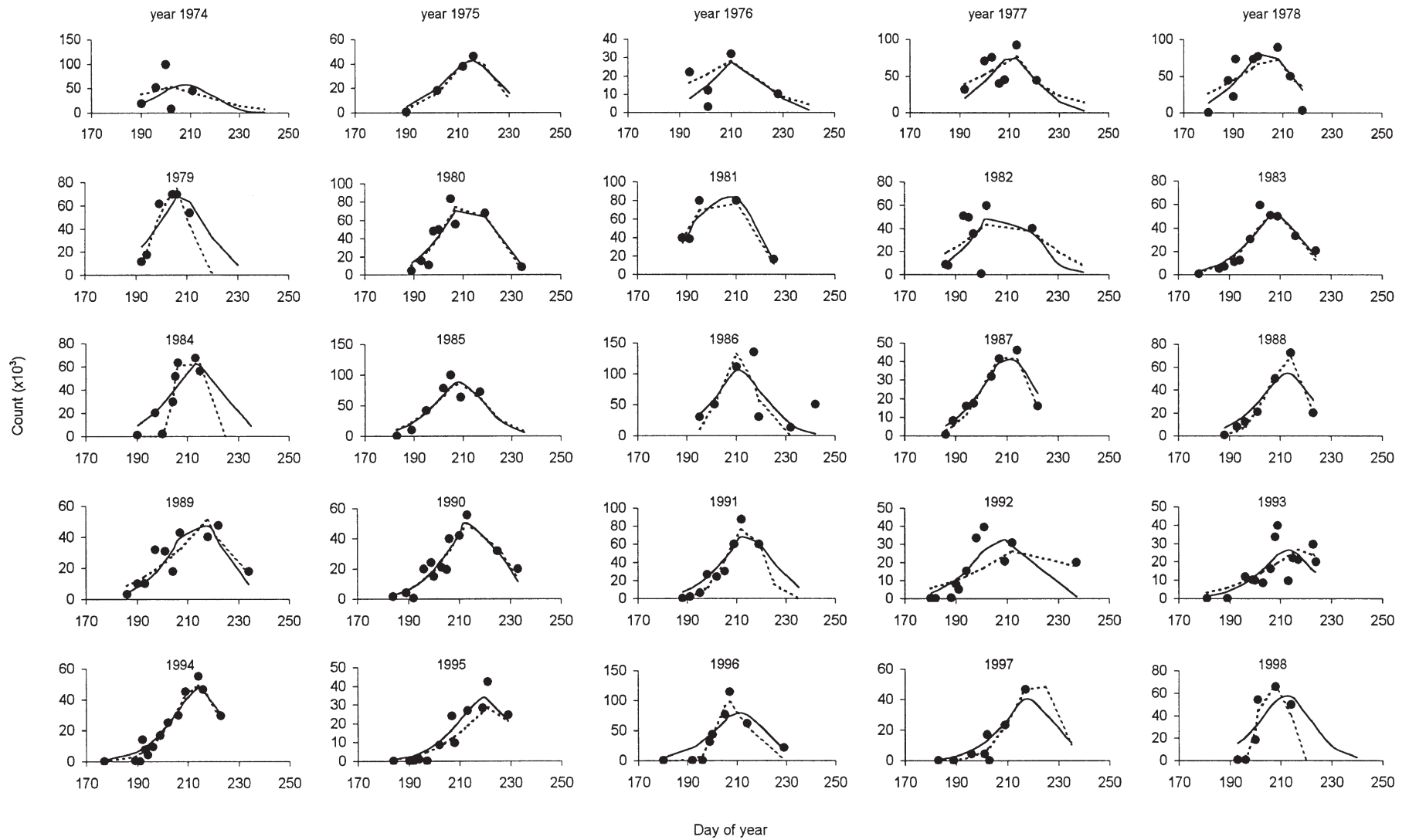
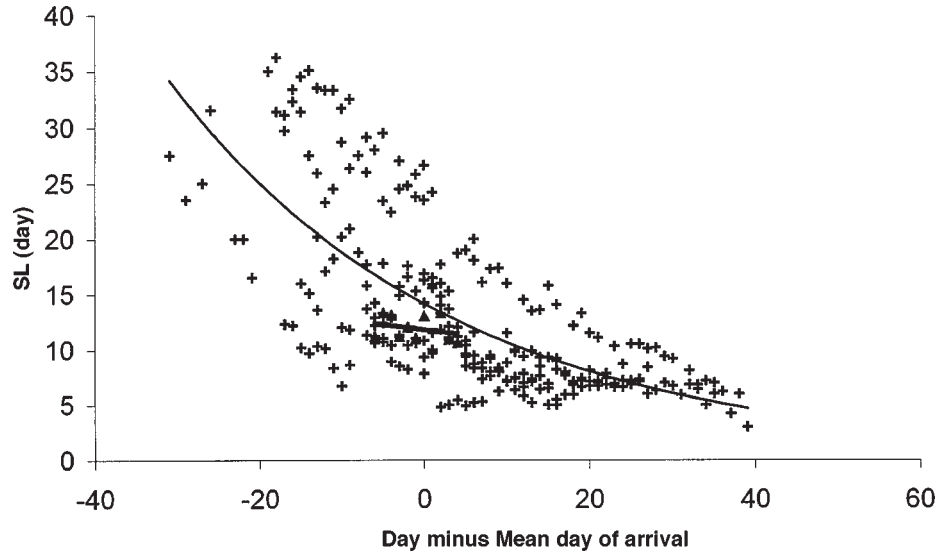


Fig. 2. Combined stream life (SL) data and those for Kadashan Creek from Dangel and Jones (1988) and the exponential fitting curves for these two data sets. Crosses denote combined SL data and triangles denote SL data for Kadashan Creek. The thin line is the fit for combined data and the heavy line is for the Kadashan data.



timing parameters had relatively high autocorrelations (lag 1 autocorrelations range from 0.8 to 0.9). Their autocorrelation function decayed slowly, approaching zero only after about lag 15. In contrast, most of the escapement parameters had low autocorrelations. The Raftery and Lewis (1992) convergence diagnostic (Appendix B) indicated that τ_m^2 and τ_s^2 were the slowest to converge, requiring 35 000 iterations in order to get an accuracy for the 2.5th percentiles within ± 0.005 with 95.5% probability.

Due to these autocorrelations, we reran the Gibbs sampler for 200 000 iterations with a thinning interval of 20. This time, all diagnostics were satisfied. We then ran another Gibbs sampler sequence with different starting values, and Gelman and Rubin's (1992) diagnostic (Appendix B) indicated convergence of the chains. Thus, the output after convergence was used for subsequent inference.

Model checking and comparison

Table 1 shows the values of the coefficients of skewness and kurtosis of the standardized residuals (Appendix C). The value of skewness for all cases is near zero, indicating that the standardized residuals achieve reasonable symmetry in their distributions (Snedecor and Cochran 1982) and implying that the models are neither overfitted nor underfitted. The values of the kurtosis indicate possible heavy tails in the distribution, less so with the scaling than the SQRT transformation. We also explored a logarithmic data transformation (not shown) and found it to be worse than SQRT. In summary, scaling the counts seems slightly more appropriate for this data set than SQRT.

Table 1 also reports the values of model comparison quantities (Appendix C). With scaling, the M_y-S model has the smallest value of the negative cross-validation log-likelihood (ncvLogL) (Appendix C), indicating that it is preferred. The values of the deviance information criterion (DIC) for the M_y-S_y and M_y-S models are very close, so the M_y-S model would also be the preferred model according to the principle of parsimony. The values of both ncvLogL and DIC indicate

that the $M-S_y$ model fit the data less well than the other two models. Under SQRT, the two model comparison criteria favor the M_y-S_y model more than the simpler models. Under both data transformations, the estimated effective number of parameters (p_D) for the M_y-S_y model is in the range of 43–47 versus the nominal value of 71. For the two simpler models, the values of p_D are in the range of 38–41 versus the nominal value of 48.

Sensitivity to the prior

Preliminary analyses indicated that τ_s^2 was the parameter most sensitive to prior specifications with the M_y-S_y model. Under scaling, with two priors of τ_s^2 (IG(0.01,0.01) and IG(0.1,0.1)), the posterior medians and 95% credible intervals of τ_s^2 were 0.023 (0.004, 0.127) and 0.069 (0.024, 0.244), respectively (see the values in the bordered area in Table 2). The first prior has a peak at 0.0099 and the latter at 0.09, indicating that the priors tended to pull the posteriors towards their modes. Under SQRT transformation, the posteriors of τ_s^2 were not so sensitive, but their left tails seemed to be influenced by the priors. The other parameters were much more robust to their priors, especially the population means and the yearly parameters (Table 2).

Data transformations, model forms, and posterior estimates

For the M_y-S_y and M_y-S models, posterior inference was more similar under the same data transformation than between the data transformations (Table 2). The $M-S_y$ model obtains slightly different results compared with other two models. When using the simpler models that fix aspects of escapement timing, more variation is attributed to the remaining set of timing parameters and the measurement error (see the increased values of τ_m^2 , τ_s^2 , and (or) σ^2 compared with those of the M_y-S_y model (Table 2)).

The 95% credible intervals of E_y were the most sensitive to the data transformation (Table 2). Scaling produced narrower 95% credible intervals of E_y than SQRT for most

Table 1. Model checking and choice quantities for data from 1977 to 1998.

Measurement	Coefficient of skewness	Kurtosis	ncvLogL	\bar{D}	p_D	DIC
Scaling						
$M_y-S_y^a$	0.138	4.16	-6.3	-86.8	42.8	-43.9
$M_y-S_y^b$	0.108	4.16	-5.3	-87.7	44.6	-43.1
M_y-S	0.181	4.13	-7.3	-84.2	40.9	-43.3
$M-S_y$	0.259	3.31	3.8	-48.3	39.2	-9.1
SQRT						
$M_y-S_y^c$	-0.385	4.73	1073.5	2076.4	47.0	2123.4
$M_y-S_y^d$	-0.385	4.74	1073.0	2078.6	45.2	2123.8
M_y-S	-0.189	5.76	1074.6	2097.4	38.2	2135.6
$M-S_y$	-0.307	4.20	1078.3	2106.2	41.0	2147.2

Note: ncvLogL = $-\sum \log(\text{cpo}_{y,i})$, the negative cross-validation log-likelihood, \bar{D} is the posterior mean of deviance, p_D is the effective number of parameters, and DIC is the deviance information criterion (Appendix C).

^aPrior for $\tau_s^2 \sim \text{IG}(0.01, 0.01)$.

^bPrior for $\tau_s^2 \sim \text{IG}(0.1, 0.1)$.

^cPrior for $\tau_s^2 \sim \text{IG}(0.001, 0.001)$.

^dPrior for $\tau_s^2 \sim \text{IG}(0.01, 0.01)$.

years. The SQRT transformation has the property of down-weighting the large counts and up-weighting the small counts. Thus, the tail area of the count curves was fit better than with scaling, but fits around the peaks could be quite different. For the years 1979, 1984, 1991, and 1998, which are years with a single count shortly after the observed peak, SQRT would quite often treat this count as a value with large observation error rather than a point containing important information about escapement timing. This may be the reason that SQRT produced large variance estimates of escapements for these years.

The $M-S_y$ model yields narrower 95% credible intervals of E_y than the other two models, especially under SQRT (see bolded values in Table 2). By forcing the M_y to a common value, we greatly reduce the uncertainty in E_y for years with little postpeak data (Figs. 1 and 3).

The posterior distributions of M_y and E_y achieve reasonable symmetry for the informative years (e.g., 1980, 1983, 1989, and 1990, etc.) under both kinds of data transformations (Fig. 4). However, posterior distributions of M_y and E_y are positively skewed for years such as 1974, 1975, 1979, 1984, 1991, 1992, 1997, and 1998 (Fig. 4), which are years with few or no postpeak counts (Fig. 1).

Effects of stream life

Assumptions about stream life appeared to be very important in determining the magnitude of the escapement estimates and also influenced estimates of variance. Using a constant stream life resulted in posterior medians of the E_y that were bigger and 95% credible intervals that were wider than with a nonconstant stream life (see bolded values in the bordered area in Table 2). We separately analyzed the effects of the two parameters of our stream life model (eq. 3). For a fixed ϕ_2 , the three yearly parameters increase as ϕ_1 decreases. When ϕ_1 is fixed, M_y and S_y decrease and E_y increases as ϕ_2 decreases.

Comparison of the HBM with separate estimation of each year

Table 3 compares the posterior estimates obtained from the M_y-S_y model with scaling with the bootstrap estimates from separate estimation for four representative years. The 1990 data (Fig. 1) contain two postpeak counts, one many days past the peak, and the counts show no apparent pulses. The 1992 data contain three postpeak counts but have apparent pulses and one possible outlier that result in a large amount of “measurement” error. The 1997 data have no postpeak counts, while the 1998 data have one count shortly after the peak.

In all cases, separate estimates match the observed count data more closely than the HBM fit but differ more among years in run timing (Fig. 1). Nevertheless, the quality of the separate estimates is more sensitive to that of the data (Table 3). Where the data are informative (e.g., 1990), the point estimates (medians) from these two approaches are comparable (Table 3); confidence intervals using the HBM are wider because the HBM approach considers the properties of other years where counts suggest different run timing. In years where no postpeak counts exist (e.g., 1997) or the data have irregular features (e.g., 1992, 1998 (with an anomalous run timing)), separate estimation is unable to unambiguously determine the run timing (see bolded medians and confidence intervals in Table 3) and consequently the escapement, whereas HBM is able to use the shapes of other years’ timing curves to narrow the uncertainty considerably.

Simulation studies (Adkison and Su 2001, table 4) show that the bootstrap confidence intervals from separate estimation underestimate the uncertainty of the parameters, while the HBM provides approximately correct uncertainty estimates. This may explain some of the instances of seemingly more precise estimates for separate estimation.

Discussion

In this paper, we have extended escapement estimation methods such as the deterministic area-under-the-curve method (English et al. 1992) and statistical modeling of each year separately (Quinn and Gates 1997; Hilborn et al. 1999) to a hierarchical Bayesian paradigm. Methods that treat each year’s data separately are prone to failure in years where count data are sparse, particularly if counts are lacking after the peak in escapement (Hilborn et al. 1999; Adkison and Su 2001). Our HBM approach is able to use historical escapement data for a stream to “borrow strength” from years with good data to obtain improved annual estimates for escapement and its timing in years with poor data.

Our approach also differs from traditional escapement estimation methods in incorporating the well-known decline in longevity on the spawning grounds as the season progresses (Perrin and Irvine 1990). The Kadashan example shows that spawning abundance models are sensitive to assumptions about stream life, in particular in estimating the magnitude of escapement. Since stream life can vary by year as well as by date within a year, this is an important source of error in escapement estimation.

Observer efficiency can be another important factor influencing escapement estimation. It is known that aerial and foot surveys often undercount the number of spawners pre-

Table 2. Posterior summary for the hyperparameters and annual parameters obtained from the M_y-S_y , M_y-S , and $M-S_y$ models under two types of data transformation and two stream life models using data from 1977 to 1998.

Data transformation		$\mu_m (M)$	τ_m^2	$\mu_s (\ln(S))$	τ_s^2	σ^2	M_y	$\ln(S_y)$	$E_y (\times 10^3)$
Scaling	$M_y-S_y^a$	207.7	20.1	2.6	0.023	0.038	207.6	2.6	110.5
		(205.0, 210.9)	(7.5, 48.7)	(2.4, 2.8)	(0.004, 0.127)	(0.03, 0.05)	(10.5)	(0.7)	(68.6)
	$M_y-S_y^b$	207.8	18.6	2.6	0.069	0.038	207.7	2.6	110.1
		(205.1, 211.1)	(5.5, 47.0)	(2.4, 2.9)	(0.024, 0.224)	(0.03, 0.05)	(11.3)	(0.9)	(69.4)
	M_y-S	207.7	21.7	2.6		0.038	207.6		111.4
		(205.2, 210.8)	(10.0, 51.4)	(2.4, 2.8)		(0.03, 0.04)	(9.4)		(68.0)
SQRT	$M-S_y$	209.6		2.8	0.096	0.046		2.8	113.3
		(207.7, 212.2)		(2.5, 3.0)	(0.028, 0.268)	(0.04, 0.06)		(0.8)	(66.6)
	$M_y-S_y^c$ (CnSL)	206.5	22.5	2.4	0.029	0.035	206.4	2.4	142.8
		(204.1, 209.2)	(10.3, 52.8)	(2.2, 2.5)	(0.005, 0.128)	(0.03, 0.05)	(9.3)	(0.7)	(102.3)
	$M_y-S_y^d$	208.3	21.6	2.5	0.049	2123	208.2	2.5	108.6
		(205.7, 211.5)	(6.3, 53.6)	(2.3, 2.7)	(0.002, 0.216)	(1680, 2725)	(11.9)	(0.8)	(105.1)
SQRT	$M_y-S_y^e$	208.4	21.0	2.5	0.063	2097	208.2	2.5	108.4
		(205.7, 211.5)	(6.3, 54.2)	(2.3, 2.7)	(0.010, 0.236)	(1653, 2706)	(11.9)	(0.8)	(103.1)
	M_y-S	208.4	22.6	2.5		2328	208.4		108.9
		(205.7, 211.7)	(9.2, 57.1)	(2.4, 2.6)		(1874, 2947)	(10.8)		(111.0)
	$M-S_y$	209.2		2.6	0.122	2434		2.6	108.4
		(207.5, 211.2)		(2.4, 2.8)	(0.044, 0.325)	(1964, 3089)		(0.8)	(88.3)

Note: For the hyperparameters and σ^2 , the median of the posterior distribution is shown with the boundaries of the 95% credible interval below in parentheses. For the annual parameters, the average across years of the posterior median is given with the average width of the 95% credible interval below in parentheses.

^aPrior for $\tau_s^2 \sim \text{IG}(0.01, 0.01)$.

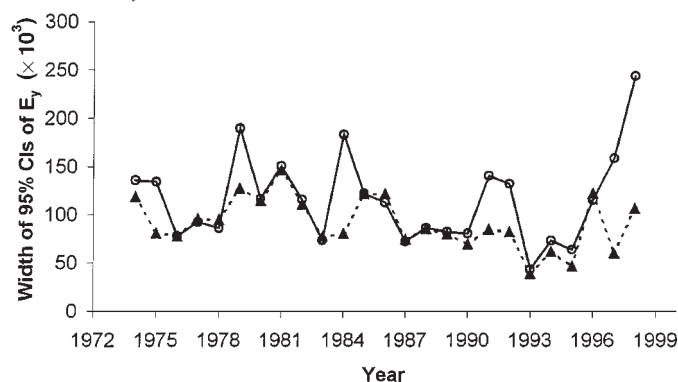
^bPrior for $\tau_s^2 \sim \text{IG}(0.1, 0.1)$.

^cAssumes that stream life is a constant (CnSL) rather than declining with date.

^dPrior for $\tau_s^2 \sim \text{IG}(0.001, 0.001)$.

^ePrior for $\tau_s^2 \sim \text{IG}(0.01, 0.01)$.

Fig. 3. Width of the 95% credible interval (CI) for E_y for each year. Results are shown for the M_y-S and $M-S_y$ models using the SQRT transformation of the counts and a nonconstant stream life. The solid line is for the M_y-S model and the dashed line is for the $M-S_y$ model.



sented in a stream, especially when the density of the fish is high (Jones et al. 1998). Within- and between-observer variability and bias of counts are also observed (Jones et al. 1998). The pink salmon counts that we used here have been partially corrected for bias (Van Alen 2000). Hilborn et al. (1999) suggested that weir studies on index streams should be used to estimate observer bias and stream life.

In our example, we found that appropriate error assumptions for the counts and simplification of the HBM were necessary for convergence of the Gibbs sampler. For example, the Gibbs sampler may fail to converge under the M_y-S_y

model setting if the counts are not transformed because the scale of the counts for different years may vary by several orders of magnitude. The scaling and SQRT transformations proposed here are both able to address this problem, but the scaling transformation seemed to work better in our example. Nevertheless, the difference between SQRT and scaling fits in these sorts of cases was smaller than that resulting from using a separate estimation approach, where data from uninformative years did not constrain escapement timing.

Further refinement of the HBM is possible. Pulsed entry (caused by irregular fishing and variability in the entry process) is typical for salmon escaping to spawning streams. Although these pulses are smoothed by the overlap of spawners who have entered the stream on different days, spikes in counts are still quite visible in much stream data. This kind of variation (caused by the variability in entry process) is not counted in the spawning abundance model explicitly but is treated as observation error. Excessive "observation" error in combination with few postpeak counts can cause convergence problems for the Gibbs sampler.

In our example, we were unable to differentiate variability in the mean date of escapement (τ_m^2) from variability in the duration of escapement (τ_s^2). By retreating to the simpler M_y-S or $M-S_y$ model, we were able to complete the analysis. Refined stream entry models that used environmental conditions, such as floods and rainfall, might explain some of the higher-frequency variation in spawner counts and thus increase the estimability of the timing parameters of the HBM.

As Millar and Meyer (2000) pointed out, "Fitting a model is just the beginning." In this paper, we addressed the important problem of model selection and adequacy using predic-

Fig. 4. Posterior median and 95% credible interval for (a) M_y and (b) E_y for each year. Results are shown for the M_y – S_y model using scaled counts and a nonconstant stream life.

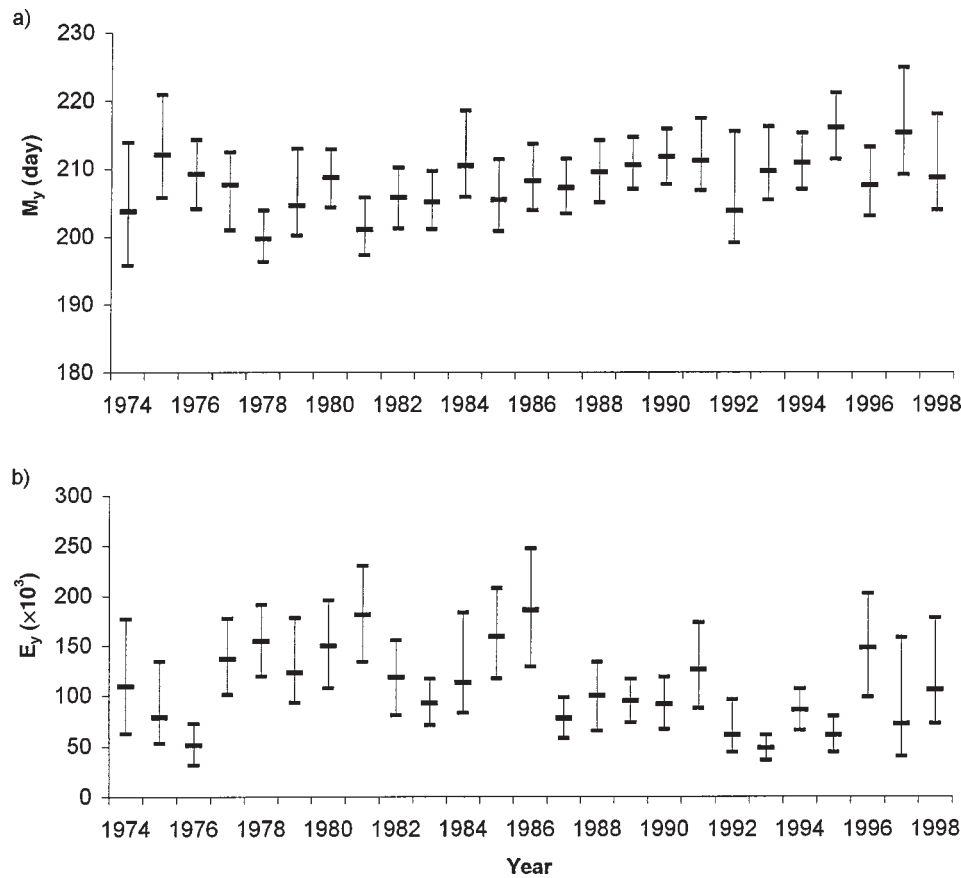


Table 3. Posterior summary (median with the 95% credible interval below in parentheses) for the annual parameters obtained from an HBM (the M_y – S_y model with scaling with priors for $\tau_m^2 \sim \text{IG}(0.001, 0.001)$ and for $\tau_s^2 \sim \text{IG}(0.01, 0.01)$) versus the bootstrap summary statistics (median with the 95% confidence interval below in parentheses) from separate estimation (SepEst) for four divergent years (see text).

Year	M_y		S_y		$E_y (\times 10^3)$	
	HBM	SepEst	HBM	SepEst	HBM	SepEst
1990	211.3 (207.4, 215.4)	212.8 (210.1, 216.6)	13.4 (9.9, 18.4)	14.7 (11.2, 19.8)	90.7 (66.6, 116.1)	97.6 (83.9, 111.9)
1992	203.4 (198.2, 215.7)	226.4 (203.8, >350.0 ^a)	13.9 (9.3, 21.0)	35.1 (14.4, >100.0 ^b)	59.8 (35.9, 93.1)	76.9 (37.4, 209.9)
1997	214.1 (208.2, 222.8)	216.4 (207.9, >350.0 ^a)	12.1 (7.8, 16.6)	10.7 (4.2, 40.1)	68.8 (39.7, 142.4)	86.5 (43.2, 1.9 × 10⁶)
1998	207.9 (202.7, 215.9)	200.8 (200.7, 200.8)	12.8 (8.3, 18.0)	1.0 (<1.0 ^c , 1.9)	97.7 (69.6, 152.1)	66.6 (60.7, 71.0)

^aThe value 350.0 is the upper bound of the constraint set for M_y .
^bThe value 100.0 is the upper bound of the constraint set for S_y .
^cThe value 1.0 is the lower bound of the constraint set for S_y .

tive model determination techniques. Such validation steps are rare in fisheries applications of Bayesian methods, probably due to the computational burden. Bayesian methods require much effort just for model fitting; sometimes, model assessment and selection are not well addressed (Millar and Meyer 2000). The Gibbs sampler and other MCMC techniques enable users to tackle complex models those are impossible to cope with by other methods. The predictive model determination quantities that we used in this example can be

easily calculated in the Gibbs sampling process, so routine model checking should not be too onerous.
An additional method of borrowing strength would be to incorporate spatial relationships as another hierarchical level in the model. In Southeast Alaska, pink salmon in neighboring streams often have similar run timing. They also experience correlated fluctuations in escapement, in response to both a common environment and a common harvest regime. Incorporating these common patterns would reduce the un-

certainty in escapement estimates even further. Weir studies on representative index streams would further increase the benefit of this approach.

Acknowledgements

We thank Drs. Terrance J. Quinn II, William Smoker, Albert Tyler, Jim Blick, Martin Liermann, Jerome J. Pella and Mike LaPointe and an anonymous reviewer for reviewing this manuscript and providing many helpful comments. This research was funded by the Rasmuson Fisheries Research Center and the Southeast Alaska Research committee.

References

- Adkison, M.D., and Su, Z. 2001. A comparison of salmon escapement estimates using a hierarchical Bayesian approach versus separate maximum likelihood estimation of each year's return. *Can. J. Fish. Aquat. Sci.* **58**. This issue.
- Browne, W.J., and Draper, D. 2000. A comparison of Bayesian and likelihood methods for fitting multilevel models. Technical report, Department of Mathematical Sciences, University of Bath, U.K.
- Carlin, B.P., and Louis, T.A. 1996. Bayes and empirical Bayes methods for data analysis. Chapman and Hall, New York.
- Cousens, N.B.F., Thomas, G.A., Swann, C.G., and Healey, M.C. 1982. A review of salmon escapement estimation techniques. *Can. Tech. Rep. Fish. Aquat. Sci.* No. 1108.
- Dangel, J.R., and Jones, J.D. 1988. Southeast Alaska pink salmon total escapement and stream life studies, 1987. Regional Inf. Rep. No. IJ88-24. Alaska Department of Fish and Game, Juneau, Alaska.
- Davidian, M., and Giltinan, D.M. 1995. Nonlinear models for repeated measurement data. Chapman and Hall, London, U.K.
- English, K.K., Bocking, R.C., and Irvine, J.R. 1992. A robust procedure for estimating salmon escapement based on the area-under-the-curve method. *Can. J. Fish. Aquat. Sci.* **49**: 1982–1989.
- Fukushima, M., and Smoker, W.W. 1997. Determinants of stream life, spawning efficiency, and spawning habitat in pink salmon in the Auke Lake system, Alaska. *Can. J. Fish. Aquat. Sci.* **54**: 96–104.
- Gelfand, A.E., and Smith, A.F.M. 1990. Sampling based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**: 398–409.
- Gelman, A., and Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **7**: 457–511.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 1995. Bayesian data analysis. Chapman and Hall, London, U.K.
- Geman, S., and Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE (Inst. Electr. Electron. Eng.) Trans. Pattern Anal. Machine Intelligence*, **6**: 721–741.
- Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. 1996. Markov chain Monte Carlo in practice. Chapman and Hall, London, U.K.
- Heard, W.R. 1991. Life history of pink salmon (*Oncorhynchus gorbuscha*). In Pacific salmon life histories. Edited by G. Groot and L. Margolis. University of British Columbia Press, Vancouver, B.C.
- Hilborn, R., Bue, B.G., and Sharr, S. 1999. Estimating spawning escapements from periodic counts: a comparison of methods. *Can. J. Fish. Aquat. Sci.* **56**: 888–896.
- Hobert, J.P., and Casella, G. 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Am. Stat. Assoc.* **91**: 1461–1473.
- Jones, E.L., III, Quinn, T.J., II, and Van Alen, B.W. 1998. Observer accuracy and precision in aerial and foot survey counts of pink salmon in a southeast Alaska stream. *N. Am. J. Fish. Manage.* **18**: 832–846.
- Liermann, M., and Hilborn, R. 1997. Depensation in fish stocks: a hierarchical Bayesian meta-analysis. *Can. J. Fish. Aquat. Sci.* **54**: 1976–1984.
- Lindley, D.V., and Smith, A.F.M. 1972. Bayes estimates for the linear model (with discussion). *J. R. Stat. Soc. Ser. B (Methodol.)*, **34**: 1–41.
- Millar, R.B., and Meyer, R. 2000. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. *Can. J. Fish. Aquat. Sci.* **57**: 43–50.
- Mosteller, F., and Chalmers, T.C. 1992. Some progress and problems in meta-analysis of clinical trials. *Stat. Sci.* **7**: 227–236.
- Perrin, C.J., and Irvine, J.R. 1990. A review of survey life estimates as they apply to the area-under-the curve method for estimating the spawning escapement of Pacific salmon. *Can. Tech. Rep. Fish. Aquat. Sci.* No. 1733.
- Quinn, T.J., II, and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.
- Quinn, T.J., II, and Gates, R. 1997. Estimation of salmon escapement: models with entry, mortality and stochasticity. *Nat. Resour. Model.* **10**: 217–250.
- Raftery, A.E., and Lewis, S. 1992. How many iterations in the Gibbs sampler? In Bayesian statistics 4. Edited by J.M. Bernardo, J. Berger, A.P. Dawid, and A.F.M. Smith. Oxford University Press, Oxford, U.K. pp. 763–773.
- Snedecor, W., and Cochran, W.G. 1982. Statistical methods. Iowa State University Press, Ames, Iowa.
- Spiegelhalter, D.J., Thomas, A., Best, N.G., and Gilks, W.R. 1997. BUGS: Bayesian inference using Gibbs sampling, version 0.60. MRC, Biostatistics Unit, Cambridge, U.K.
- Starr, P., and Hilborn, R. 1988. Reconstruction of harvest rates and stock composition in gauntlet salmon fisheries: application to British Columbia and Washington sockeye. *Can. J. Fish. Aquat. Sci.* **45**: 2216–2229.
- Tierney, L. 1994. Markov chains for exploring posterior distributions (with discussion). *Ann. Stat.* **22**: 1701–1762.
- Van Alen, B.W. 2000. Status and stewardship of salmon stocks in Southeast Alaska. In Sustainable fisheries management: Pacific salmon. Edited by E.E. Knudsen, C.R. Steward, D.D. MacDonald, J.E. Williams, and D.W. Reiser. Lewis Publishers, Boca Raton, Fla.

Appendix A. The Metropolis steps

We use a Metropolis step with a normal proposal distribution $N(\vartheta_i^* | \vartheta_i^{(t-1)}, SD_i^2)$, $i = 1, \dots, p$, to update each of those parameters with no standard-form full conditional distributions, where $\vartheta_i^{(t-1)}$ is the current value of a parameter, ϑ_i^* is a candidate value, and SD_i is a specified standard deviation for the proposal distribution. In a Metropolis step, ϑ_i^* is accepted as an update with probability $\min(1, p(\vartheta_i^* | \{\vartheta\}_{j \neq i}) / p(\vartheta_i^{(t-1)} | \{\vartheta\}_{j \neq i}))$; otherwise, it is rejected and we set $\vartheta_i^{(t)} = \vartheta_i^{(t-1)}$.

The choice of SD_i affects the efficiency of Metropolis steps and thus the whole algorithm. High values of SD_i result in large proposal moves and low acceptance rate because most of the candidate moves may step into low-density areas of the posterior distribution and thus be rejected. Small values of SD_i result in high acceptance rate and slow movement of the chain.

The performance of the Metropolis algorithm can be expressed by the acceptance rate of the candidate draws in the Metropolis steps. Theoretical and empirical results show that the acceptance rate in the range of 20–50% (depending on the number of parameters) provides optimal performance (Gelman et al. 1995). For multilevel models, Browne and Draper (2000) proposed an acceptance rate of 40–60% for univariate updating.

To increase the efficiency of the Metropolis algorithm, we adopt an adaptive tuning step similar to that of Browne and Draper (2000) to tune the SD_i before generating sample draws for inference. The goal of the tuning is to obtain a target acceptance rate of around 50%. The adaptive step is stopped after a fixed number of iterations, after which the burn-in period (the preconvergence period) and main monitoring run (the postconvergence period) are started.

Appendix A references

Browne, W.J., and Draper, D. 2000. A comparison of Bayesian and likelihood methods for fitting multilevel models. Technical report, Department of Mathematical Sciences, University of Bath, U.K.
Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 1995. Bayesian data analysis. Chapman and Hall, London, U.K.

Appendix B. Markov chain Monte Carlo convergence diagnostics

A critical issue in using MCMC methods is how to determine when random draws have converged to the posterior distribution (Gelman et al. 1995). Best et al. (1995) have developed a collection of S-plus routines called CODA for convergence diagnostics of MCMC sampling. Several general diagnostic methods have been incorporated in this free package, including methods proposed by Geweke (1992), Heidelberger and Welch (1983), Raftery and Lewis (1992), and Gelman and Rubin (1992). CODA also provides summary estimates, kernel density plots, autocorrelations, and cross-correlations for the monitored quantities. We used the results from these CODA routines to assess convergence of our MCMCs. We provide a summary of their underlining theoretical considerations, implementation, and criteria for assessing convergence in Table B1.

Table B1.

Aspect monitored	Diagnostic	No. of chains	Diagnostic steps	Theory and convergence criteria
Trend	Geweke 1992	Single	(1) Compare the mean in the first $x\%$ (e.g., 10%) portion of the simulation output of a quantity of interest with that containing the last $y\%$ (e.g., 50%) of the output by a Z score (2) If $ Z $ is extreme, the first $x\%$ can be discarded and the Geweke diagnostic be repeated and so on	If the chain is in stationary, the means of the values early and late in the sequence should be similar. The diagnostic is a Z score that is the difference of the two means divided by an asymptotic standard error of the difference of the means. As length of chain $\rightarrow \infty$, $Z \rightarrow N(0,1)$, so Geweke $ Z < 2$ means convergence
	Heidelberger and Welch 1983	Single	(1) Conduct a test for the null hypothesis of stationarity for the entire simulation output of a quantity first (2) If the null hypothesis is rejected, the test is repeated after discarding the first, e.g., 10% of the series. This process will continue until either the null hypothesis is passed or 50% of the iterations have been discarded and the null hypothesis is still rejected. In the latter case, CODA will report the failure of the stationarity test and the value of the Cramer – von Mises (C-vonM) statistic	Use a Brownian bridge theory and the C-vonM statistic for the stationarity test. The 95th percentile of C-vonM is 0.46, which can be used to judge if the stationary test is passed

Table B1 (*concluded*).

Aspect monitored	Diagnostic	No. of chains	Diagnostic steps	Theory and convergence criteria
			(3) For the portion of the chain passing the stationarity test, a half-width ($1.96 \times$ asymptotic standard error) test is further conducted to check if the half-width $<$ posterior mean $\times \epsilon$ (a small fraction, e.g., 0.1). If this test is rejected, a longer run is needed to increase the accuracy of the estimates	
Autocorrelation	Raftery and Lewis 1992	Single	<p>(1) Suppose we wish to estimate the 2.5th percentile to within ± 0.5 with probability 0.95 for each parameter</p> <p>(2) First, take a pilot run of at least N_{\min} iterates (the minimum number of iterations needed to obtain the desired accuracy of the estimate if the draws were independent). This diagnostic then provides recommendations for an initial number of iterates to discard M (burn-in), a total number of iterates N needed to obtain the desired accuracy of the estimate for which every kth is stored, and finally a "dependence factor" $I = N/N_{\min}$</p>	The value of I measures the increase in the number of iterations needed to reach convergence due to dependence between the samples in the chain. Values of I near 1.0 indicate good mixing, whereas values of I much greater than 1.0 indicate high autocorrelation within the chain (Raftery and Lewis (1992) suggested that $I > 5.0$ indicates problems)
Mixing of sequences	Gelman and Rubin 1992	Multiple	Run multiple chains with overdispersed starting points to see if all of the chains can converge to the same stationary distribution. CODA reports the median and the 97.5 percentile (estimated from the second half of each chain) of the sampling distribution for "shrink factor." Values of these two quantiles near 1.0 may mean effective convergence of all of the chains	The shrink factor involves both the between- and within-chain variances. It approaches 1.0 when the chains have mixed; at this point, the two variance components are essentially equal (Gelman et al. 1995)

Appendix B references

- Best, N.G., Cowles, M.K., and Vines, S.K. 1995. CODA manual version 0.30. MRC, Biostatistics Unit, Cambridge, U.K. (Updated in CODA version 0.4: manual addendum, 1997.)
- Gelman, A., and Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **7**: 457–511.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 1995. Bayesian data analysis. Chapman and Hall, London, U.K.
- Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *In Bayesian statistics 4. Edited by J.M. Bernardo, J. Berger, A.P. Dawid, and A.F.M. Smith.* Oxford University Press, Oxford, U.K. pp. 169–193.
- Heidelberger, P., and Welch, P.D. 1983. Simulation run length control in the present of an initial transient. *Oper. Res.* **31**: 1109–1144.
- Raftery, A.E., and Lewis, S. 1992. How many iterations in the Gibbs sampler? *In Bayesian statistics 4. Edited by J.M. Bernardo, J. Berger, A.P. Dawid, and A.F.M. Smith.* Oxford University Press, Oxford, U.K. pp. 763–773.

Appendix C. Model checking and comparison

In this appendix, we will address the important problem of model comparison and adequacy checking. Any model is necessarily a simplification of complex real-world phenomena. Hence, checking the fit of a model to data and other knowledge is always important. Nevertheless, such validation steps are rare in fisheries applications of Bayesian methods, probably due to the computational burden (Millar and Meyer 2000). We apply various predictive model evaluation techniques (Gelfand 1996) for model adequacy checking and comparison in an HBM setting. These model determination quantities can be easily calculated during the Gibbs sampling processes; therefore, routine model checking should not be too onerous.

Cross-validation residuals

Standardized cross-validation (cv) residuals are defined as

$$(C1) \quad r_{y,t}^{cv} = \frac{c_{y,t} - E(c_{y,t} | D_{(y,t)})}{\sqrt{\text{var}(c_{y,t} | D_{(y,t)})}}$$

where $c_{y,t}$ denotes the observed count and $D_{(y,t)}$ denotes all elements of the observed data set $D = \{c_{y,t}\}$ except $c_{y,t}$. The quantities $E(c_{y,t} | D_{(y,t)})$ and $\text{var}(c_{y,t} | D_{(y,t)})$ are the expected value and the variance of $c_{y,t}$ from the cross-validation predictive distribution $p(c_{y,t} | D_{(y,t)})$. Once we have obtained a sample $\{\phi_k^*\}$, where ϕ is the parameter vector and $k = 1, \dots, G$, from the posterior $p(\phi | D)$ through MCMC or other sampling techniques, $E(c_{y,t} | D_{(y,t)})$ can be calculated using the following approximation for a large data set (Carlin and Louis 1996, p. 219):

$$(C2) \quad E(c_{y,t} | D_{(y,t)}) \approx \frac{1}{G} \sum_{k=1}^G E(c_{y,t} | \phi_k^*)$$

where $E(c_{y,t} | \phi_k^*) = \eta_{y,t}(\phi_k^*)$. The variance term can be calculated using

$$(C3) \quad \text{var}(c_{y,t} | D_{(y,t)}) = E(c_{y,t}^2 | D_{(y,t)}) - [E(c_{y,t} | D_{(y,t)})]^2$$

and

$$(C4) \quad E(c_{y,t}^2 | D_{(y,t)}) \approx \frac{1}{G} \sum_{k=1}^G [\text{var}(c_{y,t} | \phi_k^*) + [E(c_{y,t} | \phi_k^*)]^2]$$

Here, $\text{var}(c_{y,t} | \phi_k^*)$ is simply the variance of the counts, or σ^2 .

We used the skewness and kurtosis of the standardized residuals to assess the normality of the residuals. Other diagnostic analyses traditionally performed on residuals can also be used.

Conditional predictive ordinate

We compute the conditional predictive ordinate (CPO), which is the actual value of the cross-validation predictive distribution $p(c_{y,t} | D_{(y,t)})$, for comparison of alternative model forms. The Monte Carlo estimate of the CPO (Gelfand 1996) is given by

$$(C5) \quad \text{cpo}_{y,t} = G \left(\sum_{k=1}^G (1/f(c_{y,t} | \phi_k^*)) \right)^{-1}$$

where

$$f(c_{y,t} | \phi_k^*) = \left(1/\sqrt{2\pi\sigma^2} \right) \exp(-(c_{y,t} - (\eta_{y,t})_k^*)^2 / 2\sigma^2)$$

is the likelihood evaluated at ϕ_k^* . The negative cross-validation log-likelihood ($\text{ncvLogL} = -\sum \log(\text{cpo}_{y,t})$) (Spiegelhalter et al. 1997) is used to compare alternative models, with a smaller value ncvLogL indicating the preferred model.

Deviance information criterion (DIC)

We also use the DIC (Spiegelhalter et al. 1998) for our model comparisons. Let the deviance $D(\theta)$ be minus twice the log-likelihood of the data and θ be the parameter vector. Then

$$\text{DIC} = \text{fit} + \text{complexity}$$

where “fit” is defined as the posterior mean of $D(\theta)$, $\bar{D} = E(D)$, and “complexity” is represented by the effective number of parameters p_D , defined as $p_D = \bar{D} - D(\bar{\theta})$, where $D(\bar{\theta})$ is the deviance evaluated at the posterior expectations $\bar{\theta}$. A smaller value of DIC indicates a better-fitting model.

DIC can be readily calculated during an MCMC run by monitoring both θ and $D(\theta)$. At the k th iteration of MCMC run, we define

$$(C6) \quad D(\theta_k^*) = -2\ln(f(c_{y,t} | \theta_k^*)) = \sum_{\{y,t\}} (\ln(2\pi\sigma^2) + (c_{y,t} - (\eta_{y,t})_k^*)^2 / \sigma^2)$$

DIC is very useful for comparing hierarchical models, where the number of parameters is not clearly defined. With non-hierarchical models, p_D is asymptotically the true number of parameters and DIC is thus equivalent to the more familiar Akaike information criterion.

Appendix C references

- Gelfand, A.E. 1996. Model determination using sampling-based methods. *In* Markov chain Monte Carlo in practice. *Edited by* W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. Chapman and Hall, London, U.K. pp. 145–161.
- Millar, R.B., and Meyer, R. 2000. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. *Can. J. Fish. Aquat. Sci.* **57**: 43–50.
- Spiegelhalter, D.J., Thomas, A., Best, N.G., and Gilks, W.R. 1997. BUGS: Bayesian inference using Gibbs sampling, version 0.60. MRC, Biostatistics Unit, Cambridge, U.K.
- Spiegelhalter, D.J., Best, N.G., and Carlin, B.P. 1998. Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Technical report, MRC, Biostatistics Unit, Cambridge, U.K.

Copyright of Canadian Journal of Fisheries & Aquatic Sciences is the property of Canadian Science Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.