

# The Efficient Semiparametric Regression Modeling of Capture–Recapture Data: Assessing the Impact of Climate on Survival of Two Antarctic Seabird Species

Olivier Gimenez and Christophe Barbraud

**Abstract** A nonparametric approach has recently been proposed for estimating survival in capture–recapture models, which uses penalized splines to achieve flexibility in exploring the relationships with environmental covariates. However, this method is highly time-consuming because it is implemented through a fully Bayesian approach using Markov chain Monte Carlo simulations. To cope with this issue, we developed a two-step approach in which the existing method is used in conjunction with a multivariate normal approximation to the capture–recapture data likelihood. The ability of our approach to capture various nonlinearities in demographic parameters was validated by carrying out a simulation study. Two examples dealing with Snow petrel and Emperor penguin capture–recapture data sets were also considered to illustrate our procedure, including the relationship between survival rate, population size and climatic covariates.

**Keywords** Auxiliary variables · Bayesian inference · Bivariate smoothing · Computational efficiency · Demographic rates · Environmental covariates · Interactions · Multivariate normal approximation · Penalized-splines · WinBUGS

## 1 Introduction

Climate change, specifically global warming, is projected to accelerate in the next century (IPCC 2001). Consequences of this on the functioning of ecosystems are at present difficult to predict, and the study of climatic fluctuations on populations is a major topic in ecology (Hughes 2000; McCarty 2001; Stenseth et al. 2002). Recent investigations show that global warming affects some animal populations, through changes in their physiology, phenology, distribution and demography (Hughes 2000; Walther et al. 2002; Root et al. 2003; Walther et al. 2005). The vast majority of studies assume that the potential effects of both climate and population

---

O. Gimenez (✉)  
Centre d'Ecologie Fonctionnelle et Evolutive/CNRS, UMR 5175, 1919 Route de Mende, 34293  
Montpellier Cedex 5, France  
e-mail: olivier.gimenez@cefe.cnrs.fr

D.L. Thomson et al. (eds.), *Modeling Demographic Processes in Marked Populations*,  
Environmental and Ecological Statistics 3, DOI 10.1007/978-0-387-78151-8\_3,  
© Springer Science+Business Media, LLC 2009

density on demographic parameters are linear. However, there is strong evidence that environmental factors may affect population dynamics in more complex ways. For instance, using a global proxy to describe climatic conditions (such as the North Atlantic Oscillation) may induce nonlinear relationships as a consequence of similar nonlinear relations between the proxy and local climatic variables (Mysterud et al. 2001). Empirical data that can be used to investigate the effects of climate change on populations is increasing. Yet, at present there is insufficient modeling methodology to investigate nonlinear relationships between environmental covariates and demographic rates, and to create reliable predictions concerning the impact that the anticipated changes might have on populations.

In this paper, we focus on a new nonparametric approach which has recently been developed to model flexible nonlinear relationships between environmental covariates and demographic rates assessed using capture–recapture/recovery models (Gimenez et al. 2006a). In the spirit of Generalized Additive Models (Hastie and Tibshirani 1990), the shape of the relationship is determined by the data without making any prior assumption regarding its form, using penalized splines (P-splines; Ruppert et al. 2003). However, the whole approach is implemented in a Bayesian framework using MCMC algorithms, and our experience shows that the model fitting process may be highly time-consuming, which can be an obstacle to model selection and to comparative analyses of the response of several species' population dynamics to environmental factors.

Here, we propose to overcome this difficulty by the use of multivariate normal approximation to the capture–recapture model likelihood in a first step (Lebreton et al. 1995; Besbeas et al. 2003). This approximation is then used in a second step in conjunction with a Bayesian approach using MCMC methods in order to implement the P-splines. This combination allows purpose-built programs (e.g. M-SURGE, Choquet et al. 2005; or MARK, White and Burnham 1999) to be used for analyzing capture–recapture data with maximum flexibility and results in a considerable reduction in the computational burden. To validate the ability of our approach to capture various nonlinearities in demographic parameters, we carry out a simulation study. Two examples are also considered to illustrate our approach, including the relationship between survival rate, population size and climatic covariates. Using this new approach we reanalyzed two capture–recapture data sets of Antarctic seabirds, for which previous analyses have investigated (and found) linear relationships between survival and environmental covariates (Jenouvrier et al. 2005). For the Snow petrel (*Pagodroma nivea*), we analyzed the nonlinear relationships between sex-specific adult survival and the Southern Oscillation Index (SOI). For the Emperor penguin (*Aptenodytes forsteri*), we investigated nonlinear relationships between sex specific adult-survival, sea ice extent and population size.

## 2 Efficient Nonparametric Regression in Capture–Recapture Modeling

In this section, we introduce our approach following two steps. First, the data are analyzed using standard capture–recapture models in a Frequentist framework.

The survival parameter estimates and the associated estimated variance–covariance matrix are then used to approximate the likelihood of the model best supported by the data (*Step 1*). This allows us to adopt a Bayesian approach using MCMC algorithms to implement the nonparametric approach using P-splines (*Step 2*).

### 2.1 Step 1: Handling the Capture–Recapture Data

We used standard capture–recapture models (Lebreton et al. 1992) to get maximum likelihood estimates (MLEs) for the probability  $\phi_i$  that an individual survives to occasion  $i + 1$  given that it is alive at time  $i$ , and for the probability  $p_j$  that an individual is recaptured at time  $j$ . All models were fitted using program M-SURGE (Choquet et al. 2005), but program MARK could have been used instead (White and Burnham 1999).

Using program U-CARE (Choquet et al. 2003), we assessed the fit of the most general time-dependent CJS model to determine whether it provided an adequate description of the data. In both examples (see the Section 2.4), we detected a trap-dependence effect on capture (Pradel 1993), meaning that capture probability at occasion  $j + 1$  was different for individuals captured at occasion  $j$  than for individuals not captured at occasion  $j$ . Such a trap-dependent effect in long-lived species is common and partly reflects heterogeneity in the quality of individuals in a population. For emperor penguins and snow petrels, trap-dependence was at least partly caused by heterogeneity between individuals in their capacity to breed at the colony every year and therefore to be captured. Consequently, we used a multistate capture–recapture model to cope with this departure from the null hypothesis that the CJS fits the data (Gimenez et al. 2003). We distinguished two states whether a capture occurred on the prior occasion (say state A) or not (say state B). In practice, we considered a separate formulation (i.e., the transition probabilities are split into survival and movement probabilities – see Hestbeck et al. 1991). The survival probabilities were time-dependent while the capture probabilities in the states A and B were set constant and fixed to 1 and 0 respectively, and the transition probabilities were state- and time-dependent. By using this formulation, the transition probabilities between states A and A were the capture probabilities given a capture on the prior occasion, and the transition probabilities between states B and A were the capture probabilities given no capture on the prior occasion. See Gimenez et al. (2003) for further details. If any lack of fit remained, we applied a correction to the estimates and their estimated variance–covariance based on the calculation of the coefficient of overdispersion (Lebreton et al. 1992).

As is seen above, we conducted modeling in two steps (Lebreton et al. 1992). We first focused on a model that described the nuisance parameter – i.e., the capture probabilities – in the most parsimonious way, while survival remained time-dependent. Then, preserving the most parsimonious structure of the nuisance parameters, we worked out the survival probabilities using P-splines. Note that for simplicity, we analyzed males and females separately for both data sets (see Section 2.4).

### Recapture

ps. First, the data  
entist framework.

We now turn to the approximation of the capture–recapture likelihood, which will be denoted  $L(\phi, p)$ . Lebreton et al. (1995) and Besbeas et al. (2003) proposed to use a multivariate normal to approximate the function  $L(\phi, p)$ . More precisely, the maximum likelihood estimates of the parameters on the logit scale,  $\hat{\theta}$ , and the associated estimated variance–covariance matrix,  $\hat{\Sigma}$ , both obtained from fitting an appropriate capture–recapture model (see above), are used to approximate the log-likelihood as:

$$2 \log \{L(\phi, p)\} = \text{constant} - (\hat{\theta} - \theta)^T \hat{\Sigma}^{-1} (\hat{\theta} - \theta). \quad (1)$$

Note that Besbeas et al. (2003) showed that it is only necessary to make the approximation for the parameters of interest, which are the survival probabilities in our case. Obviously, using a multivariate normal distribution in place of the usual product of multinomial distributions (where cells are complex nonlinear functions of the survival and recapture probabilities) results in a much simpler form for the likelihood  $L(\phi, p)$ , which in turn greatly speeds up the Bayesian fitting process using MCMC algorithms.

Nevertheless, the use of Eq. (1) may be made difficult by numerical issues. Indeed, some parameters may be estimated close to or on a boundary (0 or 1 as we are dealing with probabilities), resulting in the impossibility to properly quantify the variability associated to the MLEs using standard methods. Technically, the dispersion matrix  $\hat{\Sigma}$  is ill-conditioned which prevents us from obtaining its inverse as required in Eq. (1). We circumvent this issue by neglecting the covariances, and considering the diagonal  $\hat{\Sigma}$  matrix of the estimated variances with off-diagonal terms all zeros. Still, calculating variances for boundary estimates remains problematic. One option is to use profile-likelihood intervals (Gimenez et al. 2005), the problem being that this approach does not formally provide a point estimate nor a standard error. In this paper, we decided to assign a large variance (10,000) to those parameters estimated close to or on the boundary, thus affecting relative negligible weights to the corresponding MLEs (see Eq. (1)). This ad-hoc procedure was used in the Section 2.4 only.

## 2.2 Step 2: Semiparametric Modeling of the Survival

### 2.2.1 Univariate Smoothing

We consider the following regression model for the survival probability  $\phi_i$ :

$$\text{logit}(\phi_i) = \log\left(\frac{\phi_i}{1 - \phi_i}\right) = f(x_i) + \varepsilon_i, \quad (2)$$

where  $x_i$  is the value of the covariate applying between occasions  $i$  and  $i + 1$ ,  $f$  is a smooth function and  $\varepsilon_i$  are i.i.d. random effects  $N(0, \sigma_\varepsilon^2)$ . The function  $f$  specifies a nonparametric flexible relationship between the survival probability and

the covariate that allows nonlinear environmental trends to be detected. Following Gimenez et al. (2006a), we used a truncated polynomial basis to handle  $f$ :

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_P x^P + \sum_{k=1}^K b_k (x - \kappa_k)_+^P, \quad (3)$$

where  $x$  is the covariate,  $\beta_0, \beta_1, \dots, \beta_P, b_1, \dots, b_K$  are regression coefficients to be estimated,  $P \geq 1$  is the degree of the spline,  $(u)_+^P = u^P$  if  $u \geq 0$  and 0 otherwise, and  $\kappa_1 < \kappa_2 < \dots < \kappa_K$  are fixed knots. We considered  $K = \min(\frac{1}{4}I, 35)$  knots to ensure the desired flexibility, and let  $k_k$  be the sample quantile of  $x$ 's corresponding to probability  $\frac{k}{K+1}$ . To avoid overfitting, we penalized the  $b$ 's by assuming that the coefficients of  $(x - \kappa_k)_+^P$  are normally distributed random variables with mean 0 and a certain variance  $\sigma_b^2$  to be estimated. This is the reason why this approach is referred to as penalized splines (Ruppert et al. 2003). For further details see Gimenez et al. (2006a) and references therein.

### 2.2.2 Bivariate Smoothing

To incorporate the interaction between two continuous environmental covariates, we opted for bivariate smoothing using thin-plate splines (Green and Silverman 1994). The main challenge here was to achieve the ideal balance between roughness and smoothness, which is controlled by a parameter  $\delta$  usually referred to as the smoothing parameter. We considered the restricted maximum likelihood (REML) criterion to choose this amount of smoothing using the data (Searle et al. 1992), which allows the whole modeling exercise to be easily implemented in a mixed model framework (Ruppert et al. 2003; Crainiceanu et al. 2005; Gimenez et al. 2006a). Specifically, we consider a nonparametric model for the survival with respect to environmental covariates as follows:

$$\text{logit}(\phi_i) = f(\mathbf{x}_i) + \varepsilon_i \quad (4)$$

where  $\mathbf{x}_i = (x_i^1, x_i^2)^T$  is the value of the vector of two covariates  $\mathbf{x}^1$  and  $\mathbf{x}^2$  for year  $i$ ,  $T$  denotes transpose,  $\varepsilon_i$  are i.i.d  $N(0, \sigma_\varepsilon^2)$  and  $f$  is a smooth function. Because they have good numerical properties, we used radial basis functions to handle  $f$  (Ruppert et al. 2003):

$$f(\mathbf{x}) = \mathbf{X}\mathbf{b} + \mathbf{Z}_K\mathbf{v}, \quad (5)$$

where  $\{1, x_i^1, x_i^2\}$  is the  $i$ th row of matrix  $\mathbf{X}$ ,  $\{C(\|\mathbf{x}_i - \kappa_1\|), \dots, C(\|\mathbf{x}_i - \kappa_K\|)\}$  is the  $i$ th row of matrix  $\mathbf{Z}_K$ , the  $\kappa_k$ 's are bi-dimensional vectors of fixed knots, the function  $C(\|\mathbf{r}\|) = \|\mathbf{r}\|^2 \log \|\mathbf{r}\|$  with  $\|\mathbf{r}\| = \sqrt{\mathbf{r}^T \mathbf{r}}$  handles the nonlinear structure of the survival surface;  $\mathbf{b} = (b_1, b_2, b_3)^T$  and  $\mathbf{v} = (v_1, \dots, v_K)^T$  are vectors of fixed and random regression parameters respectively to be estimated with  $\text{Cov}(\mathbf{v}) = \sigma_u^2 \mathbf{\Omega}_K^{-1}$

where  $\Omega_K$  has  $(k, k')$ th element  $C(\|\kappa_k - \kappa_{k'}\|)$ . Using the re-parameterization  $u = \Omega_K^{1/2} v$  and defining  $Z = Z_K \Omega_K^{-1/2}$ , Eq. (5) becomes equivalent to

$$f(x) = \mathbf{Xb} + \mathbf{Zu}, \quad (6)$$

where  $\mathbf{u}$  is assumed to be normally distributed, independent from  $\varepsilon$ , with  $\text{Cov}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_K$ . It can be shown that the optimal amount of smoothing using the REML criterion is given by  $\delta = \sigma_u^2 / \sigma_\varepsilon^2$ , which turns out to be also the case in the univariate smoothing (Ruppert et al. 2003). To choose the number and the location of the knots, we considered  $K = \max\{20, \min(I/4, 150)\}$  knots as suggested by Ruppert et al. (2003) and used the space-filling algorithm of Nychka and Saltzman (1998) to select the location of these knots. This algorithm automatically places knots in regions with high density of observed values while maximizing the average spacing between knots of those regions. Finally, to plot the fitness surface, we obtained contours and perspectives views by generating a  $30 \times 30$  grid of predicted values.

### 2.2.3 Bayesian Inference

Vague prior distributions were provided for all parameters. Specifically, we chose uniform distributions on  $[0,1]$  for the detection probabilities, normal distributions with mean 0 and variances 1,000 for the  $\beta$ 's and normal distributions with mean 0 and variances  $\sigma_u^2$ ,  $\sigma_b^2$  and  $\sigma_\varepsilon^2$  for the  $u$ 's,  $b$ 's and  $\varepsilon$ 's respectively. The priors for the hyperparameters  $\sigma_u^2$ ,  $\sigma_b^2$  and  $\sigma_\varepsilon^2$  were chosen as inverse-gamma with both parameters equal to 0.001. We generated two chains of length 100,000, discarding the first 50,000 as burn-in. Convergence was assessed using the Gelman and Rubin statistic which compares the within to the between variability of chains started at different and dispersed initial values (Gelman 1996). All covariates were standardized to improve convergence. The simulations were performed using WinBUGS (Spiegelhalter et al. 2003). The R (Ihaka and Gentleman 1996) package R2WinBUGS (Sturtz et al. 2005) was used to call WinBUGS and export results in R. To implement the space-filling algorithm, we used the R package FIELDS (Fields Development Team 2006).

Whenever needed, we used the Deviance Information Criterion (DIC; Spiegelhalter et al. 2002) to discriminate between candidate models: the smaller the DIC value, the better the model. We acknowledge that the DIC is somewhat controversial in the statistical literature, and should be used with caution (see Spiegelhalter et al. 2002 and Celeux et al. 2006 and the discussion papers following these two papers). The R and WinBUGS codes are available on request from the first author.

## 2.3 Simulation Study

We conducted a simulation study to investigate the performance of our approach, in particular to check that the use of the approximation for the capture-recapture likelihood did not affect the estimation of parameters. Following Gimenez et al. (2006a),

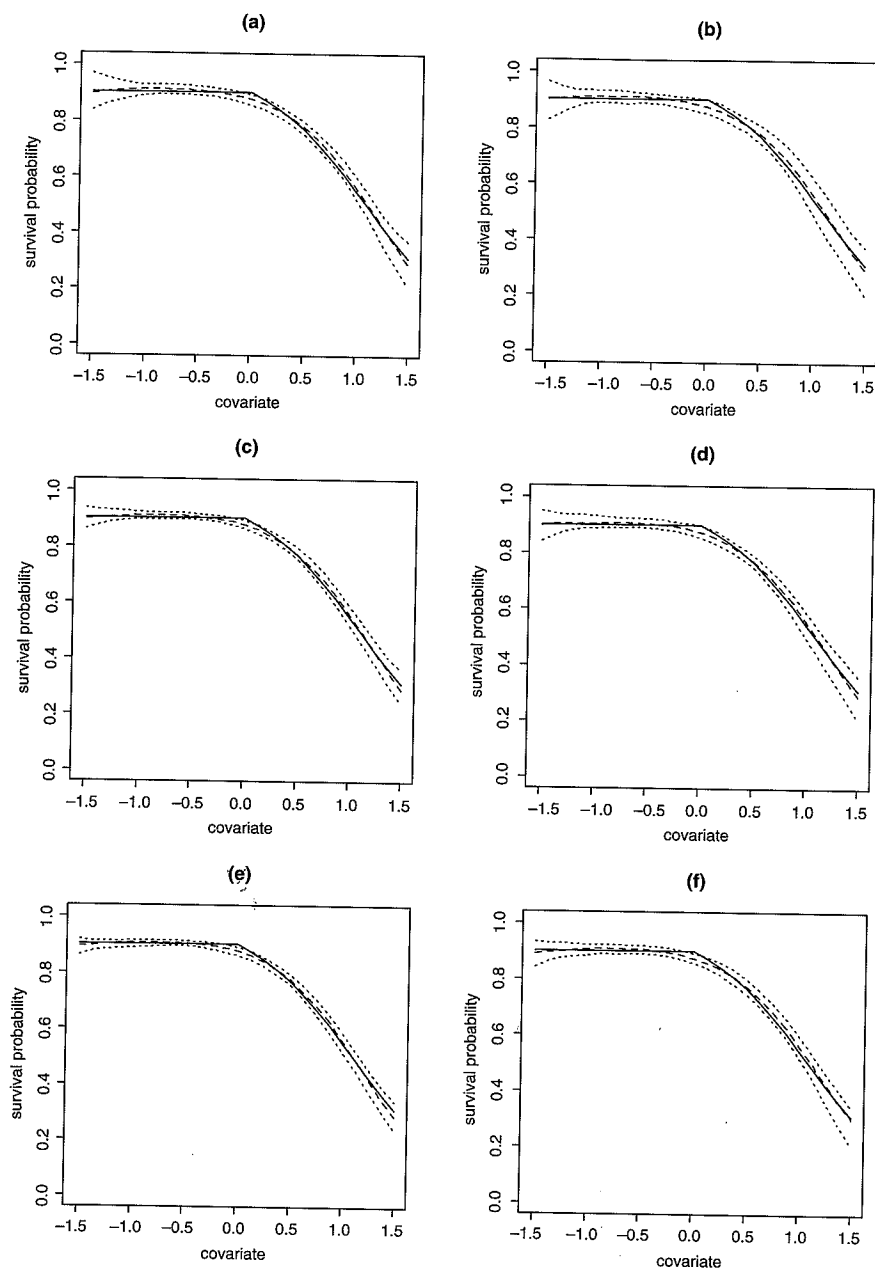
we considered two scenarios with different forms for the underlying nonlinear regression function  $f$  of Eq. (3). Study 1 used the regression function  $f(x) = 2.2$  if  $x \leq -0.06$  and  $f(x) = 2.08 - 2x$  otherwise to represent a threshold effect. The  $x$ 's were equally spaced on  $[-1.5; 1.5]$ . Study 2 used the regression function  $f(x) = 1.5g\left(\frac{x-0.35}{0.15}\right) - g\left(\frac{x-0.6}{0.1}\right)$  where  $g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$  to represent complex non-linear patterns. The  $x$ 's were equally spaced on  $[0; 1]$ . For both studies, we simulated 100 capture-recapture data sets covering 26 sampling occasions, so that 25 survival probabilities had to be estimated. We considered 50, 100 and 250 newly released individuals per occasion and two levels of variability with  $\sigma_\epsilon^2 = 0.02$  or  $\sigma_\epsilon^2 = 0.1$ . The capture probability was set constant and equal to 0.7.

For each data set, we applied our approach in two steps, first fitting a capture-recapture model with time-dependent survival probabilities and constant recapture probabilities, second using the MLEs and the variance-covariance matrix to approximate the capture-recapture likelihood of this model in order to implement the P-splines in a Bayesian framework using MCMC algorithms. Details on the practical implementation can be found in the Section 2. For each  $x$  value, we computed the median along with a 95% confidence interval for the posterior medians of  $f$  and then back-transformed in order to compare the estimated survival curve to its true counterpart. The results are displayed in Figs. 1 and 2, showing that our two-step approach does a good job in capturing the nonlinearities in the survival vs. covariate relationship. For a fixed number of newly released individuals, the greater the variance the lower the precision (both Figs. 1 and 2, left column - low variability vs. right column - high variability), the difference being clearer for Study 1. When the sample size increases, the precision gets better (both Figs. 1 and 2, going down - 50, 100 and 250 newly released individuals), although for high variability the gain was not substantial (right column in both Figs. 1 and 2). Overall, as noted by Gimenez et al. (2006a), the relationship in Study 1 was more precisely estimated than that of Study 2.

## 2.4 Applications

### 2.4.1 Snow Petrels

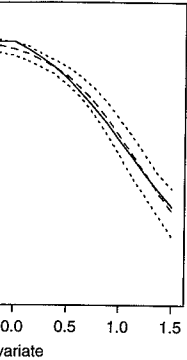
As a first example, we analyzed the data used in Gimenez et al. (2006a) to illustrate the full Bayesian implementation of the semiparametric modeling of survival probabilities. The data were obtained in a 40-year study on individually marked Snow petrels, nesting at Petrels Island, Terre Adélie, from 1963 to 2000 (see also Barbraud et al. 2000; Jenouvrier et al. 2005). We considered the Southern Oscillation Index (*SOI*) as a proxy of the overall climate condition, available from the Climatic Research Unit (<http://www.cru.uea.ac.uk/cru/data/soi.htm>). In total, we considered 563 female and 561 male capture histories (more than in Gimenez et al. 2006a who were limited by the computational burden).



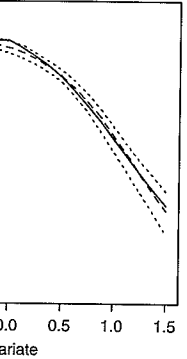
**Fig. 1** Performance of the nonparametric approach for estimating nonlinearities in the survival probability – Study 1 (see the Section 2.3 for details). We used 100 simulated capture–recapture data sets with 50, 100 and 250 newly released individuals per occasion (from top to bottom resp.) and two levels of variability,  $\sigma_\epsilon^2 = 0.02$  or  $\sigma_\epsilon^2 = 0.1$  (from left to right resp.). The solid line is the true regression function, the dashed line is the median of the 100 estimated posterior medians and the dotted lines indicate the associated 95% confidence interval



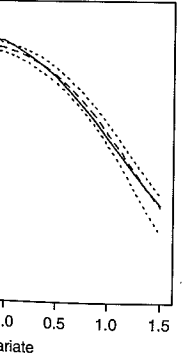
(b)



(d)

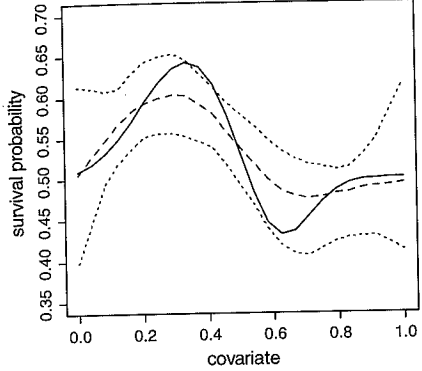


(f)

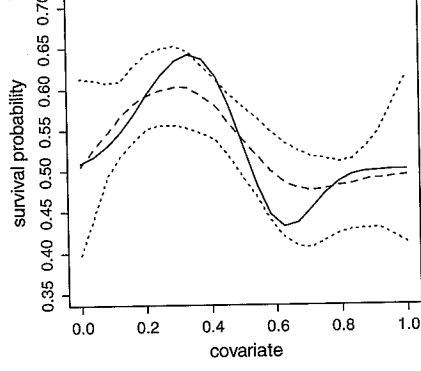


...ities in the survival  
 ...ed capture-recapture  
 ... top to bottom resp.)  
 ... The solid line is the  
 ... posterior medians and

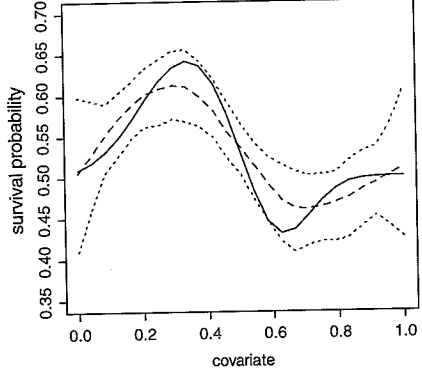
(a)



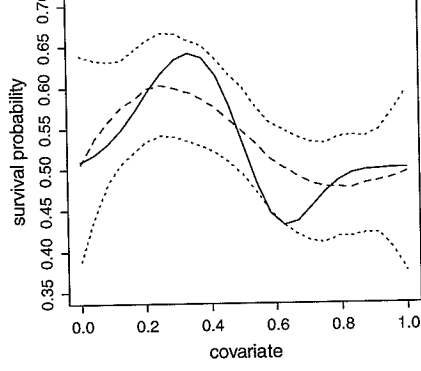
(b)



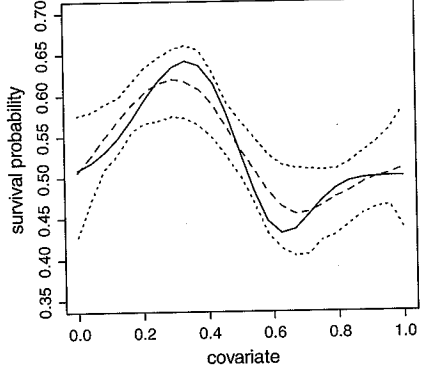
(c)



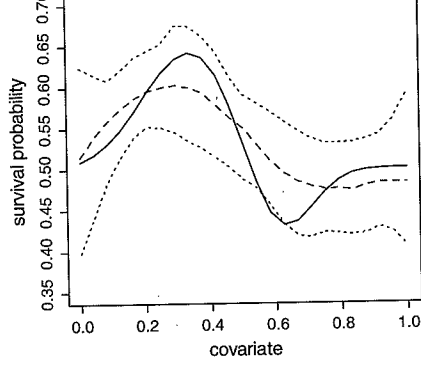
(d)



(e)



(f)



**Fig. 2** Performance of the nonparametric approach for estimating nonlinearities in the survival probability – Study 2 (see the Section 2.3 for details). We used 100 simulated capture-recapture data sets with 50, 100 and 250 newly released individuals per occasion (from top to bottom resp.) and two levels of variability,  $\sigma_\epsilon^2 = 0.02$  or  $\sigma_\epsilon^2 = 0.1$  (from left to right resp.). The solid line is the true regression function, the dashed line is the median of the 100 estimated posterior medians and the dotted lines indicate the associated 95% confidence interval

*Step 1.* After removing the first capture to remove heterogeneity, the CJS model still poorly fitted the data for both females and males (females:  $\chi^2_{167}(349.98) < 0.01$ ; males:  $\chi^2_{169}(472.20) < 0.01$ ), and a closer inspection of the results revealed that a large part of the CJS  $\chi^2$  statistic was explained by a trap-dependence effect (females:  $\chi^2_{34}(175.74) < 0.01$ ; males:  $\chi^2_{34}(248.697) < 0.01$ ). The goodness-of-fit for the model with trap-dependence was still significant (females:  $\chi^2_{133}(174.239) < 0.01$ ; males:  $\chi^2_{135}(223.503) < 0.01$ ) so we used a lack-of-fit coefficient for further analyses (females: 1.3; males: 1.7). Time-dependent survival probability estimates and the estimated variance-covariance were then obtained for both sexes using M-SURGE (Choquet et al. 2005).

*Step 2.* First, because sex differences in the survival probabilities were found before, we considered a model with an additive effect of both SEX and SIE factors. This was achieved by extending the nonparametric approach introduced above to allow a predictor to enter the model linearly (we will refer to semiparametric modeling when both linear and nonlinear effects appear in a model). To do so, we wrote:

$$\text{logit}(\phi_i^l) = \beta_0 + \gamma \text{SEX} + \beta_1 \text{SOI}_i + \sum_{k=1}^K b_k (\text{SOI}_i - \kappa_k) + \varepsilon_i, \quad (7)$$

where  $\phi_i^l$  is the survival probability between occasion  $i$  and  $i + 1$  for  $l = \text{males}$  ( $\text{SEX} = 0$ ) or  $l = \text{females}$  ( $\text{SEX} = 1$ ). Interestingly, only little adjustments to the modeling introduced in Section 2.2.1 are needed to specify the model defined by Eq. (7) (see Gimenez et al. 2006a). We also fitted a model with an interaction effect between the SEX and the SOI factors. It basically consists of considering different smooth functions according to the SEX qualitative variable (Coull et al. 2001). Table 1 shows that the model with an additive effect of both covariates is preferred to the model with interaction.

Finally we considered two further models corresponding to two biological hypotheses. First, we were interested in assessing the significance of the SEX effect, so we fitted a model without the SEX effect, while keeping the nonparametric feature of the model. This model performs better than the two models having the SEX effect (Table 1). This was also confirmed by the 95% posterior credible interval  $[-0.49; 0.15]$  of the parameter  $\gamma$  which contains 0. Second, we were interested in testing for the presence of nonlinearities in the survival probability. One way to answer this question was to fit a model with a linear effect of the SOI covariate

**Table 1** Models fitted to the Snow petrel data. DIC is the deviance information criterion, and pD the number of effective parameters.  $\Delta\text{DIC}$  is the difference between the DIC of a model and the DIC for the minimum DIC model. The model best fitting the data is in bold font

Model	DIC	pD	$\Delta\text{DIC}$
Additive effect of SEX and SOI	1129.29	1062.99	604.74
Interaction effect of SEX and SOI	1644.82	1595.38	1120.27
SOI effect only (no SEX effect)	679.27	607.34	154.72
<b>Linear effect of SOI (no SEX effect)</b>	<b>524.55</b>	<b>446.49</b>	<b>0</b>

upon the survival probability, and to compare with its nonparametric counterpart. To do so, we used:

$$\text{logit}(\phi_i) = \beta_0 + \beta_1 \text{SOI}_i + \varepsilon_i. \quad (8)$$

As already noted by Gimenez et al. (2006a), the relationship between the climatic covariate SOI and the Snow petrel survival seems to be linear (Table 1). The graphical representation of the two latter models tends to confirm this result (Fig. 3). During negative SOI, characteristic of El Niño episodes, cooler waters in the western part of the tropical Pacific and southern Australia down to the Ross Sea region seem to favor enhanced productivity in this oligotrophic area (Wilson and Adamec 2002). Therefore these oceanographic conditions may increase the food availability for snow petrels and reduce their mortality risk associated with starvation. However, the effect of SOI on adult survival is small, with only a 1–2% difference in survival between negative and positive SOI conditions, which might explain the linear relationship between survival and SOI. We will go back to the issue of formally detecting nonlinearities in Section 2.5.

In this section, we have considered an interaction between a discrete variable SEX and a continuous variable SOI. In the next section, we consider an interaction between two continuous variables using bivariate smoothing (Ruppert et al. 2003).

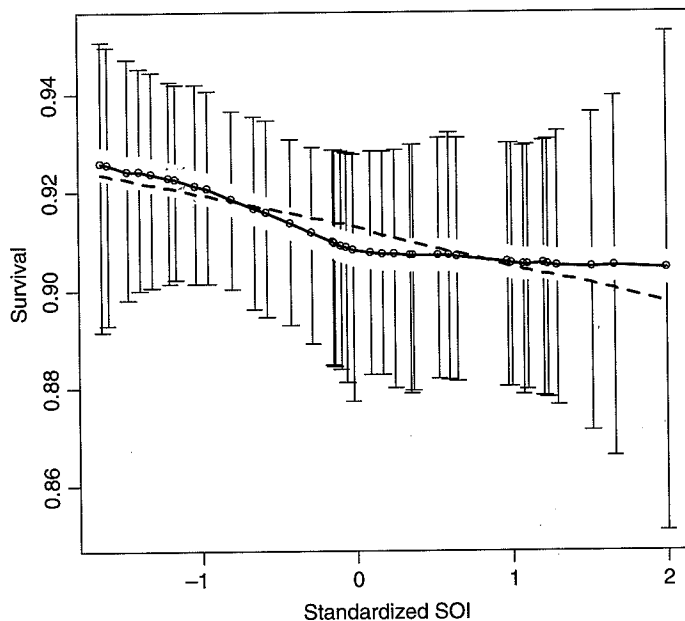


Fig. 3 Annual variations in Snow petrel survival as a function of the standardized SOI using a nonparametric model. Medians (solid line) with 95% pointwise credible intervals (vertical solid lines) are shown, along with the estimated linear effect (dotted line)

ez and C. Barbraud

, the CJS model  
 $\chi^2_{167} (349.98) <$   
 results revealed  
 dependence effect  
 the goodness-of-fit  
 $\chi^2_{133} (174.239) <$   
 efficient for further  
 ability estimates  
 both sexes using

ties were found  
 and SIE factors.  
 induced above to  
 semiparametric  
 l). To do so, we

$+ \varepsilon_i, \quad (7)$

for  $l =$  males  
 adjustments to  
 the model defined  
 with an interaction  
 of considering  
 able (Coull et al.  
 with covariates is

biological hypo-  
 SEX effect, so  
 parametric feature  
 having the SEX  
 credible interval  
 are interested in  
 ty. One way to  
 SOI covariate

criterion, and pD  
 of a model and the

$\Delta\text{DIC}$
604.74
1120.27
154.72
0

#### 2.4.2 Emperor Penguins

As a second example, we analyzed data on the emperor penguin which consist of data from a long-term study on marked individuals, nesting at Petrels Island, Terre Adélie, from 1962 to 2002 (see Barbraud and Weimerskirch 2001; Jenouvrier et al. 2005). We considered the Sea Ice Extent (SIE) as the distance from the colony to the limit of a 15% or higher sea ice concentration, which was obtained at longitude 140°E using the sea ice data available from the Antarctic CRC and Australian Antarctic Division Climate Data Sets (<http://www.antrc.utas.edu.au/~jacka/seace.C.html>). We also considered the number of breeding pairs (POPSIZE). In total, we considered 382 female and 331 male capture histories.

*Step 1.* After removing the two first captures to remove heterogeneity, the CJS model still poorly fitted the data for both females and males (females:  $\chi_{85}^2(182.05) < 0.01$ ; males:  $\chi_{79}^2(198.12) < 0.01$ ), and a closer inspection of the results revealed that a large part of the CJS  $\chi^2$  statistic was explained by a trap-dependence effect (females:  $\chi_{27}^2(112.07) < 0.01$ ; males:  $\chi_{26}^2(131.85) < 0.01$ ). The goodness-of-fit for the model with trap-dependence indicated that the fit was satisfactory (females:  $\chi_{58}^2(69.98) = 0.135$ ; males:  $\chi_{53}^2(66.28) = 0.104$ ). Time-dependent survival probabilities estimates and the estimated variance-covariance were then obtained for both sexes using M-SURGE (Choquet et al. 2005).

*Step 2.* The results of the bivariate smoothing for male and female Emperor penguins are given in Fig. 4. Overall, females survive better than males, which is in agreement with previous studies (Barbraud and Weimerskirch 2001; Jenouvrier et al. 2005). Now if we look into the relationship between survival and the interaction of the SIE and POPSIZE effects, interesting patterns emerge. Strategies differ by sex. While the survival optimum for both males and females is reached for average values of SIE, there is a marked difference regarding POPSIZE: females prefer very high POPSIZE while males survive better with relatively low POPSIZE. These differences may be interpreted in the light of the contrasting breeding strategies of males and females. After their 3.5 months fast incubating the egg, emaciated males return to sea for feeding and density dependent processes may affect their survival chances through competition for food when POPSIZE is high. This should be particularly accentuated when food resources are scarce, i.e., when sea ice extent is low. During the entire incubation, females are absent from the colony, feeding within the pack ice and below the fast ice. Males at the colony face very harsh climatic conditions and it has been shown that they also form huddles to save energy (Ancel et al. 1997). Therefore, we hypothesize that when the population is large it might be easier to find congeners and to form huddles than when the population is small, which may increase their chances of survival. However, we note that we could not formally assess sex differences since the two data sets were analyzed separately. Interestingly, it is relatively easy to get a picture of the precision associated with the survival surface as a by-product of the use of the MCMC procedure (Fig. 1, right column). Having a visualization of the precision helps us in determining to what extent the patterns we detected are supported by the data. In the present example, the standard deviations are low, except for extreme values of both covariates (Fig. 1, right column).

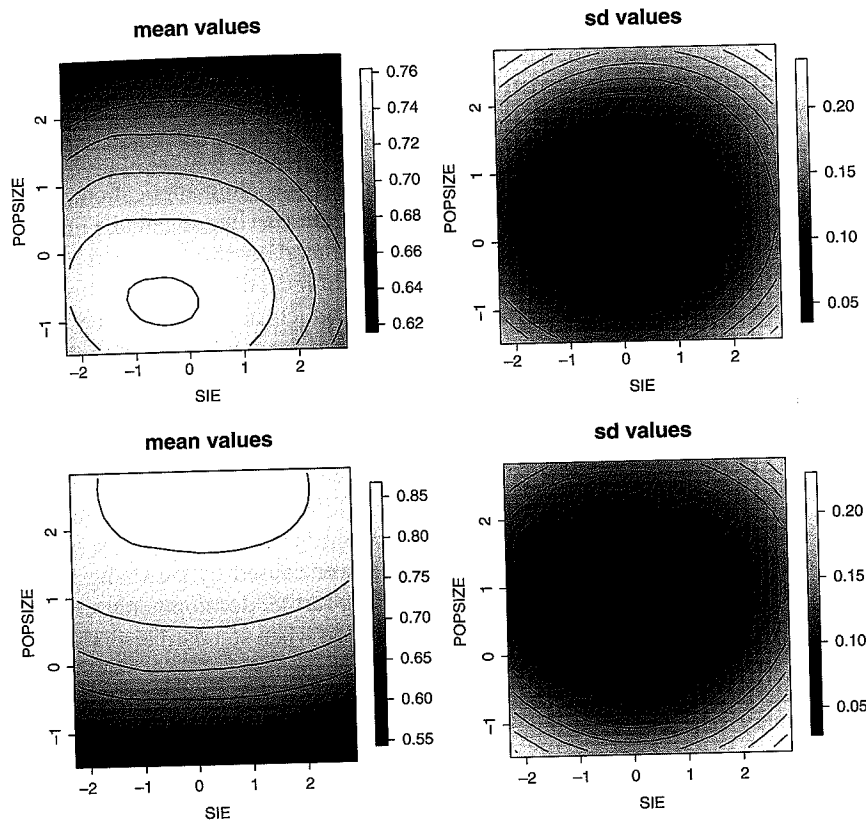


Fig. 4 Bayesian thin-plate spline visualization of the survival surface for the Emperor penguin as a function of the two external covariates sea-ice extent (SIE) and number of breeding pairs (POPSIZE). Posterior mean survival probabilities (left column) and associated posterior standard deviations (right column) are provided for males (top) and females (bottom)

## 2.5 Discussion

In this paper, we have used a combination of the Frequentist and the Bayesian approaches to implement semiparametric modeling of survival probabilities as a function of environmental covariates using capture-recapture data. Instead of opposing the two frameworks and forcing one to make a choice between the two, we have utilized the merits of each of the two approaches: the Frequentist approach was used to handle the capture-recapture data using specialist programs like M-SURGE (Choquet et al. 2005) or MARK (White and Burnham 1999) which allows flexible fitting of complex models including age, cohort and/or site effects; the Bayesian approach was used to avoid making any prior assumption regarding the form of the relationship between the survival and the covariates, while taking benefit of the automatic adjustment of the amount of smoothing in the P-splines. Besides, the combination allows the computational burden to be substantially reduced. For example, it took about 25 hours to fit the semiparametric model of Eq. (7) to the

Snow petrel data with the full Bayesian approach (Gimenez et al. 2006a), while only 5 minutes were required to obtain the MLEs with the estimated variance-covariance matrix and to fit the semiparametric model of Eq. (7) using the multinormal likelihood approximation.

Apart from the gain in time of calculation, the use of a normal approximation to the capture-recapture data likelihood has another appealing application. We can think of using the information published in the literature to investigate the impact of climatic conditions on demographic rates, in the general context of a meta-analysis. The MLEs and the associated standard errors could indeed be extracted from relevant papers and then used to form a likelihood, which in turn, could be used to relate the demographic rates to climatic conditions, for which measurements are often freely available from the Internet. Maximum flexibility in describing those relationships would be assured by the use of the approach advocated here.

Although the results of the simulations are encouraging, the fact that we did not detect a sex effect in the Snow petrel analysis is in contradiction with a previous study (Jenouvrier et al. 2005), although no sex differences were found in earlier studies (Chastel et al. 1993; Barbraud et al. 2000). Possible explanations are very small differences in survival and/or a loss of power caused by assuming that the covariances are all zeros (see Section 2.1). Pending further developments, extensive simulations are needed to assess the loss of precision when standard errors are used in place of the whole estimated variance-covariance matrix.

Regarding the Emperor penguin example, our analysis should be considered as a preliminary step towards a more comprehensive study. We envisage that model selection will be a crucial issue, as we would like to incorporate additional climatic variables (e.g., SOI and SEX) to POPSIZE and SIE, making the number of scenarios numerous. Besides, determining whether nonlinearities are required in the model still needs to be properly addressed. A Reversible-Jump MCMC procedure is a promising solution to that aim (Bonner et al. this volume).

Finally, so far we have considered environmental covariates only, i.e., variables with values changing over time. A semiparametric approach to incorporate individual covariates, i.e., variables with values changing at the individual level, has recently been proposed to assess natural selection on a single quantitative trait (e.g. body mass: Gimenez et al. 2006b) as well as estimating and visualizing fitness surfaces (Gimenez et al. submitted) using capture-recapture data. There is high interest in considering both types of covariates in a model (e.g. Coulson et al. 2001), and the normal approximation might be useful to reduce the computational burden.

## References

- Ancel A, Visser M, Handrich Y, Masman D, Le Maho Y (1997) Energy saving in huddling penguins. *Nature* 385:304-305.
- Barbraud C, Weimerskirch H, Guinet C, Jouventin P (2000) Effect of sea-ice extent on adult survival of an Antarctic top predator: the snow petrel *Pagodroma nivea*. *Oecologia* 125: 483-488.

- Barbraud C, Weimerskirch H (2001) Emperor penguins and climate change. *Nature* 411: 183–186.
- Besbeas P, Lebreton J-D, Morgan BJT (2003) The efficient integration of abundance and demographic data. *Applied Statistics* 52:95–102.
- Celeux G, Forbesy F, Robert CP, Titterton DM (2006) Deviance information criteria for missing data models. *Bayesian Analysis* 1:651–674.
- Chastel O, Weimerskirch H, Jouventin P (1993) High variability in demographic parameters in the Snow petrel *Pagodroma nivea*; a 27-year study. *Oecologia* 94:278–285.
- Choquet R, Reboulet AM, Pradel R, Gimenez O, Lebreton JD (2003) U-CARE user's guide, Version 2.0. Mimeographed document CEFE/CNRS, Montpellier.
- Choquet R, Reboulet AM, Pradel R, Gimenez O, Lebreton JD (2005) M-SURGE: new software specifically designed for multistate capture-recapture models. *Animal Biodiversity and Conservation* 27:207–215.
- Coull BA, Ruppert D, Wand MP (2001) Simple incorporation of interactions into additive models. *Biometrics* 57:539–545.
- Coulson T, Catchpole EA, Albon SD, Morgan BJT, Pemberton JM, Clutton-Brock TH, Crawley MJ, Grenfell BT (2001) Age, sex, density, winter weather, and population crashes in Soay sheep. *Science* 292:1528–1531.
- Crainiceanu C, Ruppert D, Wand MP (2005) Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software* 14(14).
- Fields Development Team (2006) fields: Tools for Spatial Data. National Center for Atmospheric Research, Boulder, CO. <http://www.cgd.ucar.edu/Software/Fields>.
- Gelman A (1996) Inference and monitoring convergence. Pp. 131–143 in Gilks WR, Richardson S, Spiegelhalter DJ (eds.) *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- Gimenez O, Choquet R, Lebreton J-D (2003) Parameter redundancy in multistate capture-recapture models. *Biometrical Journal* 45:704–722.
- Gimenez O, Choquet R, Amor L, Scofield P, Fletcher D, Lebreton et, J-D, Pradel R (2005) Efficient profile-likelihood confidence intervals for capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics* 10:184–196.
- Gimenez O, Crainiceanu C, Barbraud C, Jenouvrier S, Morgan BJT (2006a) Semiparametric regression in capture-recapture Modelling. *Biometrics* 62:691–698.
- Gimenez O, Covas R, Brown CR, Anderson MD, Bomberger Brown M, Lenormand T (2006b) Nonparametric estimation of natural selection on a quantitative trait using capture-mark-recapture data. *Evolution*. 60:460–466.
- Gimenez O, Grégoire A, Lenormand T, Submitted. Estimating and visualizing fitness surfaces using mark-recapture data.
- Green PJ, Silverman BW (1994) *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, New York.
- Hastie T, Tibshirani R (1990) *Generalized additive models*. Chapman and Hall, London.
- Hestbeck JB, Nichols JD, Malecki RA (1991) Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology* 72:523–533.
- Hughes L (2000) Biological consequences of global warming; is the signal already apparent? *Trends in Ecology and Evolution* 15:56–61.
- Ihaka SP, Gentleman EA (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5:299–314.
- IPCC (2001) *Climate change; a synthesis report. A contribution of Working Groups I, II and III to the third assessment report of the IPCC*. Cambridge University Press, Cambridge.
- Jenouvrier S, Barbraud C, Weimerskirch H (2005) Long-term contrasted responses to climate of two Antarctic seabird species. *Ecology* 86:2889–2903.
- Lebreton JD, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case-studies. *Ecological Monographs* 62:67–118.

- Lebreton J-D, Morgan BJT, Pradel R, Freeman SN (1995) A simultaneous survival rate analysis of dead recovery and live recapture data. *Biometrics* 51:1418–1428.
- McCarty JP (2001) Ecological consequences of recent climate change. *Biological Conservation* 15: 320–331.
- Mysterud ACSN, Yoccoz NG, Langvatn R, Steinheim G (2001) Nonlinear effects of large-scale climatic variability on wild and domestic herbivores. *Nature* 410:1096–1099.
- Nychka D, Saltzman N (1998) Design of Air-Quality Monitoring Networks in Case Studies in Environmental Statistics, Lecture Notes in Statistics. Nychka D, Cox L, Piegorsch W (eds.) Springer Verlag, New York.
- Pradel R (1993) Flexibility in survival analysis from recapture data: handling trap-dependence. Pages 29–37 in J.D. Lebreton and North PM (eds.) *Marked Individuals in the Study of Bird Population*. Birkhäuser, Basel.
- Root TL, Price JT, Hall KR, Schneider SH, Rosensweig C, Pounds JA (2003) Fingerprints of global warming on wild animals and plants. *Nature* 421:57–60.
- Ruppert D, Wand MP, Carroll RJ (2003) *Semiparametric Regression*. Cambridge University Press, Cambridge.
- Searle SR, Casella G, McCulloch EC (1992) *Variance Components*. Wiley, New York.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Lind A (2002) Bayesian measures of complexity and fit. *Journal of the Royal Statistical Society, Series B* 64:583–639.
- Spiegelhalter D, Thomas A, Best N, Lunn D (2003) *WinBUGS User Manual*. Version 1.4 (<http://www.mrc-bsu.cam.ac.uk/bugs>). Technical report, Medical Research Council Biostatistics Unit, Cambridge.
- Stenseth N, Mysterud A, Ottersen G, Hurrell JW, Chan K-S, Lima M (2002) Ecological effects of climate fluctuations. *Science* 297:1292–1296.
- Sturtz S, Ligges U, Gelman A (2005) R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software* 12:1–16.
- Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin J-M, Heogh-Guldberg O, Bairlein F (2002) Ecological responses to recent climate change. *Nature* 416:389–395.
- Walther G-R, Berger S, Sykes MT (2005) An ecological 'footprint' of climate change. *Proceedings of the Royal Society of London B* 272:1427–1432.
- White GC, Burnham KP (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study* 46:120–139.
- Wilson C, Adamec D (2002) A global view of bio-physical coupling from SeaWiFS and TOPEX satellite data, 1997–2001. *Geophysical Research Letters* 29:1–4.

redstart  
garden  
spring a  
could p  
change  
spring a  
graphic  
of the  
tion size  
depende  
region,  
winter. I  
can be c  
lated be  
ping nu  
analysis  
dynamic  
by distin  
difficult  
other me