Smoothing Population Size Estimates for Time-Stratified Mark–Recapture Experiments Using Bayesian P-Splines

Simon J. Bonner¹*, and Carl J. Schwarz²**

¹Department of Statistics, University of Kentucky, Lexington, Kentucky 40506, U.S.A.
²Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

*email: simon.bonner@uky.edu
**email: cschwarz@stat.sfu.ca

SUMMARY. Petersen-type mark–recapture experiments are often used to estimate the number of fish or other animals in a population moving along a set migration route. A first sample of individuals is captured at one location, marked, and returned to the population. A second sample is then captured farther along the route, and inferences are derived from the numbers of marked and unmarked fish found in this second sample. Data from such experiments are often stratified by time (day or week) to allow for possible changes in the capture probabilities, and previous methods of analysis fail to take advantage of the temporal relationships in the stratified data. We present a Bayesian, semiparametric method that explicitly models the expected number of fish in each stratum as a smooth function of time. Results from the analysis of historical data from the migration of young Atlantic salmon (Salmo salar) along the Conne River, Newfoundland, and from a simulation study indicate that the new method provides more precise estimates of the population size and more accurate estimates of uncertainty than the currently available methods.

KEY WORDS: Atlantic salmon; Bayesian inference; Hierarchical model; Mark–recapture; OpenBUGS; Penalized spline; Stratified Petersen.

1. Introduction

Single-recapture, Petersen-type mark–recapture experiments are commonly used to monitor populations of Atlantic and Pacific salmon (Salmo salar and Onchorhynchus sp.) as they migrate between their freshwater spawning grounds and the oceans where they spend their adult lives. In these experiments, samples of fish are captured at one location, marked, and returned to the population, and new samples containing both marked and unmarked fish are captured from a second location farther along the migration route. Estimates of abundance are then derived by modeling the numbers of marked and unmarked fish captured at the second location. If all fish have the same probability of being captured at the second location then the simple Petersen estimator computed from the overall numbers of marked and unmarked fish captured at the second location provides a valid estimate of population size. However, salmon migrations can last for weeks or months and the probability of capture can change considerably over this time. As a result, simple Petersen estimates of the population size, and the degree of uncertainty in these estimates, may be severely biased (see Seber, 2002, p. 85).

One way to allow for such variations in catchability is to stratify the population by time—essentially, to divide the experiment into shorter time strata (days or weeks) and then to estimate the number of individuals passing the second location separately for each stratum. Darroch (1961) provided the first rigorous treatment of the stratified-Petersen model and derived maximum likelihood estimators by conditioning on the numbers of individuals marked and released in each stratum at the first location. Similar methods were developed by Macdonald and Smith (1980) for the modified experiment in which trapping occurs at only one location and marked fish are introduced by transporting them upstream to be released. Plante, Rivest, and Tremblay (1998) developed unconditional likelihood methods by modeling the number of individuals at the first location, and further allowed the capture probabilities to depend on stratum-specific covariates like rate of flow. Estimators of abundance from the stratified-Petersen model have also been obtained by the methods of moments (Chapman and Junge, 1956) and by least squares (Banneheka, Routledge, and Schwarz, 1997).

Although stratification reduces the bias of the abundance estimates, it also increases the number of parameters, which leads to a bias/variance trade-off. In general, marked individuals may pass the second location in any stratum so that it is necessary to model the movements of these fish between the trapping locations. The number of parameters increases as the product of the number of strata at the two locations, and if few marked fish are recaptured then the movements are difficult to model and the resulting estimates of population size will be imprecise.

The number of parameters can be reduced by partially pooling the data, but there are several pitfalls. Proposed methods for pooling strata essentially entail testing for differences in the capture probabilities between neighboring strata and then combining these strata if the null hypothesis is not rejected (see Darroch, 1961; Schwarz and Taylor, 1998; Bjorkstedt, 2000). However, such tests will have low power, because
the number of fish marked in each strata is small, and important differences are likely to be overlooked. Guidelines as to how far the data should be pooled are not available, and, moreover, standard errors and confidence intervals computed from these methods do not account for the pooling decisions and will underestimate the true variability in the estimated population size (Steinhorst et al., 2004).

Schwarz and Dempson (1994) introduced another method to reduce the number of parameters by assigning parametric distributions to the travel times of the marked individuals. Their particular formulation assumes that travel times between the trapping locations follow independent log-normal distributions with mean and variance parameters that depend only on the release strata. Probabilities governing the movement of the marked fish can then be computed from standard normal calculations, and the number of parameters in the model can be greatly reduced relative to the model of Darroch (1961), which allows complete flexibility in the movements between strata. Mantyniemi and Romakkaniemi (2002) extended this model to the hierarchical Bayesian framework by modeling the capture probabilities, the means and variances of the travel times, and the numbers of unmarked fish in each stratum as random draws from hyperpriors with unknown parameters. This allows for sharing of information between the strata through estimation of the hyperparameters and further decreases the uncertainty in total abundance.

One shortcoming of all of these methods is that they fail to account for the natural, temporal ordering of the data. Most salmon runs follow a fairly predictable pattern with small numbers of fish passing the traps early in the migration followed by a steady increase to one or two peaks near the middle of the migration and then a slow decrease toward the end. Accounting for these patterns has the potential to improve estimates of the total population size, and our primary contribution is the implementation of a Bayesian, semiparametric method that takes advantage of the temporal ordering by explicitly modeling the expected number of unmarked fish in each stratum as a smooth function of time. Further to this, we develop a hierarchical, nonparametric model for the movements of the marked fish that reduces the number of parameters by building on similarities in the movements between strata rather than by assuming a specific parametric form. In the following sections, we develop the general model, describe methods for assessing goodness of fit and comparing candidate models, and provide results from an application to historical data and a simulation study that illustrate the primary advantages of our method.

2. Methods
We describe the implementation of the Bayesian P-spline model with specific application to programs monitoring the productivity of salmon populations, though the methods are generally applicable to temporally stratified data. Most species of salmon are anadromous, meaning that they are born in freshwater lakes and rivers but spend the majority of their adult lives in the ocean before returning to freshwater to spawn. Every year, the young fish born 1 or 2 years earlier (called smolts) leave the spawning grounds, and the productivity of the population can be quantified by the total number of out-migrating smolts. To estimate the size of the migration (also called the run), fish are trapped at one location, marked, and released back into the population. Samples containing both marked and unmarked individuals are then trapped at a second location farther along the migration route, and the size of the population is estimated by modeling the capture of marked and unmarked individuals at the second trap. As discussed in Section 1, these runs may last for weeks or months and the data are often stratified to account for changes in the capture probabilities over this time. We assume that the data are stratified by day for notational convenience, but the methods apply equally if the data are grouped in other ways.

Data obtained from these experiments consist of the daily numbers of marked fish released at the first trap, and the daily numbers of both marked and unmarked fish captured at the second trap. The two traps need not be operated for the same length of time, and it is common to continue trapping at the second location for several days after the last marked fish are released to allow time for these individuals to move past the second trap. We denote the number of days of capture at the first and second traps by s and t, the number of marked fish released on day i at the first trap by \( n_i \), the number of these fish recaptured on day j by \( m_{i,j} \), and the number of unmarked fish captured at the second trap on day j by \( u_j \). The total number of unmarked fish that pass the second trap on day j is denoted by \( U_j \), the sum over all days by \( U_{\text{tot}} \), and the total population size by \( N_{\text{tot}} = \sum_{i=1}^{s} n_i + U_{\text{tot}} \). Vectors and matrices are used to simplify the notation where possible such that \( \mathbf{n} \) represents the vector \((n_1, \ldots, n_s)\), \( \mathbf{u} \) the vector \((u_1, \ldots, u_t)\), and \( \mathbf{M} \) the matrix with ij entry \( m_{i,j} \).

Common assumptions for modeling such data essentially imply that the samples of marked and unmarked fish captured at the second trap constitute simple, random samples from the populations available on each day. Specifically, it is assumed that:

1. All fish passing the second trap on day j have the same probability of capture, \( p_j \), and
2. Whether or not any one fish is captured at the second trap is independent of whether or not any other fish is captured.

Further to this, it is necessary to model the movement of the marked fish between the two traps. Standard assumptions concerning the movements of marked fish are that:

1. They do not emigrate or die between the two traps,
2. Marks are not lost or overlooked when fish are recaptured,
3. All marked fish released at the first trap on day i have the same probability of passing the second trap on day j, \( \theta_{i,j} \), and
4. The movements of marked fish are independent of one another.

Given these assumptions, the recaptures of fish marked on day i will follow the multinomial distribution:

\[
\mathbf{m}_i \sim \text{MN} \left\{ n_i, \left( \theta_{i,1}p_1, \ldots, \theta_{i,t}p_t, 1 - \sum_{j=1}^{t} \theta_{i,j}p_j \right) \right\}, \tag{1}
\]

- **Biometrics**
where \( \mathbf{m} = (m_{i,1}, \ldots, m_{i,t}, n_i - m_{i,t}) \) and \( m_{i,t} = \sum_{j=1}^{t} m_{i,j} \). Conditioning on \( \mathbf{U} \), as in the original formulation of Darroch (1961), the number of unmarked fish captured at the second trap on day \( j \) will be distributed according to the binomial model:

\[
u_j \sim \text{Binomial}(U_j, p_j).
\]

The likelihood function is computed by multiplying the contributions from (1) and (2) over all days and is proportional to:

\[
L(\mathbf{U}, \mathbf{p} | \mathbf{n}, \mathbf{M}, \mathbf{u}) = \prod_{j=1}^{t} \left\{ \prod_{j=1}^{t} \left( \theta_{i,j} p_j \right)^{y_{i,j}} \left[ 1 - \sum_{j=1}^{t} \theta_{i,j} p_j \right]^{n_{i,j} - y_{i,j}} \right\} \times \prod_{j=1}^{t} \left( \frac{U_j}{u_j} \right)^{y_{j}} (1 - p_j)^{U_j - y_{j}}.
\]

A complete Bayesian formulation is then obtained by specifying prior distributions for the parameters \( \mathbf{U} \), \( \mathbf{p} \), and \( \Theta \). In their formulation, Mantyniemi and Romakkaniemi (2002) (hereafter MR02) define the prior for \( \mathbf{U} \) by first assigning \( U_{tot} \) a vague prior distribution and then modeling \( \mathbf{U} \) conditional on \( U_{tot} \). Specifically, they define the improper Jeffrey’s prior \( \pi(U_{tot}) \propto 1/U_{tot} \) and model the individual elements of \( \mathbf{U} \) by the multinomial:

\[
\mathbf{U} \mid U_{tot}, \mathbf{p} \sim \text{MN}(U_{tot}, (p_1, \ldots, p_t)).
\]

The cell probabilities, \( \mathbf{p} \), are then assigned a Dirichlet prior with fixed parameter \( \alpha \).

In theory, individual elements of \( \alpha \) could be chosen to encode prior information about the elements of \( \mathbf{U} \), but this would require the user to provide explicit details regarding the exact shape of the run. The simplest prior, with \( \alpha_j = \alpha \) for all \( j = 1, \ldots, t \), implies that the marginal prior distribution of \( U_j \) is the same for all \( j \) and ignores the ordering of the strata in that the joint prior distribution of \( (U_{j1}, U_{j2}) \) is the same for all \( j_1, j_2 \). If the run was expected to peak on day \( k \) then one could increase the values of \( \alpha_j \) in some neighborhood of \( k \), but this would require specific prior knowledge about when the peak was expected and how high it would be relative to the other days. Instead, we sought a prior that would incorporate the natural ordering of the data without the need for such specific information.

To accomplish this, we construct an alternative prior for \( \mathbf{U} \) which directly models the number of unmarked individuals passing the second trap each day. In particular, we model the expected log daily run size, \( E(\log(U_j)) \), as a smooth function of \( j \) using the Bayesian penalized spline or P-spline model of Lang and Brezger (2004). Briefly, splines are semiparametric regression functions formed as linear combinations of sets of basis functions associated with a set of knot points. The shape of a spline is very flexible and to avoid overfitting the data the complexity of the spline is controlled by two factors: the number of knot points and their locations, and the relationship among the regression coefficients. The Bayesian P-spline algorithm uses the cubic B-spline basis functions (often chosen for their computational simplicity) and addresses overfitting by fixing a relatively large number of knots but specifying a prior which favors small changes in the coefficients.

Letting \( B_1(\cdot), \ldots, B_K(\cdot) \) denote the cubic B-spline functions for the chosen set of knots, we model \( \mathbf{U} \) as:

\[
\log(U_j) = \sum_{k=1}^{K} b_k B_k(j) + \epsilon_j,
\]

where the daily deviations from the spline, \( \epsilon_j \), are assumed to be independent, mean zero normal random variables with constant variance, \( \sigma^2_\epsilon \). Following the recommendations of Lang and Brezger (2004), we space the knot points evenly across the data at intervals of 4 days (or as close to this as \( t \) allows) and model the regression coefficients by the second difference prior:

\[
b_{k+1}|b_1, \ldots, b_k \sim N(b_k + (b_k - b_{k-1}), \tau^2_k),
\]

for \( k = 3, \ldots, K \) with \( b_1 \) and \( b_2 \) assigned the improper flat prior, \( \mathcal{U}(-\infty, \infty) \). The parameter \( \tau^2_k \) controls the smoothness of the spline and is assigned an inverse gamma prior distribution with parameters \( \alpha_\tau \) and \( \beta_\tau \) chosen so that \( E(\tau^2_k) \) is small, favoring small differences in the \( b_k \) and hence a smooth fit, but \( \text{Var}(\tau^2_k) \) is large. Lang and Brezger (2004) recommend setting \( \alpha_\tau = 1 \) and choosing a small value for \( \beta_\tau \), but note that there is no common choice for all models. We found that the values \( \alpha_\tau = 1 \) and \( \beta_\tau = 0.05 \) worked well in our application.

Concerning the capture probabilities, the core model of MR02 defines a hierarchical prior for \( \mathbf{p} \), which ignores possible temporal associations, and we have adopted exactly the same prior. Specifically, the prior distribution for \( \mathbf{p} \) assumes that:

\[
\text{logit}(p_j) \sim N(\eta_p, \tau^2_p),
\]

independent but with common mean, \( \eta_p \), and variance, \( \tau^2_p \), assigned diffuse normal and inverse gamma priors, respectively. As an extension, MR02 do allow for the effect of covariates such that \( E(\text{logit}(p_j)) = X_j'\beta \). The same approach can easily be incorporated in our model, but we have ignored such effects to keep the models simple.

Finally, we consider two possible priors for the matrix of movement probabilities. The first is exactly that of MR02, which extends the log-normal model of Schwarz and Dempson (1994) (hereafter SD94) to the hierarchical Bayesian framework and the second provides a nonparametric alternative. SD94 proposed to model the times that marked fish take to move between the two traps as independent log-normal random variables with stratum-specific parameters. Denoting the mean and variance of the log travel times for fish marked on day \( i \) by \( \mu^{(LN)}_i \) and \( \sigma^{(LN)}_i \), the probability that a fish marked on day \( i \) passes the second trap on day \( i + k \) is:

\[
\theta_{i,i+k} = \begin{cases} 
0 & k < 0 \\
\Phi \left( \frac{\log(1 + d) - \mu^{(LN)}_i}{\sigma^{(LN)}_i} \right) & k = 0 \\
\Phi \left( \frac{\log(k + 1 + d) - \mu^{(LN)}_i}{\sigma^{(LN)}_i} \right) & k > 0,
\end{cases}
\]
where $\Phi(\cdot)$ is the standard-normal cumulative distribution function and $d$ corrects for the continuous release of marked fish (see p. 100 of SD94 for details). MR02 extend this to the hierarchical Bayesian framework by assigning $\mu_i^{(LN)}$ and $\sigma_i^{(LN)}$ common normal and inverse gamma priors for all $i$. We call this the log-normal prior for $\Theta$ and use the abbreviation LN to identify parameters specific to this prior.

To provide more flexibility in modeling the movement probabilities, we introduce an alternative prior based on the continuation ratio for ordinal response variables (see, e.g., Dobson, 2002, section 8.4.4). Assuming the existence of a known continuation ratio for ordinal response variables, we introduce an alternative prior based on the continuation ratio for ordinal response variables (see, e.g., Dobson, 2002, section 8.4.4). Assuming the existence of a known continuation ratio for ordinal response variables (see, e.g., Dobson, 2002, section 8.4.4). Assuming the existence of a known continuation ratio for ordinal response variables (see, e.g., Dobson, 2002, section 8.4.4). Assuming the existence of a known continuation ratio for ordinal response variables (see, e.g., Dobson, 2002, section 8.4.4). Assuming the existence of a known continuation ratio for ordinal response variables (see, e.g., Dobson, 2002, section 8.4.4). Assuming the existence of a known continuation ratio for ordinal response variables (see, e.g., Dobson, 2002, section 8.4.4).

The models Simple/Hier/NP and Hier/Hier/NP were almost equivalent to the model of MR02 with different choices of $p$. As with most complex Bayesian models, the posterior distributions are not tractable and inferences were obtained via MCMC simulation implemented in OpenBUGS (Thomas et al., 2006). Five parallel chains starting from diffuse initial values were run for each model and convergence was monitored via the Gelman–Rubin–Brooks diagnostics for both the mean and upper 97.5% quantile of each parameter (Brooks and Gelman, 1998). Each chain was run for a total of 500,000 iterations, the first 100,000 iterations were discarded, and the remaining iterations were thinned by a factor of 50 to save storage space and reduce autocorrelations, producing a final sample size of 8000 values from each of the five chains.

Figure 1. Directed acyclic graph illustrating the structure of the Bayesian P-spline model with NP travel times (P-Spline/Hier/NP), the best-fitting model for the subsampled Conne River data. Ellipses represent nodes that are stochastic and diamonds nodes that are deterministic conditional on their parents. The rectangular frames represent structures that are repeated over the days of trapping at either the first or second trap locations. Nodes outside of these frames do not repeat.

are Bayesian equivalents of the classical pooled-Petersen estimators which ignore changes in $p$ and which are commonly used when recapture probabilities are low and data are sparse (Seber, 2002, p. 60).

The complete structure of model P-spline/Hier/NP (the best-fitting model in Section 3) is illustrated by a directed acyclic graph in Figure 1 and directed acyclic graphs for the remaining models are provided in Web Figures 9–18. Prior distributions at the highest level of each model were chosen to be weakly informative about the respective parameters. For example, the prior assigned to $\eta_p$ in equation (6) was chosen such that the prior median was close to 0.1 but the 5th and 95th percentiles covered a broad range from 0.0 to 0.5.
Comparisons between the candidate models were assessed with the deviance information criterion (DIC; Spiegelhalter et al., 2002). Denoting the deviance by:

\[ D(U, p; \Theta | n, M, u) = -2 \log(L(U, p; \Theta | n, M, u)) \]

the DIC for any model is:

\[ \text{DIC} = D(\hat{U}, \hat{p}; \hat{\Theta} | n, M, u) + 2p_D, \]

where \( \hat{U} \), \( \hat{p} \), and \( \hat{\Theta} \) are point estimates (taken to be the posterior means) and \( p_D \) is a measure of the effective number of parameters computed as the mean value of the deviance over all MCMC iterations minus the value of the deviance at the posterior means (see Spiegelhalter et al., 2002, p. 591) for further details). In essence, the DIC provides a Bayesian analog of the AIC that is easily computed for hierarchical models for which the number of parameters is poorly defined.

Fit of the models was assessed by computing posterior predictive p-values (also called Bayesian p-values) (Gelman and Meng, 1996; Gelman et al., 2003, pp. 157–177). For a given discrepancy measure, \( DM(\Omega, X) \) with \( \Omega \) representing the complete set of parameters and \( X \) the observed data, the Bayesian p-value is the probability that new data, \( X' \), generated from the posterior predictive distribution produces a greater discrepancy than the observed data

\[ (p = P(DM(\Omega, X') > DM(\Omega, X) | X)). \]

Values of \( p \) close to 0 or 1 indicate that the observed data are unlikely under the assumed model. Different discrepancy measures can be chosen to assess different components of the model, and we have chosen three specific measures:

\[ DM_1(\Omega, X) = -2 \cdot L(U, p, \Theta | n, M, u) \]

providing an overall assessment of the model:

\[ DM_2(\Omega, X) = \sum_{j=1}^{t} \left( \sum_{i=1}^{n} m_{i,j} - \sum_{i=1}^{n} n_i \hat{\theta}_{i,j} \hat{p}_j \right)^2 \]

comparing the observed and expected number of marked fish recaptured in each strata, and:

\[ DM_3(\Omega, X) = \sum_{i=1}^{n} \sum_{j=1}^{t} \left( \sqrt{m_{i,j}} - \sqrt{n_i \hat{\theta}_{i,j} \hat{p}_j} \right)^2 \]

assessing the number of fish in each cell of \( M \). The latter two are based on the Freeman–Tukey statistic as recommended for mark–recapture data by Brooks, Catchpole, and Morgan (2000).

An R package called BTSPAS (Bayesian time-stratified population analysis system), which implements the Bayesian P-spline model of \( U \) and both models of \( \Theta \) along with computations of the DIC and Bayesian p-values is available from the Comprehensive R Archive Network (http://cran.r-project.org/index.html). This package also includes a demonstration based on the analysis of the following section and implements extensions of this model to the case of diagonal \( M \), separation of stocks (e.g., hatchery and wild fish), and inclusion of covariates for \( p \).

3. Application

Here we apply the models described in the previous section to analyze modified data from the study of Atlantic salmon smolts migrating along the Conne River, Newfoundland, in 1987. In this study, smolts were trapped at two sites along the river for 46 days between April 26 and June 10. A total of 4975 smolts were marked at the first trap of which 998 (20%) were recaptured at the second trap along with 13,363 unmarked smolts.

SD94 and MR02 both used this data set to illustrate their methods. SD94 excluded data from the first 3 and final 17 days of the study because small numbers fish were marked in these periods and estimated the total size of the run in the remaining period to be 75,000 fish with 95% CI (68,82) thousand. MR02 were able to analyze the entire data set and also estimated the total run size at 75,000 with 95% CI (69,81) thousand.

Although SD94 found it necessary to exclude some strata, data sparsity is not a severe issue. Our focus is to improve estimation for sparse data sets, and so we simulated new data with smaller capture probabilities by artificially reducing the numbers of fish captured at the second trap. To do this, we generated new values of \( m_{i,j} \) and \( u_j \) as:

\[ m_{i,j} \sim \text{Binomial}(m_{i,j}^*, 0.2) \]
\[ u_j \sim \text{Binomial}(u_j^*, 0.2), \]

where \( m_{i,j}^* \) and \( u_j^* \) are the original data values. This effectively reduced the capture probability on each day of the study by a factor of 5, but maintained the other relationships in the data. The new data set comprised the same number of marked fish released, but the total numbers of marked and unmarked smolts captured at the second site were reduced to 183 and 2697, respectively. A complete listing of the modified data set is provided in Web Table 1.

Table 1 summarizes inferences about the total population size from the 10 candidate models, and Table 2 presents the DIC and Bayesian p-values. Summaries of the posterior distributions for \( N_{\text{Tot}} \) obtained from the 10 models fit to the subsampled Conne River Atlantic salmon data. The first two columns provide the posterior mean (\( \hat{N}_{\text{Tot}} \)) and 95% highest posterior density credible interval (95% CI) in thousands of fish. The final two columns give the width of this credible interval absolutely (CI width) and as a percentage of the posterior mean (% CI width). Each model is labeled by a triplet identifying the components of the prior as described in the text.

<table>
<thead>
<tr>
<th>Model</th>
<th>( N_{\text{Tot}} ) (95% CI)</th>
<th>CI width</th>
<th>% CI width</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple/Pooled/LN</td>
<td>77 (67, 89)</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>Simple/Pooled/NP</td>
<td>80 (69, 92)</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>Simple/Hier/LN</td>
<td>87 (10, 107)</td>
<td>37</td>
<td>42</td>
</tr>
<tr>
<td>Simple/Hier/NP</td>
<td>90 (73, 111)</td>
<td>37</td>
<td>42</td>
</tr>
<tr>
<td>Hier/Hier/LN</td>
<td>81 (67, 100)</td>
<td>33</td>
<td>40</td>
</tr>
<tr>
<td>Hier/Hier/NP</td>
<td>85 (70, 103)</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>Cubic/Hier/LN</td>
<td>76 (64, 90)</td>
<td>26</td>
<td>35</td>
</tr>
<tr>
<td>Cubic/Hier/NP</td>
<td>79 (67, 92)</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>P-spline/Hier/LN</td>
<td>76 (64, 89)</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>P-spline/Hier/NP</td>
<td>78 (66, 92)</td>
<td>26</td>
<td>33</td>
</tr>
</tbody>
</table>
Model selection and goodness-of-fit results for the 10 Bayesian models fit to the subsampled Conne River Atlantic Salmon smolt data. The left half of the table presents the values of $p_0$ and DIC for each model, and the right half presents the Bayesian $p$-values for each discrepancy measure. The models are labeled by triplets identifying the components of the prior as described in the text.

<table>
<thead>
<tr>
<th>Model selection</th>
<th>$p_0$</th>
<th>DIC</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple/Pooled/LN</td>
<td>83.4</td>
<td>691</td>
<td>0.26</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Simple/Pooled/NP</td>
<td>80.7</td>
<td>671</td>
<td>0.05</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Simple/Hier/LN</td>
<td>94.4</td>
<td>672</td>
<td>0.41</td>
<td>0.29</td>
<td>0.32</td>
</tr>
<tr>
<td>Simple/Hier/NP</td>
<td>89.5</td>
<td>656</td>
<td>0.19</td>
<td>0.19</td>
<td>0.29</td>
</tr>
<tr>
<td>Hier/Hier/LN</td>
<td>93.0</td>
<td>676</td>
<td>0.47</td>
<td>0.20</td>
<td>0.24</td>
</tr>
<tr>
<td>Hier/Hier/NP</td>
<td>88.0</td>
<td>659</td>
<td>0.17</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>Cubic/Hier/LN</td>
<td>84.6</td>
<td>670</td>
<td>0.58</td>
<td>0.25</td>
<td>0.31</td>
</tr>
<tr>
<td>Cubic/Hier/NP</td>
<td>79.8</td>
<td>650</td>
<td>0.28</td>
<td>0.21</td>
<td>0.32</td>
</tr>
<tr>
<td>P-spline/Hier/LN</td>
<td>85.1</td>
<td>666</td>
<td>0.55</td>
<td>0.26</td>
<td>0.33</td>
</tr>
<tr>
<td>P-spline/Hier/NP</td>
<td>79.7</td>
<td>644</td>
<td>0.28</td>
<td>0.24</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Simple/Hier/NP provided relatively precise estimates at the peak of the run, but the posterior of $U_j$ was highly variable for $j$ near 1 or 4. Model Hier/Hier/NP decreased the variability of these values, as well as $N_{\text{tot}}$, but increased the DIC (Table 2) indicating that the hierarchical prior with constant mean is also not appropriate for this data. In contrast, models Cubic/Hier/NP and P-spline/Hier/NP lowered the posterior variability of the individual $U_j$ and of $N_{\text{tot}}$ (Figure 2 and Table 1), and also decreased both $p_0$ and the DIC (Table 2). Compared to each other, the curves fit to $E(\log(U_j))$ were difficult to distinguish by eye (Figure 2) and the estimates of $N_{\text{tot}}$ were almost coincidental, but the DIC for model P-spline/Hier/NP was lower by 6 units (Table 1). Bayesian $p$-values for these models presented no evidence to question their fit to the data (Table 2).

Convergence diagnostics indicated that the Markov chains for all models converged quickly to stable distributions and that the samples obtained represented the posterior distributions well. Trace plots and GRB diagnostics comparing the five parallel chains for each model indicated that the burn-in period was more than sufficient. Web Appendix A provides details for model P-spline/Hier/NP (the model with the smallest DIC) and results were similar for the remaining models.

Effects of the choice of prior for the P-spline parameters were assessed through a sensitivity analysis. We refit model P-spline/Hier/NP a total of 25 times for all combinations of five inverse gamma priors for $\tau_1^2$ and $\tau_2^2$ that varied according to their location and relative spread. Details of the selected priors and complete results are provided in Web Appendix B. In summary, the sensitivity analysis indicated that the shape of the fitted spline and the posterior distribution of $N_{\text{tot}}$ were robust to the choice of priors. However, the posterior distributions of the individual $U_j$ were affected by the choice of prior for $\tau_1^2$, and so we recommend that any analysis be repeated using a range of prior parameterizations—particularly if one is interested in local characteristics of the run (e.g., peak day).

### 4. Simulation Study

To examine the P-spline model further, we conducted a simulation study based on the results of the analysis of Section 3. In short, 300 data sets were simulated from the model described by equations (1) and (2) with $\log(U_i)$ generated from a smooth curve similar to the spline fit in Section 3, $\theta_{i,j}$ generated from the LN model but truncated so that $\theta_{i,i+k} = 0$ for $k > 6$, and $p_0$ generated from a distribution approximating the posterior obtained in Section 3. The 10 models described previously were fit to each data set and compared based on (1) the resulting bias and mean-squared error (MSE) of the posterior mean of $N_{\text{tot}}$ as a point estimate of the true population size and (2) the width and coverage of the associated 95% credible intervals. Performance of the DIC for model selection was also assessed. Full details of the simulation study are presented in Web Appendix C; here, we summarize the key results.

Overall, the Bayesian P-spline models produced estimates of $N_{\text{tot}}$ that were nearly unbiased and had small MSE relative to the other models, and also gave 95% credible intervals with exactly the nominal coverage. The only models which performed similarly were those fitting a cubic polynomial to $E(\log(U_j))$. These models actually had smaller bias and MSE,
Smoothed Time-Stratified Petersen Estimates

Figure 2. Estimated daily abundance of unmarked fish for the subsampled Conne River Atlantic Salmon data. Each panel summarizes the posterior distribution of $\log(U_j)$, $j = 1, \ldots, 46$, resulting from one of the five base models combined with the NP model of the travel time probabilities, as indicated. Points represent the daily posterior medians and the error bars represent the 95% credible intervals. The curves in the final two panels represent the functions fit to $E(\log(U_j))$.

though coverage of their credible intervals was slightly below the nominal value. The models with the simple or hierarchical fit to $E(\log(U_j))$ performed poorly in that they had large bias and MSE, and although point estimates from the models with pooled capture probabilities were nearly unbiased, the coverage of the credible intervals was only $\frac{2}{3}$ of the nominal value.

In accordance with these results, the DIC selected the Bayesian P-spline model of $E(\log(U_j))$ for 66% of the simulated data sets (26% and 40% in combination with the LN and NP models of $\Theta$, respectively). The cubic model was selected for 30% of data sets (11% and 19% in combination with the LN and NP models of $\Theta$), and of the remaining 4% of the simulated data sets, model Simple/Pooled/NP was selected eight times, Simple/Hier/NP three times, and Hier/Hier/NP once. The fact that the cubic model was selected for almost $\frac{1}{3}$ of the data sets was not surprising given that a cubic polynomial fits the true $E(\log(U_j))$ very closely and requires fewer parameters than the spline. Despite the similar performance of the two models, we prefer the Bayesian P-spline because of its abilities to fit a wide range of curves and to properly account for model uncertainty. These issues are discussed further in Section 5. Similar DIC values were also to be expected from the alternative priors for $\Theta$ as the true $\theta_{ij}$ were truncated, as in the NP model, but were well approximated by the LN model. These results confirm that the differences in DIC seen in Section 3 provide strong evidence for selecting between models and further support our conclusion that P-Spline/Hier/NP is the best of the candidate models for the subsampled Atlantic salmon data.
5. Conclusion

The results of Sections 3 and 4 clearly demonstrate the advantages of the P-spline model of $U$ for estimating the size of a population from temporally stratified mark–recapture data. The P-spline accounts for the natural ordering of the data by explicitly modeling the expected (log) population size per strata as a smooth function of time. The result is a nearly unbiased estimate of $N_{\text{Tot}}$ that is as or more precise than those from the other models.

The model that produced results most similar to those of the P-spline model was the cubic model of $E(\log(U_j))$. The cubic and P-spline curves fit to $E(\log(U_j))$ in Section 3 were almost identical and these models performed almost equally well in the simulation study. However, it is easy to imagine data for which the cubic would perform poorly, for example if the run size had two distinct peaks. Several strategies might be used to fit higher degree polynomials to such data, but each has drawbacks. One might test polynomials with increasing degree until an adequate model was found, but this would constitute data-snooping and variance estimates would not account for the trial-and-error process. Alternatively, a high degree polynomial could be fit to all data, but this would lead to overfitting, and reversible jump MCMC could be used to select the degree of the polynomial (Green, 1995), but this requires more complex computation. The Bayesian P-spline solves the problem of...
unknown model complexity by allowing flexibility in the degree of the curve while penalizing overly complex models. Moreover, by including $\tau^2$ as an unknown parameter in the model the variance estimates properly account for uncertainty in the complexity.

This is the first time, to our knowledge, that splines have been applied to model changes in population size over time from mark-recapture data, though similar methods have been used to model other aspects of population dynamics. Methods using splines to allow flexibility in the relationship between covariates and survival probabilities have been presented by Gimenez et al. (2006); Bonner, Thomson, and Schwarz (2009); Gimenez and Barbraud (2009); and Gimenez, Grégoire, and Gimenez et al. (2009). Although the nature of these models are similar, and all apply Bayesian inference, the specific spline implementations vary. For example, Gimenez et al. (2006) applied a penalized spline approach using the truncated polynomial basis functions, whereas Bonner et al. (2009) applied a free-knot method in which the knot locations were chosen as part of the inference procedure. Gimenez and Barbraud (2009) introduced a two-stage inference procedure to reduce computation by approximating the likelihood function, and Gimenez et al. (2009) developed fitness surfaces that simultaneously model the effects of two or more variables on survival with multivariate splines. Each method has advantages and disadvantages, and there is a large body of literature on splines (see, e.g., Ruppert, Wand, and Carroll, 2003). Although it is not clear if one method is particularly suited to modeling mark-recapture data, the Bayesian P-spline model of Lang and Brezger (2004) has worked well in the examples we have tried thus far.

As mentioned previously, there is a variant of the two-sample experiment, which requires only one trap. When only one trap is used, a subset of the individuals captured in each strata are marked and released back into the population upstream so that they pass the trap a second time. The only difference in modeling is that the marked fish do not need to be recounted when they pass the trap a second time and so the total population size satisfies $N_{\text{Tot}} = U_{\text{Tot}}$. An example of the one trap experiment is provided in Bonner (2008). In some studies it may also be known that $\theta_{i,j} = 1(i = j)$ and so it is not necessary to model the movements of the marked fish. Such data are called diagonal because $M$ is a diagonal matrix and examples of fitting the P-spline model to such data are provided by Schwarz et al. (2009).

The P-spline model also provides a way to deal with some common data anomalies. Fisheries traps often cannot be operated continuously because of adverse weather conditions, and the data from such studies may contain missing values in some strata. Classical estimators ignore these strata completely, whereas standard hierarchical Bayesian models can produce inference by drawing on the data from other strata but treat all other strata equally rather than weighting information from the neighboring strata more heavily. The P-spline model can easily be fit to data with missing days and essentially smooths the run size across the strata with unobserved data, allowing for uncertainty by addition of the error term. Of course, this relies on the untestable assumption that the patterns in the observed data continue over the missing strata.

One drawback of the P-spline approach is that MCMC simulations can be time consuming and sometimes misleading. Although the algorithms implemented in OpenBUGS generally performed well, we did encounter one data set in the simulation study in which the single chain became trapped in a local maximum of the posterior density. The error was fixed simply by choosing a new seed value, but this emphasized the importance of running multiple chains to assess convergence, as we did in Section 3. Gains from the model will also be small if the run size has a very simple underlying shape or if the number of strata is small so that the smoothness cannot be well estimated. The model may also produce misleading results if $E(\log(U_{i,j}))$ is not smooth—for example, if there are jumps in $U_i$ associated with sudden outmigration or the release of hatchery fish. If the timing of such events are known then they can be accommodated by incorporating breaks in the P-spline, and an example is provided in Bonner (2008).

Another concern is that the model treats the expected number of unmarked fish in each stratum, $E(\log(U_{i,j}))$, as the parameter of interest rather than the total number of fish, $E(\log(N_{i,j}))$. The difficulty is that $N_i$ does not appear in the likelihood in equation (3), which conditions on $n$. One could adopt a likelihood that models capture at both locations, as in Plante et al. (1998), and apply the P-spline to the numbers of fish passing the first trap in each strata, but this likelihood effectively doubles the number of parameters. Instead, we have considered applying the Bayesian melding algorithm of Poole and Raftery (2000) that allows specification of prior information for derived parameters like $N_i$ or $N_{\text{Tot}}$. Note, however, that if relatively few fish are marked then $N_i \approx U_i$ and whether the prior is defined for $N_i$ or $U_i$ will have little effect on inference.

Another minor issue is that the coefficients of the P-spline do not have direct interpretations. However, inference about quantities like run timing can easily be generated from the MCMC output. For example, inference about the day by which 50% of the population has passed the traps can be generated simply by computing this quantity for each iteration and then computing posterior summary statistics. Inference about many other biological quantities, like the peak day of the run, can be obtained in exactly the same way.

Although we have focused on smoothing abundance, the Bayesian P-spline could also be applied to smooth $p$ or the elements of $\Theta$. If $p_j$ was expected to change smoothly with time then it would be sensible to model $E(\log(p_j))$ in a similar manner. The Bayesian P-spline model could also be applied to multiple sets of parameters simultaneously, though we found that such models did not perform well in application to the Conne River data set. Specifically, models that smoothed both $U$ and $p$ produced poor estimates of abundance and DIC values that were much higher than those of the other models. Developing techniques to smooth multiple parameters simultaneously is a continuing aspect of our research.

6. Supplementary Materials

Web Appendices, Figures, and Tables referenced in Sections 2, 3, and 4 are available under the Paper Information link at the Biometrics website http://www.biometrics.tibs.org.
ACKNOWLEDGEMENTS

This research was supported by the National Science and Engineering Research Council of Canada, the Pacific Institute for the Mathematical Sciences, and the National Science Foundation (Grant 0814194). Further funding and data for model development were provided by the Trinity River Restoration Program. We also thank the editor and two anonymous referees for their constructive comments on earlier versions of the manuscript.

REFERENCES


Received March 2010. Revised February 2011. Accepted February 2011.