Hierarchical Bayesian estimation of unobserved salmon passage through weirs

Article in Canadian Journal of Fisheries and Aquatic Sciences · September 2017

DOI: 10.1139/cjfas-2016-0398

4 authors, including:

James R Jasper
Alaska Department of Fish and Game

Margaret Short
University of Alaska Fairbanks

William Stewart Grant
University of Alaska Fairbanks

NRC Research Press

ARTICLE

Hierarchical Bayesian estimation of unobserved salmon passage through weirs

James R. Jasper, Margaret Short, Chris Shelden, and W. Stewart Grant

Abstract: We developed a hierarchical Bayesian model (HBM) to estimate missing counts of Chinook salmon (*Oncorhynchus tshawytscha* (Walbaum in Artedi, 1792)) at a weir on the Kogrukluk River, Alaska, between 1976 and 2015. The model assumed that fish passage during a breach of the weir was typical of passage during normal operations. Counts of fish passing the weir were missing for some days during the runs, or only partial counts for a given 24-hour period were available. The HBM approach provided more defensible estimates of missing data and total escapement than ad hoc or year-by-year model estimates, because estimates of passage for a given year were informed not only by counts for the current year, but also by counts for all previous years. The results of the HBM yielded less variable estimates of escapement than did ad hoc or year-by-year model estimates. The HBM represents a standardized approach to estimate missing counts and total escapement and eliminates the need for ad hoc estimates of missing counts. The model also provides managers with a measure of uncertainty around estimates of escapement and around estimates of hyper-parameters to define run curves in subsequent years with incomplete fish counts.

Résumé : Nous avons développé un modèle hiérarchique bayésien (MHB) pour estimer les dénombrements manquants de saumons quinnats (*Oncorhynchus tshawytscha* (Walbaum in Artedi, 1792)) dans une fascine sur la rivière Kogrukluk, en Alaska, de 1976 à 2015. Une des hypothèses du modèle est que le passage de poissons après la formation d'une brèche dans la fascine est typique du passage dans des conditions normales de fonctionnement. Il manque des dénombrements de poissons passant par la fascine pour certains jours durant les migrations, et seuls des dénombrements partiels pour des périodes de 24 heures données sont disponibles. L'approche MHB produit des estimations plus valables des données manquantes et de l'échappement total que les estimations ad hoc ou issues de modèles d'estimation année par année, parce que les estimations du passage pour une année donnée reposent non seulement sur les dénombrements pour l'année concernée, mais aussi sur les dénombrements pour toutes les années précédentes. Les résultats du MHB produisent des estimations moins variables de l'échappement que les estimations ad hoc ou année par année. Le MHB représente une approche normalisée pour estimer les dénombrements manquants et l'échappement total et élimine la nécessité d'estimations ad hoc des dénombrements manquants. Le modèle fournit aussi aux gestionnaires une mesure de l'incertitude associée aux estimations de l'échappement et aux estimations d'hyperparamètres pour définir les courbes de migration pour les années subséquentes avec des dénombrements incomplets. [Traduit par la Rédaction]

Introduction

The migration of wild Pacific salmon through fishery management districts (escapement) is a key management goal to ensure that adequate numbers of maturing adults reach spawning areas to maintain healthy population sizes. Methods used to estimate escapement fall into two broad categories: (*i*) the counting of fish passing a particular point in the river, often aided by funneling migrating fish through an opening in a barrier such as a weir (Cousens et al. 1982), and (*ii*) the periodic estimation of fish abundances along a stretch of a river, which can be used as an index of escapement or expanded to infer total escapement (e.g., Irvine et al. 1992; Su et al. 2001). For the first category, fish are counted by direct observation at weirs or from counting towers or, in some cases, by sonar (Eggers 1994). In the second category, abundances are estimated by stream walkers, aerial surveys, or mark–recapture experiments (Labelle 1994; Manske and Schwarz 2000).

Counts of fish moving through a weir potentially provide more reliable estimates of escapement than do spawning-area surveys (e.g., Irvine et al. 1992) and mark–recapture methods (Schwarz et al. 1993; Labelle 1994), because counts at a weir remove uncertainties about residence times, which can introduce error into other estimates of escapement. However, estimates based on weir counts are reliable only when daily counts are made over most of the run (Adkison and Su 2001). In practice, many datasets lack counts after the peak run timing or after escapement goals have been reached. The reliability of estimates from weir counts depends on the fraction of the run observed, as well as on the accuracy of estimates of unobserved fish passage. Some fish passing a weir are inevitably uncounted. Procedures for estimating these uncounted fish vary among and within management agencies and are often undocumented in the technical literature.

Previous methods used to estimate unobserved passage depend on the nature of the missing component of the run. These components can be classified by the temporal segments of the missing counts: (*i*) the beginning or ending tail of a run, (*ii*) the building or waning portions of a run, and (*iii*) the peak of a run. In the first case, missing tail counts can be estimated by an exponential curve fit to the available ascending or descending counts. In the second case, missing ascending or descending counts can be estimated by linear or scaled interpolations between the counts for days

Received 2 September 2016. Accepted 8 September 2017.

J.R. Jasper, C. Shelden, and W.S. Grant. Division of Commercial Fisheries, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK 99518, USA.
M. Short. Department of Mathematics and Statistics, University of Alaska Fairbanks, P.O. Box 756660, Fairbanks, AK 99775-6660, USA.
Corresponding authors: James R. Jasper (email: Jim.Jasper@alaska.gov); W. Stewart Grant (email: william.grant@alaska.gov).

Published at www.nrcresearchpress.com/cjfas on 15 September 2017.

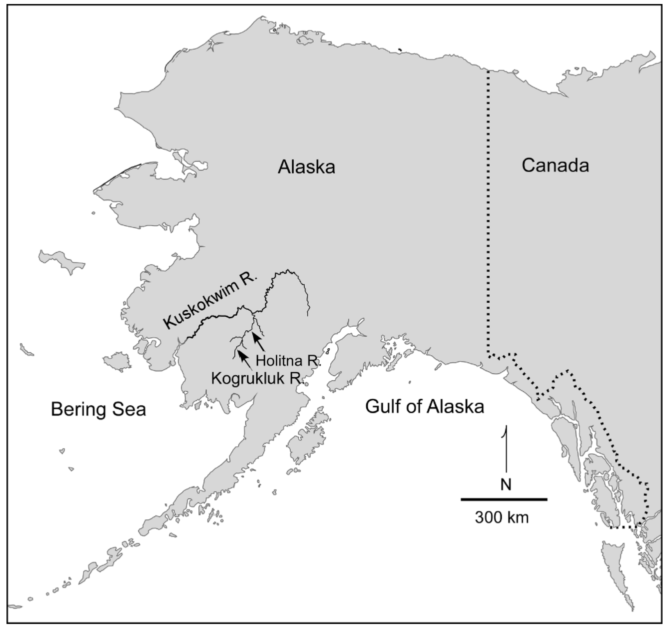
preceding and following a breach in the weir from flooding or floating debris, weekends off, or intrusions by bears. Linear interpolation may provide adequate estimates when the period of missing counts is short, but scaled interpolation may be needed to produce more realistic curves when observations have been interrupted for several days.

Estimation of unobserved passage in the third case — at the peak of the run — is most difficult and most prone to error. Ad hoc estimates of the peak passage usually depend on “expert opinion” by selecting a more complete run from another single year of data at the same weir or a run distribution from a neighboring stream that correlates highly with the observed portion of the run in the year being estimated. When run timing and abundances are not closely comparable, a surrogate run distribution can be rescaled to improve accuracy. The use of a surrogate run may provide more accurate reconstructions of missing counts than linear interpolations for some parts of missed passage, especially when counts over several days are missing. However, the use of data for a single year or of surrogate datasets from other years to estimate missing components of a run fails to incorporate all available information.

Several models have been used to improve estimates of escapement from run timing curves of observed counts and ad hoc estimates of fish passage. One approach is to estimate passage using likelihood or Bayesian methods for one year at a time. A Bayesian approach has an advantage over a likelihood-based approach by using prior information to produce more accurate estimates (Hilborn et al. 1999; Adkison and Su 2001) and by using Markov chain Monte Carlo (MCMC) methods to produce credible intervals around estimates of escapement. Sethi and Bradley (2016) used a Bayesian approach to estimate arrival curves of pink, chum, coho, and Chinook salmon from passage counts at weirs that had not been continuously monitored. To evaluate the accuracies of their reconstructed passage curves, simulations were made under various scenarios of missing data, including weekends off when no counts were made, missing counts in the first 15% of the run, and missing counts during the peak of the run. These simulations showed that a negative binomial model best fit the counts, because it allowed for overdispersion to accommodate patchy data resulting from groups of fish sporadically passing the weir. The negative binomial model of discrete, highly variable counts flexibly allows for overdispersion in the counts but reduces to a Poisson distribution model in the absence of overdispersion. Although the models in Sethi and Bradley (2016) improved reconstruction of a passage curve, and hence the overall estimate of escapement, these models provided run curves for only one year at a time and did not incorporate information from previous years.

The use of all available data over years is a major feature of a hierarchical Bayesian model (HBM). HBMs, together with MCMC methods, have also been used in many other contexts to improve parameter estimation from sparse datasets. HBMs have been used to address problems in such diverse disciplines as fishery dynamics (Rivot and Prévost 2002; Prévost et al. 2003), wildlife diseases (Cross et al. 2010), ecology (Wikle 2003; Aderhold et al. 2012; Bal et al. 2014), marketing (Bradlow and Rao 2000), and medicine (McCormick et al. 2012), among many other disciplines. HBMs of time-series data are even more useful than model estimates for a single year, because they extract higher level abstractions from specific observations (Cressie et al. 2009). These higher level abstractions place individual observations (daily counts) into a broader context (counts in other years) and improve the efficiency of a model to produce accurate estimates. In this case, the process giving rise to the daily counts of fish stems from an underlying run-timing curve, whose parameter values are similar among years. The hierarchical model of passage at a weir builds informative Bayesian priors for the run-timing curve parameters, so that estimates of passage in years with missing counts are informed by more complete run curves from other years.

Fig. 1. Map of Alaska showing location of counting weir on the Kogrukluk River, a tributary of the Kuskokwim River.



Our approach was similar to an HBM incorporating historical abundance priors for spawning aerial observations (Su et al. 2001). Unlike weir counts, aerial counts are confounded by stream residence times of the salmon before spawning and the possibility of missing or double counting a fish. Hence, they modeled the mean date of arrival using an exponential function (curve). Assumptions about stream residency proved to be important in estimating total escapement and its variance. Aerial estimates of pink salmon in spawning areas in Southeast Alaska for 25 years from 1974 to 1998 were analyzed with the HBM (Su et al. 2001). Many counts did not extend past, or include, the peak portion of the run curve, so that annual counts varied widely and could not be sampled efficiently without a scaling transformation of the counts.

In the present study, we examined a 40-year time series of fish passage in the Kogrukluk River with an HBM. The Alaska Department of Fish and Game (ADF&G) has operated a fish counting weir at a remote site on the Kogrukluk River since 1976, and this historical time series is the longest sequence of weir counts for several species of salmon in the Kuskokwim River drainage. The Kogrukluk River weir counts have served as an important indicator of Chinook salmon (*Oncorhynchus tshawytscha* (Walbaum in Artedi, 1792)) abundance in the Kuskokwim River, which supports substantial commercial and subsistence fisheries. The Kogrukluk River is a tributary of the Holitna River, which is the largest salmon-producing tributary in the Kuskokwim River drainage (Fig. 1; Shelden et al. 2004). The methods used to generate this dataset are typical of most weir counting procedures for Chinook salmon, as well as for other species of salmon for which weir counts are used to estimate escapement. However, this time series is not complete because of the challenges associated with operating a weir under harsh conditions in a dynamic river environment. The dataset consists of daily weir counts, with counts that were either missing or incomplete on several days.

The goal of the present study was to develop an HBM to estimate parameters from yearly fish passage curves at the Kogrukluk River weir and to estimate unobserved components of salmon passage. We created a generative HBM to estimate run-timing curves from imperfect weir counts that could be used to estimate fish passage during periods when counts are missing. A pivotal assumption of our model is that, when the weir has been breached, migrating

Table 1. Explanations of model parameters.

Parameter	Description
t	Year index
$\mathrm{Log}(\alpha_t)$	Log expected escapement
β_t	Run-protraction parameter
m_t	Mode of run
κ_t	Negative–binomial dispersion parameter
μ_α	Prior mean of α_t
σ^2_α	Prior variance of α_t
μ_β	Prior mean of β_t
σ^2_β	Prior variance of β_t
μ_m	Prior mean of m_t
σ^2_m	Prior variance of m_t
μ_κ	Prior mean of κ_t
σ^2_κ	Prior variance of κ_t

fish pass the weir site as they do during normal operating days. We then compared the results of this model with historical ad hoc estimates and with year-by-year Bayesian estimates. The HBM approach provides more defensible estimates of escapement than either of these methods, because it incorporates information from previous years and produces estimates with smaller mean squared errors (Casella 1985).

Methods

Count data

Counting procedures varied over the course of the time series. In the early years of the time series, only the total count for one day was reported by ham radio the next day, because the weir was located in a remote area of Alaska. The closest village was about 200 km downstream and the nearest road was about 600 km away. Hence, the original tallies and detailed information about the number of hours the fish were counted were unavailable. Simple daily counts were used in this study.

Model

The observed daily counts in the model are described by a probability distribution whose expectation changes deterministically over time. This characterization was then used to iteratively generate missing counts. Descriptions of parameters used in the model appear in Table 1.

Likelihood

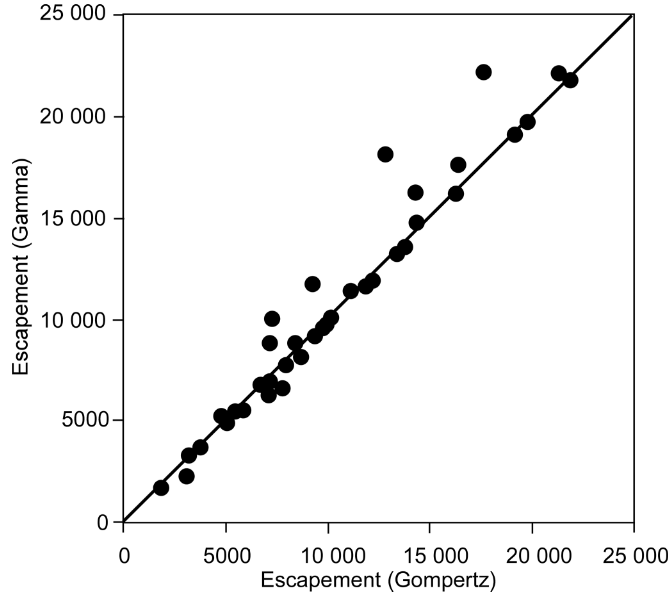
We modeled the highly variable daily fish-passage counts as a negative binomial random variable (Ross and Preece 1985) with JAGS coding following the gamma–Poisson mixture approach (Congdon 2003) and with time-varying expectation (Supplementary Table S1[1]). A negative binomial distribution rather than a normal distribution was used to model counts at a weir, because variability in passage was expected to be correlated with the magnitude of the passage (Sethi and Bradley 2016). For day $j \in \{1, 2, \ldots, J\}$ in year $t \in \{1, 2, \ldots, T\}$, the distribution of the daily count $y_{t,j}$ is

$$(1a)\qquad y_{t,j} \sim \mathrm{Poisson}(\mu_{t,j}\nu_{t,j}) I\left(y_{t,j}^{(\mathrm{partial})} \le y_{t,j} \le \infty\right)$$

$$(1b)\qquad \nu_{t,j} \sim \mathrm{Gamma}(\kappa_t, \kappa_t)$$

where $\mu_{t,j}$ is the expected count and κ_t is a dispersion parameter. We use the notation $I(y_{t,j}^{(\mathrm{partial})} \le y_{t,j} < \infty)$ to represent the support of $y_{t,j}$, where $y_{t,j}^{(\mathrm{partial})}$ is a partial count on day j in year t. Partial counts are incomplete counts when the weir is unattended or during a breach in the weir and are "right censored" providing a minimal count of passage for that day. During normal operations, complete counts for a day are "uncensored" but are set at $y_{t,j}^{(\mathrm{partial})} = 0$ for ease of coding.

Fig. 2. A comparison between estimates of escapement using the Gompertz and gamma distributions. The line represents where the estimates are equal ($y = x$).



To model the expected counts, μ_{ij}, we needed a flexible curve that captures the dynamics of expected salmon arrivals, in which the run builds quickly, peaks, and then slowly dies off, resembling a unimodal probability distribution with a heavy tail. Both the gamma and Gompertz curves potentially meet these requirements. However, a comparison of estimates of escapement with the HBM showed that the Gompertz curve consistently provided more conservative estimates (Fig. 2). The use of the gamma curve in the model may over-estimate escapement in some years and lead to inappropriate fishery management action. We therefore used a scaled version of a Gompertz curve

$$(2)\qquad \mu_{t,j} = E(y_{t,j}) = \frac{\alpha_t}{\beta_t}\exp\left(\frac{m_t - j}{\beta_t}\right)\exp\left(-\exp\left(\frac{m_t - j}{\beta_t}\right)\right)$$

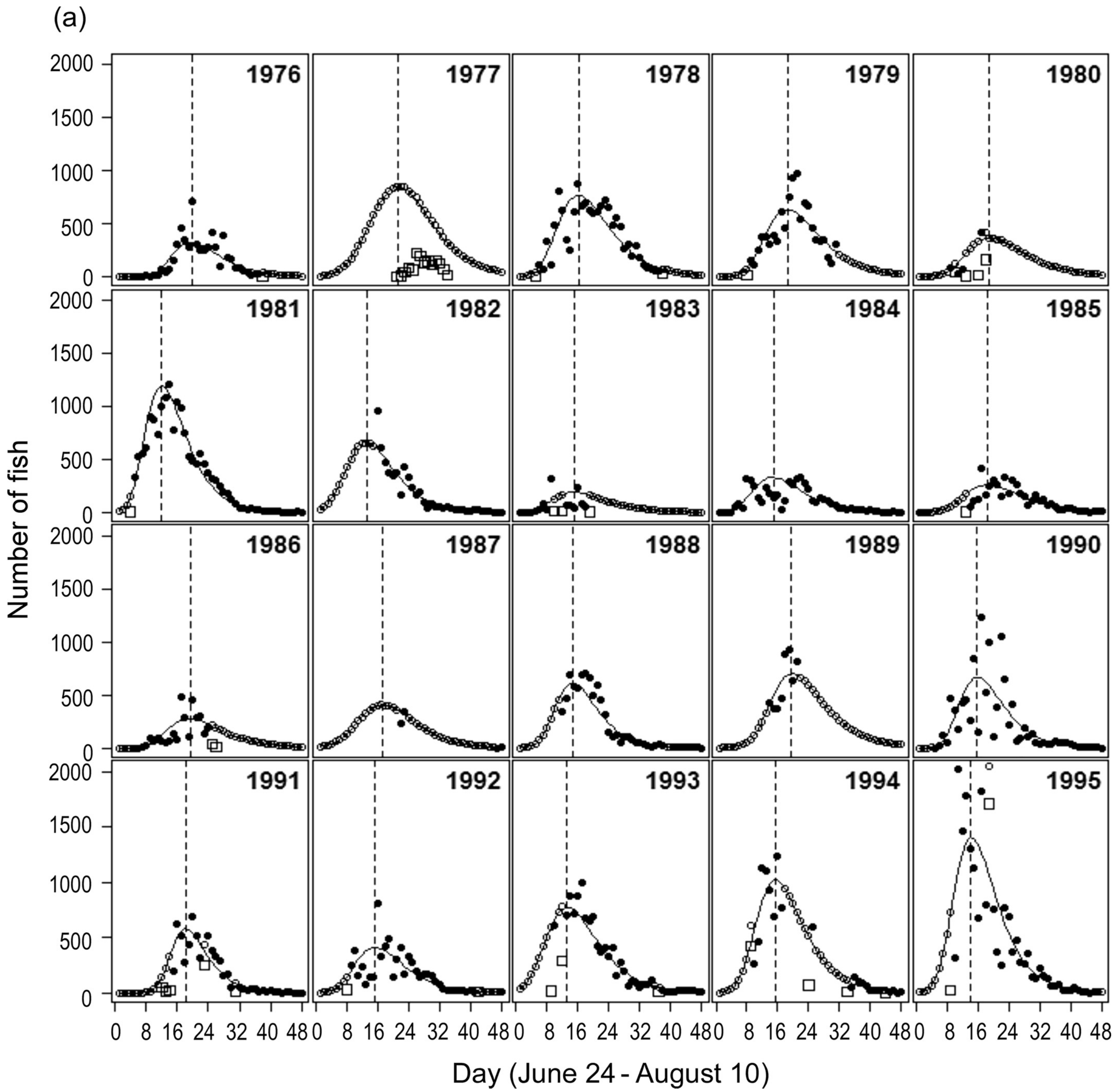
where $\mu_{t,j}$ is the expected count on day j in year t, α_t is a scalar that determines the magnitude of the curve, m_t corresponds to the mode of the run, and β_t accommodates run protraction (i.e., the right skew of the curve). We sought a function that produced a flexible unimodal curve that could incorporate variation in run timing characteristics such as protraction and mode of passage.

Bayesian priors

There are four parameters per year: $\log(\alpha_t)$, β_t, m_t, and κ_t from eqs. 1 and 2 (Table 1). We used $\log(\alpha_t)$ instead of α_t to keep this parameter positive. Each year was given a hierarchical gamma prior with its own mean and variance. This construct required eight hyperprior parameters: μ_α, σ^2_α, μ_β, σ^2_β, μ_m, σ^2_m, μ_κ, and σ^2_κ (Table 1). Each global mean μ_θ and variance σ^2_θ was given an uninformative exponential prior with a mean of 10 000, which produced a flat distribution of priors intended to have minimal effect on the results. Experimental analyses with values other than 10 000 produced similar results (results not shown). Our parameterization of the model was written in terms of biologically interpretable

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2016-0398.

Fig. 3. Daily counts (complete and partial) of Chinook salmon passage at the Kogrukluk River weir, Alaska, and associated model run-time distributions from 1976 to 2015, where (*a*) 1976–1995 and (*b*) 1996–2015. Closed circles represent observed daily counts. Open circles represent estimates of daily passage. Open squares represent partial daily counts. Vertical dashed lines indicate the peak of the run as indicated by model estimates. Solid curves represent model expectations.



means and variances to provide an intuitive understanding of the equations.

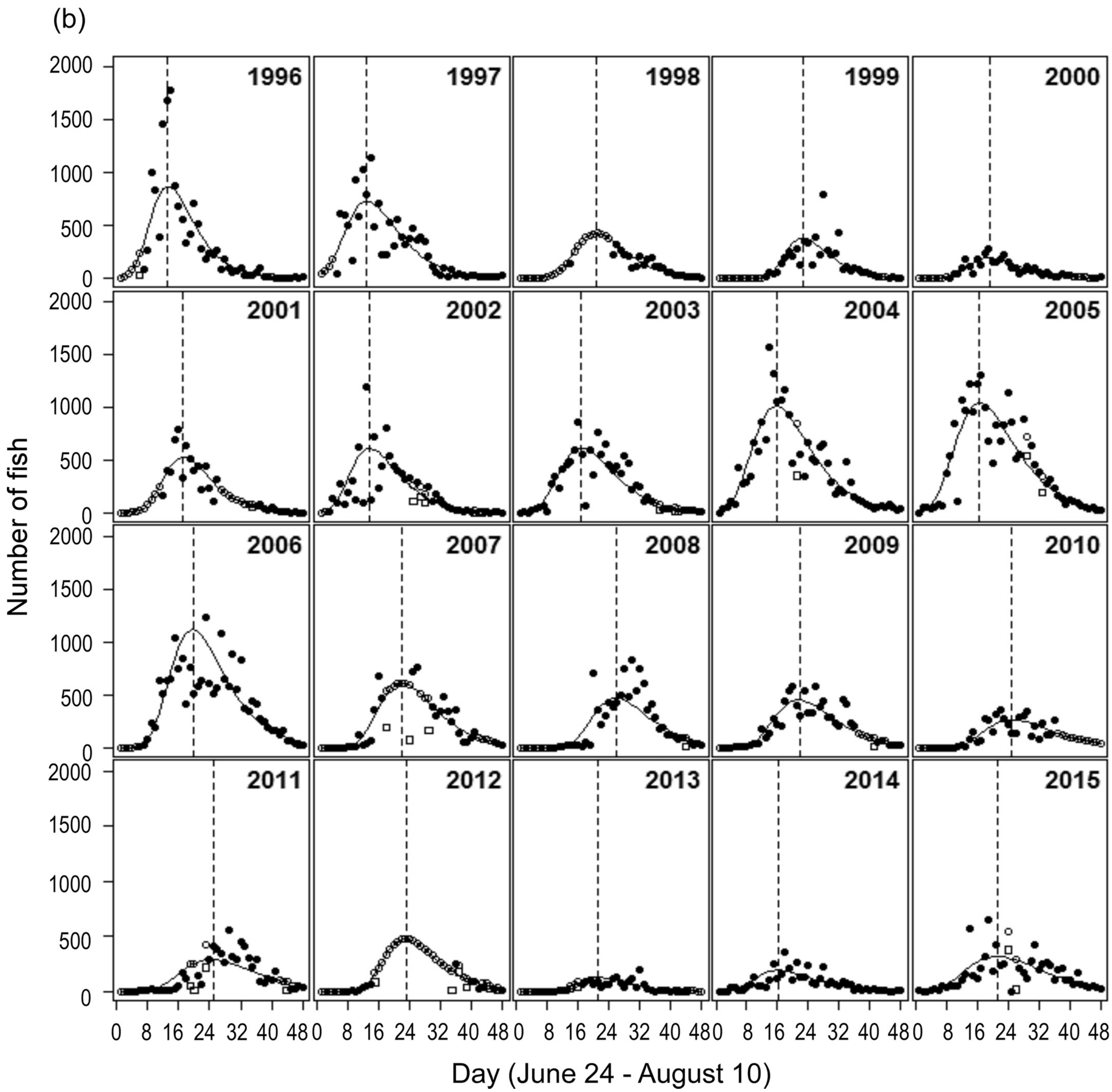
Kogrukluk River time series

We applied this model to a long-term set of counts of Chinook salmon at a weir across the Kogrukluk River. The data consisted of a 40-year time series (1976–2015) of daily fish counts made, sometimes sporadically, depending on the condition of the weir, over 48 days from June 24 to August 10 each year. While attempts were made to standardize counting, counts were not always made every day, or on the hour when the weir was operational, especially when no fish were passing the weir gate. The shapes of the curves indicate that this 48-day observation period included most, if not all, of the run in all years. Breaches in the weir occurred in all but two years (1984 and 2014) and averaged roughly 16 days among years, or about one-third of the run timing. In 1977, the weir was never fully operational so that only partial counts were made. Partial counts were used to inform the model to prevent the partial count from exceeding the model estimated count. These “right censored” observations set the minimal count for that day.

Implementation

The model was coded in JAGS language (Supplementary Table S1), and chains of 1 million iterations (thinning by 10) with different

Fig. 3 (*concluded*).



starting values were run with a burn-in of 10 000 iterations. One million iterations were used to ensure convergence. Five chains were run simultaneously and required about 24 h in a cluster computing environment. Convergence of the MCMC runs was tested using the Gelman–Rubin shrink factor (Gelman and Rubin 1992). Total escapement was estimated at each iterate as the sum of actual and generated daily counts. Point estimates for total escapement each year were taken as the posterior means, and 95% credible intervals were constructed by taking the 2.5% and 97.5% quantiles of the posterior distribution.

Evaluation of model

We used two complete annual datasets from 1984 and 2014 to evaluate the HBM. These two years included observations over all of the 48-day period the weir was in operation. The mean time that the weir was inoperable from breaches was 15.9 days, with no complete daily counts in in 1977 and complete daily counts in years 1984 and 2014. Hence, we used 16 days as a realistic period of unobserved counts in this evaluation. Accordingly, we removed counts in the first (16 days), middle (16 days), and last (16 days) parts of the run before estimating run curves for each experimental removal with the full model informed by historical data and compared these results with escapements estimated from the observed counts.

Results

The Gelman–Rubin shrink factors converged to approximately one, indicating that HBM chains estimating total escapement converged for the datasets from all 40 years. Each year, passage at the

Table 2. Posterior means and standard deviations of the 16 hyper-prior parameters for the hierarchical run timing model over the 40-year dataset.

Parameter (θ)	$\hat{\mu}_\theta^{(40)}$	$s_{\mu_\theta}^{(40)}$	$\hat{\sigma}_\theta^{2(40)}$	$s_{\sigma_\theta^2}^{(40)}$
ln α	9.09	0.10	0.38	0.10
β	6.60	0.21	1.38	0.42
m	18.12	0.68	16.65	4.79
κ	4.65	0.44	4.54	2.12

weir tended to build, reach a peak, and then attenuate until the run was exhausted in late summer (Figs. 3*a* and 3*b*). Table 2 shows the means and variances of estimates of model hyper-prior parameters, and α is the expected escapement but varies greatly across years. We therefore parameterized α as log (α) (mean ± standard deviation, 9.09 ± 0.10) to accommodate this variability and produce a more unimodal distribution. The parameter β (6.60 ± 0.21) is the mean of a measure of the protracted shape of the run-timing curves. The parameter κ (4.65 is ±0.44) is the dispersion parameter for the negative binomial measuring daily passage through the weir. The values of these parameters for individual years (Supplementary Table S2) represent draws from these higher level distributions.

The three methods of estimating escapement produced different patterns of variability (Table 3; Fig. 4). The HBM estimates of total escapement resolved peaks in 1977, 1981, 1989, 1995, and 2005 (Fig. 4, open circles). Escapement in these years exceeded 13 000 Chinook salmon but exceeded 20 000 fish in 1995 and 2005. The ad hoc estimates of escapement were similar to the HBM estimates but lacked measures of uncertainty (Fig. 4, closed circles). The year-by-year Bayesian estimates yielded run-timing curves that were ambiguous or produced large uncertainties in estimates of total escapement in several years (Fig. 4, open triangles).

Annual escapements estimated with the HBM showed positive autocorrelation, $\rho = 0.60$, with a one-year lag. The estimated peaks of the runs differed by as much as 16 days among years over the 40-year dataset. The peak in run timing tended to be earlier in years of large runs (Pearson's $r = -0.33$; Fig. 5).

As a test of the efficiency of the HBM, we removed 16-day periods of counts in the first, middle, and last third of the run for each dataset and used our HBM to estimate escapement. In both years, the credible intervals included the estimate of escapement from the full dataset (Fig. 6). Estimates of error with credible intervals were largest for the loss of counts from the first part of the run and the smallest for the loss of counts at the last third of the run in both years.

Discussion

The goal of this study was to develop a standard method with reasonable computation time to estimate escapement from sporadic counts of fish passing a weir. Weir counts in index streams are widely used, but estimates are often based on ad hoc linear or exponential regressions of sporadic or partial counts. The HBM approach used here incorporates summaries of weir passage in previous years as priors and is expected to produce more defensible estimates than ad hoc methods or year-by-year Bayesian (or maximum likelihood) estimates, because the mean squared error is expected to be smaller (Casella 1985; Su et al. 2001). We found a close association between model estimates and observed counts in years when counts were missing on some days. In 2 years of the series for which counts were made on all days, HBM estimates of periods with counts artificially removed were not significantly different from the observed counts.

Although the model represents a standard approach for estimating escapement from counts of fish passing the weir, other factors influencing weir counts can affect estimates of escapement. For example, the model assumed that fish pass a breached weir at the same rate they would have passed through an opening in an operational weir. When a weir is operational, Chinook salmon tend to congregate downstream before swimming through the counting gate (personal observation; Hansen and Blain 2013; Lisa Stuby, Alaska Department of Fish and Game, Fairbanks, AK, unpublished data, 2007). When a weir is breached, these fish move rapidly upstream as a group, and the number of fish subsequently passing the weir drops considerably the day after a breach has occurred. Hence, models estimating missing counts may underestimate passage in some instances following a major breach in the weir.

Passage can also be underestimated when minor breaches occur by erosion around the weir and provide an alternative gateway through the weir. On the other hand, passage may be overestimated when flooding hinders fish migration upstream. Despite these potential sources of error, weir counts of migration provide more accurate estimates of escapement than aerial or stream-side counts in spawning areas, because weir counts are based on observations of individual fish.

An additional source of error in the daily counts arises from procedures used to count fish passing the weir. The counting procedure at many weirs is regimented by counting hourly for a given number of minutes. However, counting times at the Kogrukluk River weir were informally set by the pattern of Chinook salmon behavior throughout the day. Fish tended to accumulate behind a weir and often passed the weir in groups only after a "leader" swam through the observation gate while it was open. Hence, because several hours can elapse without fish moving through the counting gate, counts were made when fish were actually moving through the weir. This practice adds to the sporadic nature of the count data.

In addition to providing a more consistent approach to estimating passage than ad hoc estimates, a hierarchical Bayesian approach offers an important advantage by using prior information from runs in all years with data, in addition to data for the current year, to improve estimates of passage in the current year. Ad hoc and year-by-year estimates fail to incorporate this store of information. Analyses of long-term datasets are more defensible than estimates based on a single season, because limited observations in one year are augmented by "borrowing strength" from other years (Su et al. 2001). In addition to the use of information from previous years' escapements, the model can be modified to incorporate other variables influencing run timing and run strength such as temperature and rainfall, spawning area carrying capacity (e.g., Geiger and Koenings 1991), and demographic information (e.g., Lessard et al. 2008).

The posterior distribution of a Bayesian analysis (HBM or year-by-year approaches) also provides an estimate of the precision of the total escapement, which is not possible with ad hoc estimates. Nevertheless, not all sources of error are incorporated into the model's credible intervals, including counting errors during the daily operation of a weir, whether by direct observation or by sonar. Estimates of error from our model also do not account for model error, or process error, inherent in using one approach (Cressie et al. 2009).

Another potential problem with the use of time-series data to support an HBM is a shifting baseline because of biological responses to long-term environmental change such as the Pacific Decadal Oscillation (PDO; Mantua and Hare 2002). These changes, however, can be incorporated into an HBM as an additional source of variability at a higher hierarchical level. In our case, the PDO shift in the late 1970s did not appear to be a major source of variability in the historical data set of Chinook salmon passage at the Kogrukluk River weir.

We tested the ability of our model to handle missing counts by examining the behavior of the model after removing segments of the weir counts from 2 years that had a full set of counts for the

Table 3. Comparison of ad hoc, year-by-year, and hierarchical Bayesian estimates of Chinook salmon passage at the Kogrukluk River weir.

Ad hoc estimate		Year-by-year Bayesian model estimate					Hierarchical Bayesian model estimate				
				Credible interval					Credible interval		
Year	Estimate	Estimate	Standard deviation	Lower	Upper	Shrink factor	Estimate	Standard deviation	Lower	Upper	Shrink factor
1976	5 600	5 746	92	5 616	5 975	1.00	5 744	84	5 620	5 945	1.00
1977	NA	20 175	14 447	4 561	61 805	1.46	17 678	10 573	6 285	44 962	1.00
1978	13 667	13 642	139	13 433	13 957	1.00	13 640	137	13 428	13 957	1.00
1979	11 338	12 069	746	11 194	13 884	1.03	11 820	468	11 087	12 927	1.00
1980	6 572	10 145	4 280	4 137	20 542	1.08	7 205	3 259	3 432	14 988	1.00
1981	16 809	16 235	82	16 110	16 430	1.00	16 241	109	16 091	16 508	1.00
1982	10 993	25 330	9 436	9 202	45 208	1.09	11 118	2 695	7 364	17 713	1.00
1983	3 025	3 642	5 095	1 500	22 781	1.42	3 034	1 014	1 825	5 735	1.00
1984	4 928	4 909	0	4 909	4 909	NA	4 909	0	4 909	4 909	NA
1985	4 625	4 843	246	4 476	5 427	1.01	4 915	268	4 511	5 545	1.00
1986	5 038	6 747	3 597	3 500	16 957	1.02	4 765	858	3 632	6 882	1.00
1987	4 063	10 664	7 108	2 816	29 891	1.39	7 743	3 201	3 631	15 705	1.00
1988	8 520	9 070	887	8 020	11 367	1.01	9 204	1 040	8 034	11 964	1.00
1989	11 940	30 018	12 011	10 713	53 704	1.64	12 771	2 676	8 718	18 986	1.00
1990	10 214	10 072	7	10 067	10 087	1.01	10 072	6	10 067	10 087	1.00
1991	7 850	6 978	724	6 455	8 309	1.25	7 125	651	6 477	8 753	1.00
1992	6 755	6 865	463	6 480	8 212	1.08	6 781	278	6 481	7 501	1.00
1993	12 333	19 660	7 073	12 479	40 074	1.10	14 266	1 705	11 922	18 524	1.00
1994	15 227	16 897	1 689	14 326	20 886	1.01	16 396	1 481	13 901	19 721	1.00
1995	20 651	22 069	2 589	19 702	29 095	1.26	21 267	1 404	19 610	24 855	1.00
1996	14 198	14 680	1 165	13 843	18 530	1.06	14 273	453	13 839	15 495	1.00
1997	13 285	13 609	495	13 139	15 011	1.01	13 432	258	13 125	14 090	1.00
1998	12 107	6 887	1 279	5 364	9 824	1.05	7 072	1 181	5 374	9 955	1.00
1999	5 570	5 501	339	5 441	5 527	1.30	5 473	21	5 441	5 522	1.00
2000	3 310	3 218	25	3 179	3 276	1.00	3 218	26	3 178	3 279	1.00
2001	9 298	8 376	525	7 731	9 772	1.05	8 391	574	7 706	9 846	1.00
2002	10 103	9 967	190	9 695	10 429	1.00	9 960	181	9 696	10 390	1.00
2003	11 770	11 719	80	11 589	11 903	1.00	11 717	80	11 588	11 899	1.00
2004	19 650	19 792	319	19 344	20 554	1.01	19 796	328	19 340	20 587	1.00
2005	22 000	21 840	220	21 551	22 384	1.00	21 840	216	21 550	22 371	1.00
2006	19 414	19 125	7	19 115	19 143	1.00	19 124	7	19 115	19 143	1.00
2007	13 029	12 115	1 287	10 020	14 980	1.00	12 192	1 254	10 103	14 987	1.00
2008	9 730	9 653	80	9 564	9 861	1.00	9 640	64	9 563	9 802	1.00
2009	9 702	9 251	134	9 023	9 548	1.00	9 228	138	8 993	9 535	1.00
2010	5 690	5 887	522	5 222	7 206	1.02	5 547	285	5 108	6 219	1.00
2011	NA	7 106	331	6 602	7 904	1.00	7 157	381	6 606	8 081	1.00
2012	NA	8 428	2 707	4 909	14 735	1.09	8 641	2 093	5 481	13 643	1.00
2013	NA	1 821	75	1 721	2 010	1.00	1 836	84	1 727	2 044	1.00
2014	NA	3 705	0	3 705	3 705	NA	3 705	0	3 705	3 705	NA
2015	NA	7 865	246	7 529	8 464	1.00	7 894	257	7 541	8 522	1.00

Note: NA, not available.

48 days the weir was in operation. All of the model estimates of passage for periods with missing data had credible intervals that included the estimates of escapement in the model (Fig. 6). This simulation indicated that the hierarchical Bayesian approach provided robust estimates of escapement when some counts were missing. However, it was not possible to compare the results of either the ad hoc, year-by-year analyses or the HBM estimates with a golden standard to determine accuracy. Although we cannot say that the one approach produces more accurate results than the other, we can say that HMB estimates are statistically more defensible, because HBM estimates are expected to have smaller mean squared errors than ad hoc or year-by-year estimates.

In previous models, run-timing variability among years was estimated for some salmon runs but interannual cycles were not well resolved (Heard 1991; Su et al. 2001). The results of our model showed that run-timing peaks for Chinook salmon returning to the Kogrukluk River differed by as much as 16 days. These differences in run-timing peaks were not random among years but showed autocorrelation with a 1-year lag (ρ = 0.60). For example, run peaks occurred earlier in the season as run strength gradually built up in the late 1970s and peaked in 1981, before a gradual decline. Abundances also peaked in 1990, 1995, and 2005.

The model results also showed that the timing of peak passage through the weir was negatively correlated with total run size (r = –0.33; Fig. 5). The peaks of runs in years when abundances were large tended to occur earlier in the season. For example, run peaks during high abundances in 2004 and 2005 occurred on days 14–16, but run peaks in subsequent years gradually shifted to day 21 then to day 26 by 2010 when runs were weaker. In concordance with our model results for Kogrukluk River Chinook salmon, early run timing occurred with large returns, and late run timing was associated with small runs of Chinook salmon in the Columbia River (Keefer et al. 2008) and of sockeye salmon in rivers around Bristol Bay (Adkison and Cunningham 2015).

The results of our study point to several areas for future research. One is to better understand the behavior of fish passage past a weir during normal operation and during breaches in the weir. Our HBM does not account for fish accumulating behind a weir and passing en masse during a breach. Some method is needed to account for the number of fish behind the weir during

Fig. 4. Yearly ad hoc (closed circles), year-by-year Bayesian (open triangles), and hierarchical Bayesian (open circles) estimates of escapement of Chinook salmon in the Kogrukluk River from 1976 to 2015. Error bars represent 95% credible intervals. No ad hoc estimates were made from 2011 to 2015, when HBM estimates were used exclusively.

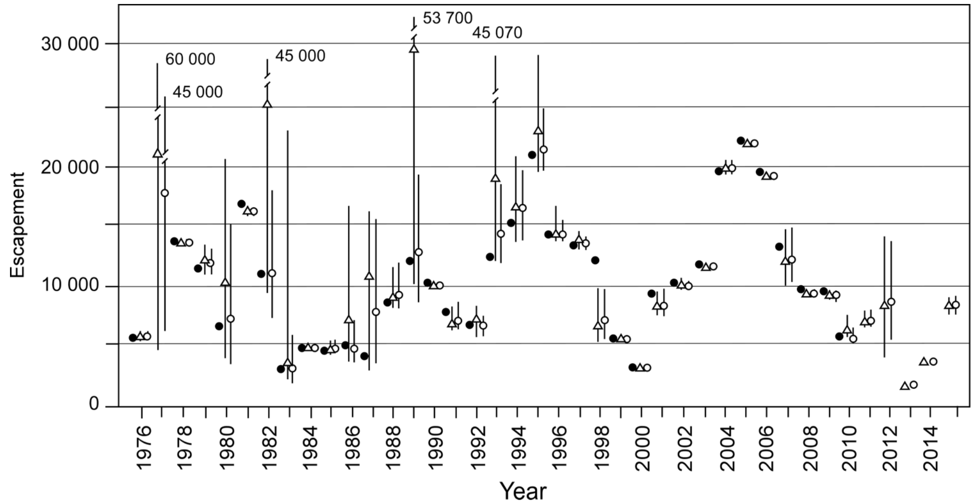


Fig. 5. Relationship between day of peak of passage counts beginning 24 June each year at the Kogrukluk River weir and estimate of total escapement. Correlation coefficient was $r = -0.33$.

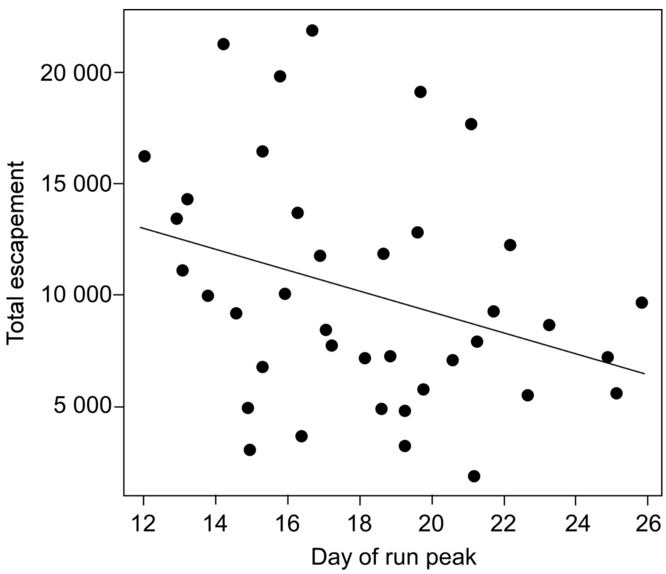
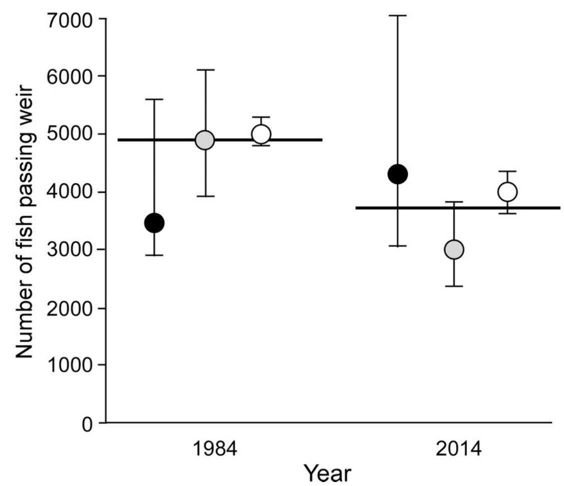


Fig. 6. Estimates of escapement at the Kogrukluk River weir for two datasets with complete 48-day counts in 1984 and 2014 (horizontal lines) based on the hierarchical Bayesian model of the 40-year dataset from 1976 to 2015. Circles represent estimates for the years 1984 and 2014, in which early (16 days, closed circle), peak (16 days, grey circle), and late (16 days, open circle) portions of the runs were deleted to simulate unobserved passage through the weir. Vertical lines represent 95% credible intervals.



a breach in the weir. Another potential problem is to assess model uncertainty, which can be addressed with various approaches to model selection. Another avenue for future research might be further comparisons of year-by-year and HBM approaches. Finally, the results of this time series point to the problem of understanding what environmental variables control annual run strength and timing in the Kuskokwim River system.

In conclusion, the HBM implemented here provides a standard, more defensible approach over ad hoc and year-by-year procedures for estimating missing fish passage counts. Missing passage can be estimated in a standardized way, eliminating ad hoc and often subjective applications of various interpolation methods. The accuracies of estimates of missing counts greatly influence the accuracies of estimates of overall escapement in a particular year. Model estimates of missing counts are also expected to be more accurate than ad hoc methods or year-by-year analyses, because a hierarchical model incorporates prior information from all years, in addition to counts for the current year. The resolution of run timing and abundance also provides insights into factors influencing population abundances. Importantly for management, Bayesian methods provide estimates of precision, which are lacking in ad hoc methods, but which are important for identifying uncertainties in recommendations for fishery management. Our HBM approach has been adopted by several fishery management agencies to estimate escapement.

Acknowledgements

Milo Adkison, Jan Conitz, Chris Habicht, Andrew Munro provided insightful comments on various drafts of the manuscript. We extend special thanks to Allen Bingham (retired from Alaska Department of Fish and Game) for his interest and support of this project.

References

Aderhold, A., Husmeier, D., Lennon, J.J., Beale, C.M., and Smith, V.A. 2012. Hierarchical Bayesian models in ecology: reconstructing species interaction networks from non-homogeneous species abundance data. Ecol. Inform. **11**: 55–64. doi:10.1016/j.ecoinf.2012.05.002.

Adkison, M.D., and Cunningham, C.J. 2015. The effects of salmon abundance and run timing on the performance of management by emergency order. Can. J. Fish. Aquat. Sci. **72**(10): 1518–1526. doi:10.1139/cjfas-2015-0174.

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**(8): 1663–1671. doi:10.1139/f01-100.

Bal, G., Rivot, E., Baglinière, J.-L., White, J., and Prévost, E. 2014. A hierarchical Bayesian model to quantify uncertainty of stream water temperature forecasts. PLoS ONE, **9**: e115659. doi:10.1371/journal.pone.0115659. PMID:25541732.

Bradlow, E.T., and Rao, V.R. 2000. A hierarchical Bayes model for assortment choice. J. Market. Res. **37**: 259–268. doi:10.1509/jmkr.37.2.259.18733.

Casella, G. 1985. An introduction to empirical Bayes data analysis. Am. Stat. **39**: 83–87. doi:10.1080/00031305.1985.10479400.

Congdon, P. 2003. Applied Bayesian modelling. Wiley, West Sussex.

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. 1108. 122 p.

Cressie, N., Calder, C.A., Clark, J.S., Ver Hoef, J.M., and Wikle, C.K. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecol. Appl. **19**: 553–570. doi:10.1890/07-0744.1. PMID:19425416.

Cross, P.C., Heisey, D.M., Scurlock, B.M., Edwards, W.H., Ebinger, M.R., and Brennan, A. 2010. Mapping brucellosis increases relative to elk density using hierarchical Bayesian models. PLoS ONE, **5**: e10322. doi:10.1371/journal.pone.0010322. PMID:20428240.

Eggers, D.M. 1994. On the discrimination of sockeye and Chinook salmon in the Kenai River based on target strength determined with 420 kHz dual-beam sonar. Alaska Fish. Res. Bull. **1**: 129–139.

Geiger, H.J., and Koenings, J.P. 1991. Escapement goals for sockeye salmon with informative prior probabilities based on habitat considerations. Fish. Res. **11**: 239–256. doi:10.1016/0165-7836(91)90004-Y.

Gelman, A., and Rubin, D. 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. **7**: 457–472. doi:10.1214/ss/1177011136.

Hansen, T.R., and Blain, B.J. 2013. Kogrukluk River salmon studies, 2011. Alaska Department of Fish and Game, Fishery Data Series, No. 13-13, Anchorage, Alaska.

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**(5): 888–896. doi:10.1139/f99-013.

Irvine, J.R., Bocking, R.C., English, K.K., and Labelle, M. 1992. Estimating coho salmon (*Oncorhynchus kisutch*) spawning escapements by conducting visual surveys in areas selected using stratified random and stratified index sampling designs. Can. J. Fish. Aquat. Sci. **49**(10): 1972–1981. doi:10.1139/f92-219.

Keefer, M.L., Peery, C.A., and Caudill, C.C. 2008. Migration timing of Columbia River spring Chinook salmon: effects of temperature, river discharge, and ocean environment. Trans. Am. Fish. Soc. **137**(4): 1120–1133. doi:10.1577/T07-008.1.

Labelle, M. 1994. A likelihood method for estimating Pacific salmon escapement based on fence counts and mark-recapture data. Can. J. Fish. Aquat. Sci. **51**(3): 552–566. doi:10.1139/f94-058.

Lessard, R.B., Hilborn, R., and Chasco, B.E. 2008. Escapement goal analysis and stock reconstruction of sockeye salmon populations (*Oncorhynchus nerka*) using life-history models. Can. J. Fish. Aquat. Sci. **65**(10): 2269–2278. doi:10.1139/F08-130.

Manske, M., and Schwarz, C.J. 2000. Estimates of stream residence time and escapement based on capture-recapture data. Can. J. Fish. Aquat. Sci. **57**(1): 241–246. doi:10.1139/f99-203.

Mantua, N.J., and Hare, S.R. 2002. The Pacific Decadal Oscillation. J. Oceanogr. **58**: 35–44. doi:10.1023/A:1015820616384.

McCormick, T.H., Rudin, C., and Madigan, D. 2012. Bayesian hierarchical rule modeling for predicting medical conditions. Ann. Appl. Stat. **6**: 652–668. doi:10.1214/11-AOAS522.

Prévost, E., Paerent, E., Crozier, W., Davidson, I., Dumas, J., Gudbergsson, G., Hindar, K., McGinnity, P., McLean, J., and Sættem, L.M. 2003. Setting biological reference points for Atlantic salmon stocks: transfer of information from data-rich to sparse-data situations by Bayesian hierarchical modelling. ICES J. Mar. Sci. **60**: 1177–1193. doi:10.1016/j.icesjms.2003.08.001.

Rivot, E., and Prévost, E. 2002. Hierarchical Bayesian analysis of capture-mark-recapture data. Can. J. Fish. Aquat. Sci. **59**(11): 1768–1784. doi:10.1139/f02-145.

Ross, G.J., and Preece, D.A. 1985. The negative binomial distribution. J. R. Stat. Soc. Ser. D Stat. **34**: 323–335. doi:10.2307/2987659.

Schwarz, C.J., Bailey, R.E., Irvine, J.R., and Dalziel, F.C. 1993. Estimating salmon spawning escapement using capture-recapture methods. Can. J. Fish. Aquat. Sci. **50**(6): 1181–1197. doi:10.1139/f93-135.

Sethi, S.A., and Bradley, C. 2016. Statistical arrival models to estimate missed passage counts at fish weirs. Can. J. Fish. Aquat. Sci. **73**(8): 1251–1260. doi:10.1139/cjfas-2015-0318.

Shelden, C.A., Gilk, S.E., and Molyneaux, D.B. 2004. Kogrukluk River weir salmon studies, 2003. Reg. Inform. Rep. No. 3A-4-22, 141 p.

Su, Z., Adkison, M.D., and Van Alen, B.W. 2001. A hierarchical Bayesian model for estimating historical salmon escapement and escapement timing. Can. J. Fish. Aquat. Sci. **58**(8): 1648–1662. doi:10.1139/f01-099.

Wikle, C.K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology, **84**: 1382–1394. doi:10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2.