Time-Varying Covariates and Semi-Parametric Regression in Capture-Recapture: an Adaptive Spline Approach

Simon J. Bonner[1], David Thomson[2,3], and Carl J. Schwarz[4]

1. Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada sbonner@stat.sfu.ca

2. Max Planck Institute for Demographic Research, Konrad-Zuse Str. 1, D-18057, Rostock, Germany Thomson@demogr.mpg.de

3. Netherlands Institute of Ecology, Centre for Terrestrial Ecology, Boterhoeksestraat 48, Postbus 40, 6666 ZG Heteren, Netherlands.

4. Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada cschwarz@stat.sfu.ca

Abstract

Advances in capture-recapture methodology have allowed the inclusion of continuous, time-dependent individual-covariates as predictors of survival and capture probabilities. The problem posed by these covariates is that they are only observed for an individual when that individual is captured. One solution is to assume a model of the covariate which defines the distribution of unobserved values, conditional on the observed values, and apply Bayesian methods to compute parameter estimates and to test the covariate's effect.

Previous applications of this approach have modeled the survival probability as a linear function of the covariate on some scale (e.g. identity or logistic). In some applications a linear function may not adequately describe the true relationship. Here we incorporate semi-parametric regression to allow for more flexibility in the relationship between the covariate and the survival probabilities of the Cormack-Jolly-Seber model. A fully Bayesian, adaptive algorithm is used to model the relationship with splines, in which the complexity of the relationship is governed by the number and location of the knots in a spline. A reversible jump Markov chain Monte Carlo algorithm is implemented to explore splines with different knot configurations, and model averaging is used to compute the final estimates of the survival probabilities.

The method is applied to a simulated data set and to data collected through the Dutch Constant Effort Sites ringing project to study the survival of reed warblers (*Acrocephalus scirpaceus*) as a function of condition.

Keywords: adaptive splines, Bayesian inference, capture-recapture, free-knot splines, reversible-jump Markov chain Monte Carlo, semi-parametric regression

1 Introduction

Continuous, individual variables (e.g. body mass) may often be of interest as predictors of survival or catchability of animals in capture-recapture (CR) studies. The difficulty posed by variables that are unique to each animal and change over time is that they can only be observed when an individual is captured. Because of the missing values, these variables cannot be used in standard CR models (see e.g. Lebreton, Burnham, Clobert, and Anderson (1992) who employ environmental or fixed individual covariates as predictors of survival). Bonner and Schwarz (2006) introduce one method for including continuous covariates in CR studies by developing a hierarchical model of the unobserved covariate values and then using Bayesian analysis via Markov chain Monte Carlo (MCMC) simulation to obtain parameter estimates. The method is applied to study the effect of body mass on the survival of the meadow vole (*Microtus pennsylvanicus*).

The objective of this paper is to allow more flexibility in the relationship between the covariate and the survival probability. The model of Bonner and Schwarz (2006) assumes that the link between the probability that individual i survives from occasion t to $t+1$ and the covariate is a linear function on the logistic scale:

$$\eta_{it} = \text{logit}\left(P(\text{ind. } i \text{ is alive at time } t+1 | \text{ind. } i \text{ is alive at time } t)\right) = \beta_0 + \beta_1 z_{it}$$

where z_{it} denotes the value of the covariate for individual i at time t. In this paper we model η_{it} using a large class of non-linear functions. In particular we model:

$$\eta_{it} = s(z_{it})$$

where $s(\cdot)$ is a smoothing spline fit through an adaptive spline approach.

Smoothing splines are flexible functions formed from polynomial segments which connect at selected points called knots. Unlike polynomials which form a global fit to data, splines are locally adaptive such that the coefficients of the polynomial segments may vary over the range of the data (Eubank, 1999). The amount of local change possible in a spline is determined by two factors: the number and location of the knots, and the amount by which the spline may change at each knot. This dichotomy leads to two methods for fitting splines. Penalized spline methods use a large number of fixed knots over the range of the data, but introduce a smoothing parameter which limits how the coefficients of the polynomial segments may differ on either side of a knot. Adaptive (or free knot) spline methods include estimation of the number and location of knots as part of the fitting procedure.

The adaptive method can be framed as a problem of fitting a hierarchical model in which splines with different numbers of knots form different submodels. This problem is ideally suited to Bayesian inference incorporating reversible jump Markov chain Monte Carlo (RJMCMC) to explore splines with different numbers of knots and hence different numbers of parameters (Green, 1995). Final estimates are computed by model averaging the functions sampled on each RJMCMC iteration.

Section 2 of this paper provides an introduction to adaptive splines and describes the method for fitting survival probabilities as a function of a continuous covariate. Section 3 examines a simulation study where the survival probability is known to have local dependence on the covariate. Fit of the spline model is compared with the fit of a simpler cubic polynomial model. In section 4, we apply the method to study the relationship between the survival and condition of reed warblers (*Acrocephalus scirpaceus*) captured as part of the Dutch Constant Effort Sites (CES) ringing program. The final section discusses advantages and disadvantages of the method and provides some suggestions for its use in future CR studies.

2 Methods

2.1 Notation

Observed Data & Latent Variables:

T =number of capture occasions (indexed by t). *Observed.*

a_i =occasion on which individual i is first captured and marked. *Observed.*

z_{it}=value of time-varying covariate for individual i at time t. *Partially observed.*

Parameters:

ϕ_{it} =probability that individual i is available for capture at occasion $t+1$ given that it was available for capture at occasion t

p_t =probability that any individual available for capture at occasion t is captured

$s(z)$ =spline function modelling dependence of ϕ_{it} on z_{it}

κ =number of potential knot locations

K =number of knots in a single realization of the adaptive spline

ξ_k =location of the k^{th} knot

ξ_l, ξ_u=lower and upper boundary knots

$\boldsymbol{\beta}, \boldsymbol{\gamma}$ =vectors of coefficients of $s(z)$ (truncated polynomial basis)

$\boldsymbol{\beta}$ =vector of coefficients of $s(z)$ (B-spline basis)

μ_t =population mean change in covariate between t and $t+1$

σ^2 =population variance of change in covariate between adjacent capture occasions

2.2 Cormack-Jolly-Seber Model with Individual Covariates

The purpose of our method is to allow flexibility in modelling the relationship between some process in a capture-recapture experiment (e.g. survival or capture) and an individual covariate. Here we illustrate the method by incorporating a time-varying condition measure as a predictor of survival in the Cormack-Jolly-Seber (CJS) model. We assume that animals are captured and marked on T capture occasions and released back into the population where they can be recaptured on subsequent occasions. The CJS models the probability of recapturing individuals in terms of two sets of parameters: the apparent survival probabilities and the capture probabilities. Standard assumptions for the CJS model can be found in many sources (see e.g. Seber (1982); Williams, Nichols, and Conroy (2002)).

Our model modifies the assumptions by allowing the survival probability to depend on a single time-varying covariate, denoted z_{it}. Bonner and Schwarz (2006) models the relationship between ϕ_{it} and z_{it} using a simple logistic function:

$$\eta_{it} = \text{logit}(\phi_{it}) = \beta_0 + \beta_1 z_{it}.$$

The primary difficulty with fitting this model is the large number of missing values for the covariate that may result when animals are not captured. Bonner and Schwarz (2006) solves this by defining a distribution for the unobserved values of the covariate conditional on the observed values. In particular, the model assumes that changes in the covariate between subsequent capture occasions are normally distributed according to:

$$z_{i,t+1}|z_{it} \sim N(z_{it} + \mu_t, \sigma^2) \tag{1}$$

where μ_t is allowed to vary with t and σ^2 is constant. Estimates of β_0 and β_1 are then generated from Bayesian inference via Markov chain Monte Carlo.

Here we extend the model by allowing η_{it} to be a non-linear function of z_{it}. In particular, we model $\eta_{it} = s(z_{it})$ where $s(z)$ is a spline. For simplicity, the capture probabilities are assumed to depend only on the capture occasion.

2.3 Splines

The class of splines is a set of functions that is commonly used for smoothing – finding a flexible functions, $y = f(x)$, to describe coordinate data $(x_1, y_1), \ldots, (x_n, y_n)$. Part of their appeal is that splines can be formulated as an extension of polynomial regression, and much of the methods and theory used to fit simple regression models can be applied to fitting splines. The disadvantage of a straight polynomial regression model for scatterplot smoothing is that its fit is global. The value of the polynomial over any (small) interval determines its value over the entire range of data (Schumaker, 1993, pg. 103). If the relationship between y and x is complicated then the polynomial model will not fit well or will require many terms to achieve an adequate fit.

Spline models remedy this by introducing extra predictors that allow local changes in the fitted curve. A polynomial of order q is a function formed as a linear combination of the functions $\{1, x, \ldots, x^{q-1}\}$. A spline of order q on the interval $[\xi_l, \xi_u]$ is a function formed as a linear combination of the basis functions $\{1, \ldots, x^{q-1}, (x-\xi_1)_+^{q-1}, \ldots, (x-\xi_K)_+^{q-1}\}$, for some points $\xi_l < \xi_1 < \ldots < \xi_K < \xi_u$. That is:

$$f(x) = \beta_1 + \beta_2 x + \ldots + \beta_q x^{q-1} + \sum_{k=1}^{K} \gamma_k (x - \xi_k)_+^{q-1}.$$

The function $(x-\xi_k)_+^{q-1}$ is a truncated polynomial of order q defined to be equal to 0 for $x < \xi_k$ and $(x-\xi_k)^{q-1}$ for $x \geq \xi_k$. The points ξ_l and ξ_u are the boundary knots of the spline and the points ξ_k, which partition $[\xi_l, \xi_u]$ into $k+1$ disjoint intervals, are the internal knots of the spline.

Over any interval between adjacent knots, $[\xi_k, \xi_{k+1}]$, the spline is equal to a polynomial of order q. At each knot the form of this polynomial is allowed to change, but only in a constrained manner which ensures that the entire curve over $[\xi_l, \xi_u]$ will be smooth. The curve must be continuous at each knot $(\lim_{x \to \xi_k^-} f(x) = f(\xi_k) = \lim_{x \to \xi_k^+} f(x))$, it will have $q-2$ continuous derivatives over the entire range, and the $(q-1)^{th}$ derivative will be continuous except for jumps of size γ_k at each knot. In contrast, a polynomial of order q has $q-1$ non-zero derivatives all of which are continuous. Our method considers cubic splines

$(q = 4)$ so that the spline is equivalent to a cubic polynomial between adjacent knots and has 2 continuous derivatives (Schumaker, 1993, pg.108).

The flexibility of a spline comes from the jumps in the $(q-1)^{st}$ derivative allowed at each knot. For a cubic spline, the jumps occur in the 3^{rd} derivative which means that the second derivative is continuous but may change very sharply at each knot. With q fixed, more flexibility can be introduced in the spline in one of two ways: 1) increasing the number of knots or 2) allowing for larger jumps in the derivative. This dichotomy leads to two competing approaches for fitting splines. Penalized spline methods use a large number of knots with fixed positions to allow flexibility anywhere in the range, and maintain smoothness by constraining the size of the jumps in the $(q-1)^{st}$ derivative allowed at each knot (Ruppert, Wand, and Carroll, 2003, pg.65). This requires specification or estimation of a smoothing parameter which constrains the $\boldsymbol{\gamma}_k$. Adaptive or free-knot spline methods allow $\boldsymbol{\gamma}_k$ to change more freely and include estimation of the number and location of the knots within the fitting procedure so that knots can be placed where needed. This is the approach taken in our method.

2.4 B-splines

In practice, the truncated polynomial basis $\{1, \ldots, x^{q-1}, (x-\xi_1)_+^{q-1}, \ldots, (x-\xi_K)_+^{q-1}\}$ can often lead to numerical problems. First, the columns of the design matrix can be highly correlated if ξ_k and ξ_{k+1} are close; and second, the entries of the design matrix can be very large so that matrix operations become unstable. These problems can be avoided by using an equivalent basis of functions: i.e. a set of $q+K$ functions such that every spline can be written as a linear combination of the new set of functions and vice versa. A common choice is the set of B-spline basis functions (Ruppert et al., 2003; Schumaker, 1993). This basis has several computational advantages including: 1) the basis functions are positive and sum to 1 at any single point, 2) as a result, the values of the design matrix will always be between 0 and 1, and 3) all of the functions in the basis are local in that each is positive only over a sub-interval of $[\xi_l, \xi_u]$. In the new basis, a spline of degree q with K knots can be written as:

$$s(z) = \sum_{j=1}^{q+K} b_j B_j^{(q)}(z)$$

where $B_j^{(q)}(z)$ is the j^{th} B-spline basis function of order q and b_j is its coefficient.

One important consideration in computing the B-spline basis is the choice of the values ξ_l and ξ_u which define the interval over which the spline will be computed. These points are often called the boundary knots, and are required in computing the values of the B-spline basis functions. While the basis functions can be computed even for points which lie outside of this interval, the choice of the boundary knots is crucial for numerical reasons. At any point which lies inside $[\xi_l, \xi_u]$ the value of the B-spline basis functions will all lie in $[0, 1]$ and the sum over all of the basis functions will be exactly 1. This makes computing with the basis very stable. At points outside of this interval the value of the B-spline basis functions may be negative, and may become very large which makes the algorithms prone to numerical errors. This issue is addressed further below. Further discussion and algorithms for computing B-splines are provided in (Schumaker, 1993).

2.5 Bayesian Adaptive splines

To estimate the function $s(z)$ we follow the adaptive Bayesian method of Biller (2000) for fitting B-splines to generalized linear models. First, we select a large number, κ, of potential knot locations prior to the analysis. Here we space these locations evenly across the observed range of the data. The prior distribution is then constructed so that it places all of its mass on functions in the set of splines with between 1 and κ knots at these locations. This prior distribution is then updated with information from the data, in the form of the likelihood, to generate the posterior distribution over the same space of splines. In practice, even though the number of different knot configurations is finite the likelihood cannot be computed for each model and so the posterior cannot be constructed analytically. Instead, Biller provides a Markov chain Monte Carlo algorithm for generating a sample of realizations from the posterior distribution from which inference can be made.

A single spline in the restricted space is identified by three components: K, the number of knots ($0 \leq K \leq \kappa$); $\boldsymbol{\xi}$, the vector of knot locations; and $\boldsymbol{b}$, the coefficients of the spline. To define a distribution over the space we need to define a joint density for these parameters. Following Biller (2000) again we write the joint prior density as the product of densities:

$$\pi(\boldsymbol{b}, \boldsymbol{\xi}, K) = \pi(\boldsymbol{b}|\boldsymbol{\xi}, K)\pi(\boldsymbol{\xi}|K)\pi(K)$$

where $\pi(\boldsymbol{b}|\boldsymbol{\xi}, K)$ is the conditional density of the coefficients given the number and locations of the knots, $\pi(\boldsymbol{\xi}|K)$ the density of the knot locations given the number of knots, and $\pi(K)$ the marginal density on the number of knots. First, we assign $\pi(K)$ a truncated Poisson(λ) distribution so that:

$$\pi(K) \propto e^{-\lambda}\lambda^K(K!)^{-1}.$$

Then given K knots we assume that all configurations of the knots are equally likely. As there are $\binom{\kappa}{K}$ possible splines with K knots out of the κ potential locations, this leads to the conditional density:

$$\pi(\boldsymbol{\xi}|K) = \binom{\kappa}{K}^{-1}.$$

Finally, given $\boldsymbol{\xi}$ and K we assign the elements of $\boldsymbol{b}$ independent, diffuse normal priors with mean 0 and variance τ^2.

The likelihood for the model is an extension of the CJS likelihood with time-varying covariates and is exactly as given in Bonner and Schwarz (2006) with logit $(\phi(z)) = s(z)$. Combined with the prior distribution above, this defines a posterior over the restricted space of splines. Because of the large number of models in the space it is not possible to compute summaries of the posterior distribution analytically. Instead, Biller (2000) provides a MCMC algorithm that samples different realizations from the space. Inference is then made by computing summary statistics from this sample of functions.

The major challenge in the algorithm is that moving between splines with different numbers of knots changes the dimension of the model by increasing or decreasing the length of $\boldsymbol{b}$. Moves between models of different dimension cannot be accommodated in standard MCMC algorithms, like Metropolis-Hastings (MH), and instead, adding or removing knots is performed through RJMCMC. Like the MH algorithm, each iteration of RJMCMC involves proposing a new state for the parameters conditional on the current parameter values. This new function is then accepted with a probability computed from the prior distributions of the current and proposed states, their likelihoods given the observed data, and their densities under the proposal distribution. If accepted, the proposed state becomes the current state, and the chain continues. Otherwise the current state is retained and a new proposal is generated. As in the MH algorithm, the acceptance step in RJMCMC ensures that the posterior distribution will be a stationary distribution of the Markov chain. The RJMCMC acceptance probability was first derived by Green (1995); a simplified derivation is given in Waagepetersen and Sorensen (2001) and more details are available in recent books on Bayesian analysis or statistical computation (see e.g. Chen, Shao, and Ibrahim (2000)).

The specific algorithm of Biller (2000) generates new proposals by adding or deleting one knot from the current spline. On each iteration, a random choice is made to add or delete one knot. When adding a knot, the location of the new knot is selected from all of the currently unoccupied locations with equal probability. When deleting a knot, the knot to delete is randomly chosen from the currently occupied locations. Each iteration also includes a step for moving one chosen knot to a nearby unoccupied location, which helps the chain to move across the function space.

Implementation of this algorithm for our model poses some difficulties because of the unobserved covariate values. In particular, Biller (2000) recommends choosing the boundary knots, ξ_l and ξ_u, equal to the minimum and maximum values of the covariate and spacing the κ potential knot locations between. This is not possible in our application because the minimum and maximum values are not actually observed. Setting ξ_l and ξ_u equal to the observed minimum and maximum, say $z_{\min}^{\text{obs}}$ and $z_{\max}^{\text{obs}}$, it is likely that some of the unobserved covariate values will lie beyond the boundary knots. Instead, we recommend choosing ξ_l and ξ_u to enclose a wide range about the observed data (e.g. $\xi_l = z_{\min}^{\text{obs}} - (z_{\max}^{\text{obs}} - z_{\min}^{\text{obs}})$ and $\xi_u = z_{\max}^{\text{obs}} + (z_{\max}^{\text{obs}} - z_{\min}^{\text{obs}})$) but still spacing the potential knots equally between $z_{\min}^{\text{obs}}$ and $z_{\max}^{\text{obs}}$. This arrangement allows for covariate values to lie outside of $[z_{\min}^{\text{obs}}, z_{\max}^{\text{obs}}]$, but constrains the curve to be a cubic polynomial on the intervals $[\xi_l, z_{\min}^{\text{obs}}]$

and $[z_{\max}^{\text{obs}}, \xi_u]$ where there is no observed data. Sampling from the posterior distribution also requires steps to impute the unobserved covariate values and to update the remaining model parameters.

As in Bonner and Schwarz (2006) steps must be included in the algorithm for updating the remaining parameters of the model and simulating the missing data values. The final algorithm for generating a sample from the posterior distribution of all random variables has the following structure:
Initialization:

1) Define the boundary knots, ξ_l and ξ_u, and select κ potential knot locations between $z_{\min}^{\text{obs}}$ and $z_{\max}^{\text{obs}}$.

2) Select initial values for all parameters, hyperparameters, and the missing data values for each individual.

MCMC Iteration:

1) Latent data:
   Simulate the unobserved covariates, z_{it}.

2) Parameters of covariate distribution:
   Update the parameters $\mu_1, \ldots, \mu_{T-1}$, and σ^2.

3) Spline fit:

   (a) Propose change in dimension (i.e. addition of a knot at an empty location or deletion of a randomly selected knot).

   (b) Propose movement of a single, randomly chosen knot to a vacant location in the same neighbourhood.

   (c) Update the spline coefficients, $\boldsymbol{b}$.

4) Remaining CJS parameters:
   Update the capture probabilities $p_2, \ldots, p_T$.

MCMC iterations are then repeated until the chain converges and a large sample of realizations is generated. Full details of the different MCMC steps are described in a technical report Bonner (2007).

Because there is no single set of knots, an estimate of $s(z)$ cannot be generated by plugging estimates of the coefficients into the B-spline equation. Instead, $s(z)$ is estimated by the posterior mean over the entire model space which is approximated by averaging over the functions sampled on a large number of iterations from the tail of the Markov chain. Precision of $s(z)$ is assessed with pointwise 95% highest posterior density (HPD) credible intervals. That is, for each value of z in $[\xi_l, \xi_u]$ we compute the shortest interval which covers 95% of the sampled values of $s(z)$.

3 Simulation Study

In our simulation study, capture histories for 500 individuals were generated from a CJS model with 3 capture occasions. Covariate values for each individual were simulated from the diffusion model in (1) with the initial distribution $z_{a_i} \sim N(0, 1)$ and parameters $\mu_1 = \mu_2 = 0.00$ and $\sigma^2 = 1.00$. Survival probabilities for each interval were computed from a bimodal function of the covariate with modes at $z = -1$ and $z = 1$. This function is plotted in Figure 1. Capture probabilities were $p_2 = p_3 = .85$. Marking times were assigned so that half of the individuals were first captured at occasion 1 and half at occasion 2.

Three different models of the survival probability were fit to the simulated data to study the method's performance. The first model fit a cubic polynomial to the logit of the survival probability. The second and third fit the adaptive spline model described above using two different prior distributions for the number of active knots: Poisson with mean 25 and Poisson with mean 75. Markov chains for all three analyses were run for 100,000 iterations. The initial 10,000 were discarded as burn-in, and every 10^{th} of the remaining 90,000 iterations were retained for inference.

Estimates of the survival probability as a function of the covariate for all 3 models are plotted along with their pointwise 95% HPD credible intervals in Figure 1. The cubic function clearly is too rigid to adjust to the local changes in the survival probability. Instead, the fitted curve decreases throughout the range of the covariate, and the 95% credible intervals fail to cover the true survival probability for much of the range.

In comparison, the spline fit using the Poisson(25) prior easily captures the bimodality of the survival probability. The pointwise 95% credible intervals completely cover the true function, but are between 2 and 3 times wider than the credible intervals for the cubic fit. A trace plot and histogram of the number of active knots in the spline for each MCMC iteration and a plot indicating the average number of times each knot location is occupied are shown in Figure 2. The number of knots appears to converge very quickly to a stable distribution which places 95% of the posterior probability on models with between 5 and 14 knots. The posterior median is 10 knots. Knot locations that are most often occupied are centred near the largest mode, although knots throughout the range have a minimum probability of approximately .01 of being occupied.

The posterior distribution of the spline fit the Poisson(75) prior assigns probability to more complex models with many more knots. Ninety-five percent of the posterior probability is assigned to models with between 14 and 35 knots, with a posterior median of 23 knots. The minimum posterior probability of activity is doubled to approximately .02.

The result of using the Poisson(75) prior is that more knots are included in the spline on each MCMC iteration and so the estimated survival probability as a function of the covariate is less smooth. In fact, the estimate now contains considerable noise and the most frequently occupied knot locations are associated with an anomalous local change in the survival probability at z=1.8. This spike in the survival probability is caused by a chance grouping of individuals in the data all of which are captured on one occasion with covariate values near 1.8 and fail to survive until the next capture occasion. Examination of the true survival status (available from the simulation) versus the covariate shows exactly the same result. Note that even though the estimated survival probability is far from the true value, the 95% credible intervals still cover the truth at all points.

Results for the remaining parameters are given in Table 1. Changing the model of the survival probabilities has negligible affect on their estimates, and the means and 95% credible intervals produced by all three models are remarkably similar. The reason for this is the combination of high capture and high survival probabilities. Of the 500 individuals, 300 are captured on at least 2 occasions. With this many recaptures good estimates of the capture probabilities can be obtained from direct comparison of the capture histories, and the observed covariate values allow accurate estimation of μ_1, μ_2, and σ^2, without any knowledge of the survival probability.

4 Example

Data for the study of reed warblers (*Acrocephalus scirpaceus*) was obtained from the Dutch Constant Effort Sites (CES) banding project Speek (2006). In the CES project, volunteer ringers capture birds on 12 day-long visits between April and August to each of 38 sites in Holland. Ringers optionally record demographic and biometric characteristics of captured birds including age, sex, body mass, wing length, and tarsus length. The program was initiated in 1994 and data was available for 10 years up to 2003.

This analysis applies our extended CJS model to the final 5 years of data using a measure of the birds' condition as a predictor of survival. Each year of the study was considered as a single capture occasion. Multiple captures of the same bird in one year were combined by collapsing records into a single capture indicator and averaging the biometrics measurements. The condition measure for a single bird in a single year was defined as the ratio of its average observed body mass to its average observed wing length. After cleaning the data, measurements of the body mass ranged from 9.9 g to 17.5 g and wing length from 62 mm to 71 mm. The range of the condition measure was .15 g/mm to .24 g/mm.

The CES database contains records of approximately 300,000 captures of 25,000 reed warblers captured between 1999 and 2003. The majority of these birds were observed only once, which was taken as evidence of large numbers of transients in the population. To avoid heterogeneity in the survival probability resulting from immigration, we used an ad hoc method restricting our analysis to resident individuals, including only birds captured 2 or more times – even if the 2 captures occurred in the same year. Capture histories were then conditioned on the birds' second release. Observations of juvenile birds were removed from the data because we believed that the probability of survival, and its relation to condition, was likely to differ between juveniles and adults. It was also necessary to remove many individuals for whom the condition measure could not be computed in any year of the study because of missing data on body mass or wing length. The final

data set for our analysis contained capture histories of 592 birds, with 111 captured on 2 or more years.

As in the simulation study, three models were fit to the relationship between the survival probability: a cubic model on the logit scale, and two adaptive spline models with different priors on the number of knots. For the spline models, $\kappa = 100$ potential knots were equally spaced between the minimum and maximum observed condition values and the boundary knots were located at .05 and .34 g/mm. The prior distributions on the number of knots were Poisson with mean 25 in the first run and Poisson with mean 75 in the second. Separate intercept terms were included in all three models to allow the survival probabilities to change over time and capture probabilities were also modelled separately for each capture occasion. Note that $\phi_4(z)$ and p_5 are not completely confounded because of the dependence on z, but are very weakly identifiable separately and so only their product was estimated. The covariate was again modelled according to (1).

Figure 3 compares the survival probabilities estimated from the three different models. Here the fit of the curves is similar for all three models. Some local effects do appear in the spline estimates, but these occur at extreme values of condition where few birds are observed and come at the cost of much lower precision. Indeed, the apparent peak in survival at .22 g/mm arises from two birds captured with covariate values near this point. The peak disappears from both spline models when these birds are removed (results not shown) and the credible intervals at this point range from approximately .10 to .90 indicating extreme uncertainty. As in the simulation, the Poisson(75) prior generates a posterior that places higher probability on models with more knots, which decreases the smoothness and precision of the estimated survival probability. The 95% credible intervals for all three models overlap at all points.

Estimates of the remaining parameters and a point estimate of survival at the median value of condition, .17 g/mm, are provided in Table 2. Also included are the results of fitting a Bayesian implementation of the CJS model with no effect of the covariate. The results are very similar for all four models. In all cases, the estimates suggest no significant change in the capture probability over time, though there is a slight decrease in the survival probability. Credible intervals of $\phi_t(z)$ are wider for the spline models than for the cubic model, but this does not affect the remaining parameters. The estimates of μ_t are all close to 0 indicating that there is no distinct increase or decrease in the birds' condition over any period. The estimate of σ is .012 g/mm for all models. For a bird with fixed wing length between 62 and 71 mm, this translates to an estimated standard deviation in mass between .74 and .85 g.

5 Discussion

To our knowledge, only one other author has suggested the use of splines to allow more flexibility in modelling a covariate's effect in capture recapture methods. Gimenez, Covas, Brown, and Anderson (2006a) incorporates splines to model the effect of the southern oscillation index (SOI) on the survival of Snow petrels living in Terre Adèlie, and Gimenez, Crainiceanu, Barbraud, Jenouvrier, and Morgan (2006b) to model the effect of body mass on the survival of sociable weavers in South Africa. Both applications employ Bayesian methods but differ from our model in two respects. First, there is no missing data in the covariate: in the first applications, (SOI) is an environmental covariate that can be observed regardless of the capture of individual birds, and in the second, a single, static covariate is computed by averaging all observations of body mass for each bird. The second difference is that both applications make use of penalized splines with fixed number of knots and fixed knot locations, rather than adaptive (free-knot) splines.

Another application of spline methods in a capture-recapture study is given by Fewster (2007). Their objective is to estimate the distribution of residence times for southern right whales' on their breeding grounds in the Auckland Islands from multiple sighting data. Penalized cubic splines are fit to the density function of the residence times to ensure a smooth result. In contrast to our method, and the work above, inference in this application is based on classical likelihood methods.

The advantage of an adaptive spline over a penalized spline is that the number and location of the knots are estimated as part of fitting the spline, so that knots can be located where more flexibility is needed. One consequence is that the degree of smoothing can vary across the spline by clustering knots where the curve changes most rapidly and placing no knots where the curve is most smooth. The posterior distribution computed from the simulated data set with the Poisson(25) prior assigns low probability of being occupied to most knot locations, except for those in the area of the largest mode which are 2 to 3 times more likely to be occupied. Denison, Mallick, and Smith (1998) provides some examples of adaptive spline fits to even

more rapidly changing functions with jump discontinuities. In contrast, a penalized spline fit has an upper limit to the flexibility which is determined by the spacing of the fixed knots. Of course, the added flexibility also increases the potential for overfitting as seen in both the simulation and the analysis of the reed warbler data.

The Bayesian approach to adaptive spline fitting also provide a natural way to incorporate uncertainty concerning the location of the knots. Selecting a single set of knots for the model, based on the posterior mode or any other criteria, ignores the fact that the knot locations aren't known and that different knot configurations provide different fitted curves. The Bayesian methodology provides posterior model probabilities that naturally rank the different models. Model averaging using these probabilities to weight the different models then provides a clear way to aggregate the different fitted curves into a single estimate. The posterior mean over all knot locations will favour the features of the most probable models, but also includes some features of the less probable models weighted according to their posterior probabilities. Credible intervals computed from the entire set of models also allow for uncertainty across different sets of knots and will generally be wider than those computed from a single set of knots (Hoeting, Madigan, Raftery, and Volinsky, 1999).

The primary difficulty with Bayesian adaptive spline methods is selecting the joint prior distribution on the set of models indexed by $\boldsymbol{b}$, $\boldsymbol{\xi}$, and K. As is evident from the examples in this paper, the choice of prior is a very important determinant of the smoothness of the final curve. A prior distribution that places too much mass on simple models risks ignoring important aspects of the data, while a prior that favour complex models risk overfitting. Apart from the prior distribution there is no penalty for the model complexity, and overfitting is a serious concern. In simple smoothing of a scatterplot it is possible to choose the priors subjectively and then plot the fitted curve over-top of the raw data to assess the fit. The difficulty in our application is that neither the covariate nor the response are completely observed, so the fit of the curve cannot be visualized directly.

Our recommendation is that several prior distributions be selected and the resulting curves compared to see how the fit changes and whether the changes are biologically plausible. For the reed warbler data, the obvious difference between the cubic fit to the survival probability and the spline fits with Poisson(25) and Poisson(75) prior is the peak at .22 g/mm. The size of the peak in the last set of curves in Figure 3 is striking, but the point where this occurs is well beyond the 97.5%-ile of the observed covariate values, where there is little data, and the 95% credible intervals at this point are very wide. Further analysis reveals that the peak is the effect of 2 birds and disappears once these birds are removed. Despite the size of the peak, it seems clear that there is no evidence of a jump in the survival probability at .22 g/mm.

A second subjective decision that must be made in applying our method is the choice of boundary knots $[\xi_l, \xi_u]$. In theory, as long as $[\xi_l, \xi_u]$ encloses the observed data this choice should have little effect on the fitted curve. The challenge is that if $[\xi_l, \xi_u]$ is very wide then the distance from the lower boundary to the first internal knot or from the last knot to the upper boundary will be large. This will lead to small values in the design matrix and the numerical algorithms may become unstable. Conversely, if $[\xi_l, \xi_u]$ is too narrow then it is possible for imputed values of the covariate to lie far outside this range and similar problems can occur. To assess the impact of this choice, we repeated the analysis of the simulated data with several values for of ξ_l and ξ_u and found no effect on the fitted survival probability. We also encountered no numerical problems using our default choice for the boundary knots.

One source of confusion with the adaptive-spline method might be the apparent discrepancy between the prior and posterior distribution on the number of knots. In both the simulation and example, the posterior distribution concentrates its mass on much simpler models than the prior distribution. This seems contradictory, but exactly the same behaviour appears in the original examples of Biller (2000). In his discussion of another paper on Bayesian adaptive splines, Holmes (2002) explains that the apparent discrepancy is a result of the Bartlett-Lindley paradox. Ignoring the prior distribution on $\boldsymbol{\xi}$ and K, the vague multivariate-normal prior on the spline coefficients, $\pi(\boldsymbol{b}|\boldsymbol{\xi}, K)$, induces a prior for the data whose variance increases with the dimension of $\boldsymbol{b}$. As a result any observed data has lower prior probability under more complex models and the distribution actually places less and less mass on models of higher and higher complexity.

The key in the adaptive spline model is that the prior on $\boldsymbol{b}|\boldsymbol{\xi}, K$ which favours simple models is partially offset by the Poisson prior on K which assigns very little mass to these models – when λ is large enough. The resulting prior distribution is a balance that assigns its mass to models simpler than those favoured by

$\pi(K)$ alone. In essence, there is no discrepancy; rather, the prior on the set of models has to be interpreted through the full joint distribution, $\pi(\boldsymbol{b}, \boldsymbol{\xi}, K)$, and not simply the marginal prior $\pi(K)$.

Another caution with our approach, and with methods incorporating time-varying covariates in general, is the amount of data needed to provide adequate estimates. The final data set for the reed warblers contained 592 animals, but only 111 were captured on two or more occasions. Further, the condition measure was recorded two or more times for only 77 birds. This provides very little information regarding how the covariate changes over time and how differences in condition might affect the birds' survival. Fitting a Bayesian implementation of the standard CJS model to the data (assuming time-dependent survival and capture probabilities and ignoring any effect of the covariate) yields credible intervals for the survival probabilities that are between .34 and .36 wide. In light of this uncertainty when ignoring the covariate, it seems unlikely that any model will be able to detect an effect of condition on survival.

Based on our experience, including time-dependent covariates in the CJS model requires more data (i.e. capture of more individuals at given capture and survival rates) than models assuming homogeneity of individuals, or using environmental or static individual predictors. Using splines to model the survival probability as a function of the covariate will require even more data. Whereas a parametric curve borrows information from all values of the covariate to estimate the survival probability at any given value of the covariate, the spline only uses information from a local neighbourhood of covariate values. The result is that if few birds are observed with values in a given range of the covariate then in that range the estimate from the spline model will be highly variable though the estimate from the parametric model may still be precise.

As a final note we address the removal of transient birds from the CES reed warbler data. The majority of individuals captured are never recaptured and it is likely that many of these individuals are passing through the sites while migrating to other locations. The apparent survival probability of these individuals is 0. Pradel, Hines, Lebreton, and Nichols (1997) developed a model to account for transient individuals in the standard CJS model and compared it to the *ad hoc* method of conditioning on second release. They found that the *ad hoc* method produced unbiased estimates of the survival probabilities and was almost as efficient as the more complicated model when capture probabilities are high and the model cannot be simplified. With capture probabilities of .4 they found relative efficiency greater than .8.

An added effect in our application may be filtering of the values of the covariate. Resident individuals captured with values that equate to low survival probabilities will have less chance of being recaptured and more chance to be removed from the analysis. This should not bias the estimates of survival probability, but will decrease the precision of the estimates where these individuals are removed. The analysis is not intended to be an exhaustive examination of the reed warbler data, and how to deal with transients properly remains an open question.

Acknowledgements

Simon Bonner was partially funded by a CGS-D scholarship from the National Science and Engineering Research Council of Canada (NSERC). We are indebted to Dr. Rachel Fewster, Dr. Ken Burnham, and an anonymous reviewer for their constructive comments on earlier drafts.

References

C. Biller. Adaptive Bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics*, 9(1):122–140, March 2000.

S. J. Bonner. Time-varying covariates and semi-parametric regression in capture-recapture: an adaptive spline approach. Technical report, Simon Fraser University, 2007.

S. J. Bonner and C. J. Schwarz. An extension of the Cormack-Jolly-Seber model for continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics*, 62(1):142–149, March 2006.

M.-H. Chen, Q.-M. Shao, and J. G. Ibrahim. *Monte Carlo Methods in Bayesian Computation*. Springer, New York, 2000.

D. G. T. Denison, B. K. Mallick, and A. F. M. Smith. Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society, Series B*, 60:333–350, 1998.

R. L. Eubank. *Nonparametric Regression and Spline Smoothing*. Marcell Dekker, Inc., New York, USA, 2nd edition, 1999.

R. M. Fewster. Cubic splines for estimating the distribution of residence times using indvidual resighting data. *Environmental and Ecological Statistics*, *(*):*, 2007.

O. Gimenez, R. Covas, C. R. Brown, and M. D. Anderson. Nonparametric estimation of natural selection on a quantitative trait using mark-recapture data. *Evolution*, 60(3):460–466, 2006a.

O. Gimenez, C. Crainiceanu, C. Barbraud, S. Jenouvrier, and B. J. T. Morgan. Semiparametric regression in capture-recapture modelling. *Biometrics*, 62:691–698, 2006b.

P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–7332, 1995.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical Science*, 14:382–417, 1999.

C. Holmes. Discussion of: Spline adaptation in extended linear models. *Statistical Science*, 17(1):22–24, 2002.

J.-D. Lebreton, K. P. Burnham, J. Clobert, and D. R. Anderson. Modelling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, 62(1): 67–118, March 1992.

R. Pradel, J. E. Hines, J.-D. Lebreton, and J. D. Nichols. Capture-recapture survival models taking account of transients. *Biometrics*, 53:60–72, 1997.

D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, Cambridge, UK, 2003.

L. L. Schumaker. *Spline Functions: Basic Theory*. John Wiley & Sons, Inc., 2nd edition, 1993.

G. A. F. Seber. *The Estimation of Animal Abundance and Related Parameters*. Macmillan, New York, New York, USA, 2nd edition, 1982.

H. G. Speek. Ces handleiding. Technical report, December 2006. URL `http://www.vogeltrekstation.nl/ces_handleiding.htm`.

R. Waagepetersen and D. Sorensen. A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping. *International Statistical Review*, 69:49–61, 2001.

B. K. Williams, J. D. Nichols, and M. J. Conroy. *Analysis and Management of Animal Populations*. Academic Press, San Diego, California, 2002.

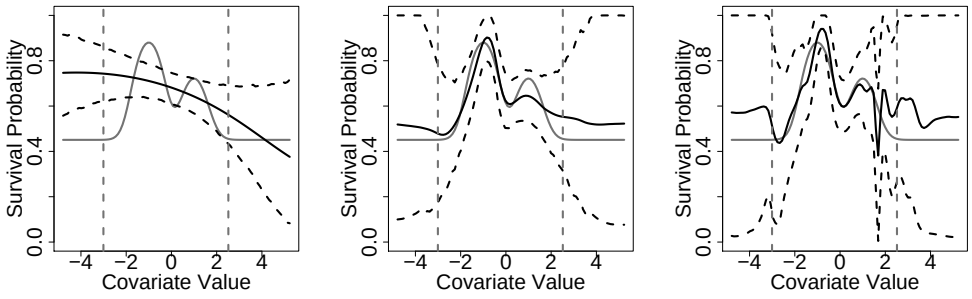

Figure 1: Estimated survival probability as a function of the covariate for the simulated data. The left plot illustrates the estimated function assuming a cubic fit, the centre plot using the adaptive spline method with a Poisson(25) prior on the number of knots, and the right plot using the adaptive spline method with a Poisson(75) prior on the number of knots. In each plot the solid grey line indicates the true function, solid black line the pointwise posterior mean fit, and dotted black lines the bounds of the pointwise posterior 95% credible interval. The vertical dotted grey lines indicate the 2.5 and 97.5 percentiles of the simulated covariate values.

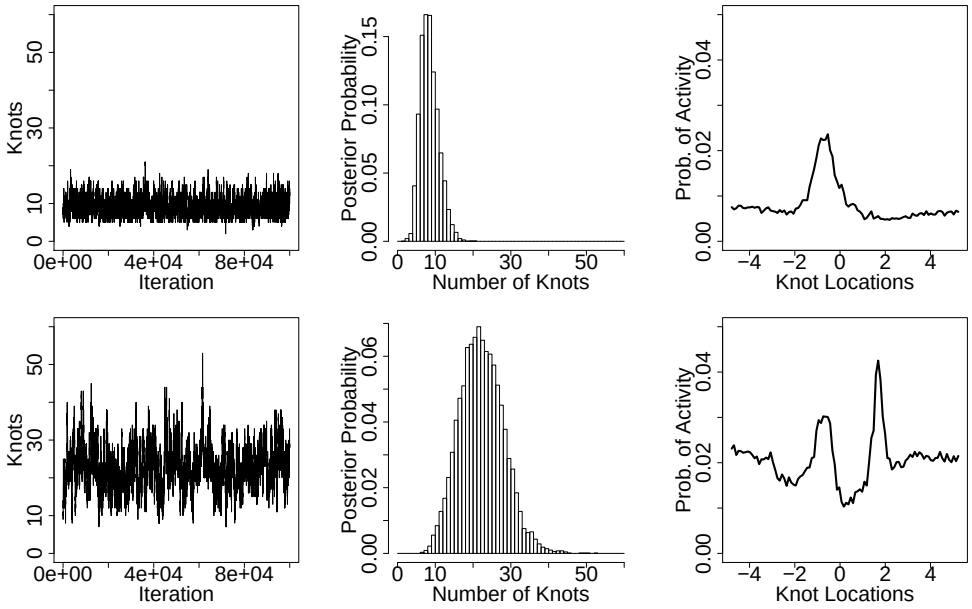

Figure 2: Number and locations of knots in the spline fits of the survival probability for the simulated data. The upper row shows results using the Poisson(25) prior distribution on the number of knots and the lower using the Poisson(75) prior. The plots illustrate, from left to right, the number of knots on each MCMC iteration, the proportion of knot locations occupied on each iteration, and the proportion of iterations for which each potential knot location is occupied.

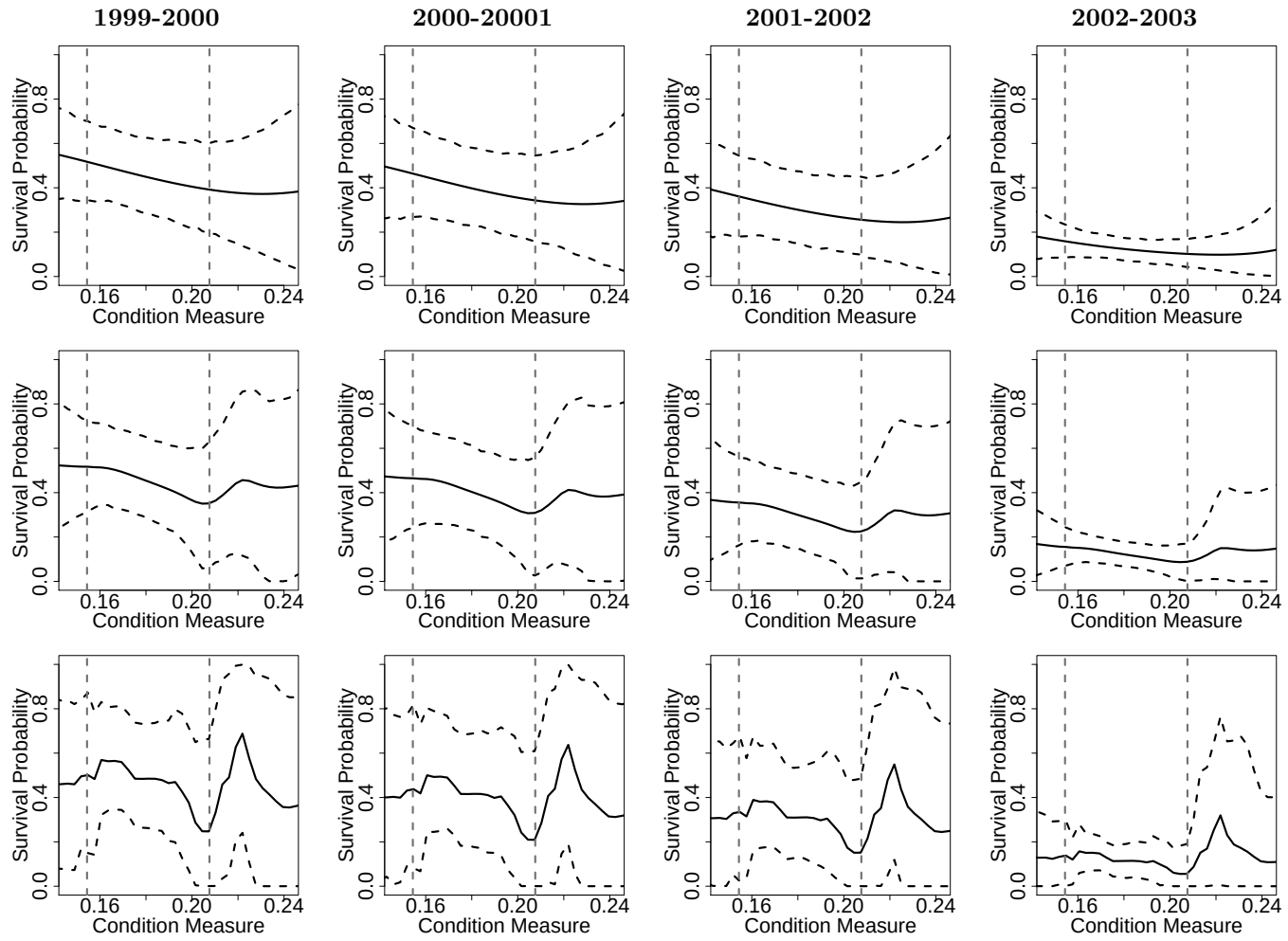

Figure 3: Estimated survival probabilities as a function of condition for the reed warbler data. Estimates from the cubic model are shown in the top row, from the spline model with Poisson(25) prior on the number of knots in the middle, and from the spline model with Poisson(75) prior on the bottom. Solid black lines indicate the pointwise posterior mean and dashed lines the bounds of the posterior pointwise 95% credible intervals. In each plot, the vertical grey dashed line indicate the 2.5-th and 97.5-th percentiles of the observed condition values. The plots for 2002-2003 actually estimate the product $\phi_4(z)p_5$ because of the weak identifiability of these parameters separately.

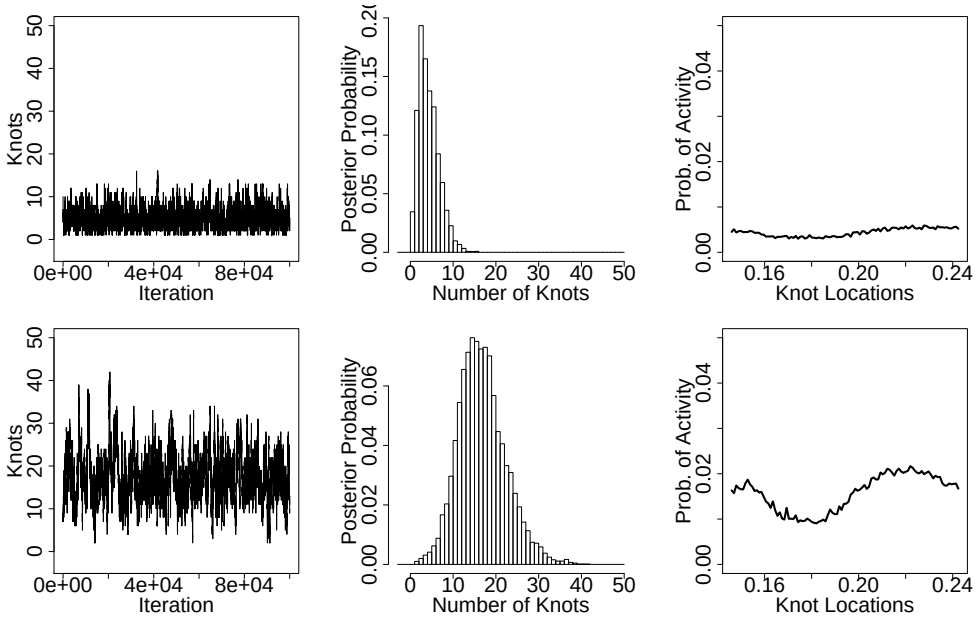

Figure 4: Number and locations of knots in the spline fits of the survival probabilities for the reed warbler data. The upper row shows results using the Poisson(25) prior distribution on the number of knots and the lower using the Poisson(75) prior. The plots illustrate, from left to right, the number of knots on each MCMC iteration, the proportion of knot locations occupied on each iteration, and the proportion of iterations for which each potential knot location is occupied.

Parameter	True Value	Cubic Polynomial	Poisson(25) Prior	Poisson(75) Prior
p_1	0.85	0.85(0.77,0.91)	0.84(0.76,0.90)	0.84(0.76,0.90)
p_2	0.85	0.84(0.73,0.96)	0.82(0.72,0.93)	0.80(0.70,0.90)
μ_1	0.00	-0.06(-0.23,0.10)	-0.06(-0.23,0.10)	-0.06(-0.23,0.10)
μ_2	0.00	0.04(-0.09,0.17)	0.04(-0.09,0.17)	0.04(-0.09,0.17)
σ	1.00	1.03(0.96,1.10)	1.03(0.96,1.10)	1.03(0.96,1.10)

Table 1: Estimates of capture probabilities and parameters of the covariate distribution for the simulated data. Results on the left are from the model assuming a cubic relationship between survival and the covariate; on the right are from the spline models. Estimates are given by the posterior mean with 95% HPD credible interval.

Quantity	CJS Model	Cubic Polynomial	Poisson(25) Prior	Poisson(75) Prior
$\phi_1(0.17)$	0.48(0.33,0.67)	0.47(0.32, 0.64)	0.48(0.32, 0.68)	0.52(0.32, 0.78)
$\phi_2(0.17)$	0.41(0.27,0.63)	0.42(0.25, 0.60)	0.43(0.26, 0.65)	0.45(0.23, 0.71)
$\phi_3(0.17)$	0.33(0.19,0.55)	0.32(0.17, 0.49)	0.32(0.17, 0.50)	0.34(0.14, 0.58)
$\phi_4(0.17)p_5$	0.13(0.09,0.19)	0.13(0.08, 0.19)	0.14(0.08, 0.19)	0.13(0.06, 0.20)
p_2	0.45(0.29,0.62)	0.46(0.29, 0.64)	0.45(0.29, 0.63)	0.42(0.26, 0.60)
p_3	0.46(0.28,0.66)	0.46(0.28, 0.67)	0.45(0.27, 0.66)	0.43(0.24, 0.65)
p_4	0.46(0.25,0.71)	0.48(0.26, 0.73)	0.48(0.26, 0.72)	0.44(0.23, 0.70)
μ_1	NA	0.000(-0.004,0.004)	0.000(-0.004,0.004)	0.000(-0.004,0.004)
μ_2	NA	0.000(-0.005,0.004)	-0.001(-0.006,0.004)	-0.001(-0.006,0.004)
μ_3	NA	-0.002(-0.007,0.004)	-0.002(-0.007,0.003)	-0.002(-0.007,0.002)
μ_4	NA	-0.002(-0.010,0.007)	-0.001(-0.010,0.007)	-0.003(-0.011,0.005)
σ	NA	0.012(0.010,0.014)	0.012(0.010,0.014)	0.012(0.010,0.014)

Table 2: Estimates of capture probabilities and parameters of the covariate distribution for the reed warbler data. Results on the left are from the cjs model assuming no effect of the covariate on survival, then from a model assuming a cubic relationship between survival and the covariate, from the spline model with Poisson(25) prior on the number of knots, and on the right are from the spline model with Poisson(75) prior on the number of knots. Posterior means with 95% credible intervals are provided for each parameter. Estimated survival probabilities are computed at the median observed value of condition, .17 g/mm. Note that p_5 is almost confounded with $\phi_4(z)$ and the their product was estimated as a single parameter.