



Combining a Bayesian nonparametric method with a hierarchical framework to estimate individual and temporal variation in growth

Douglas B. Sigourney^{a,c,*}, Stephan B. Munch^b, Benjamin H. Letcher^c

^a Department of Environmental Conservation, University of Massachusetts, Amherst, MA 01003, United States

^b Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Road, Santa Cruz, CA 95060, United States

^c S.O. Conte Anadromous Fish Research Center, Leetown Science Center, U.S. Geological Survey, P.O. Box 796, One Migratory Way, Turners Falls, MA 01376, United States

ARTICLE INFO

Article history:

Received 8 March 2012

Received in revised form 11 August 2012

Accepted 13 August 2012

Available online 20 September 2012

Keywords:

Atlantic salmon

Bayesian state space model

Growth model

Mark-recapture

Random effects

Gaussian process

ABSTRACT

Growth modeling has long played an important role in ecology, conservation and management of many species. However, adopting a statistical framework that includes both temporal and individual variability in the growth dynamics has proven challenging. In this paper, we use a Bayesian state space framework (BSSF) to estimate parameters of a discrete time model from a mark-recapture data set of age-1 juvenile Atlantic salmon. We use a Gaussian process (GP) based approach to model variation in seasonal growth potential. In addition, we use auxiliary information on the food environment as prior knowledge of seasonal fluctuations in growth. Parameters for the GP prior and measurement error variances were fixed to speed convergence. Posterior estimates of model parameters were relatively insensitive to these choices. Our model captures the seasonal growth dynamics of juvenile Atlantic salmon as evidenced by close agreement between observed and predicted lengths ($r^2 = 0.98$). In addition, the relatively narrow confidence intervals indicated significant learning in the parameters of interest. Finally, our model approach was able to accurately recover missing data points. Although this model was applied to a mark-recapture dataset of Atlantic salmon, the generality of the approach should make it applicable to a wide variety of size trajectory datasets, and thus, provides a useful tool to estimate individual and temporal variability in growth from datasets with repeated measurements.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Understanding growth rate variation is an important goal in ecological studies. Growth and consequently size variation can strongly influence population dynamics (Ebenman and Persson, 1988; DeAngelis et al., 1993; Peacor et al., 2007). As a cohort ages, variation in growth occurs at two important scales, individual and temporal. Individual variation in size can arise through a combination of intrinsic differences (Arendt, 1997; Fujiwara et al., 2004) and resource heterogeneity in the environment (Pfister and Peacor, 2003). Temporal variation occurs due to the tight relationship between size and metabolism (Peters, 1983; Schmidt-Nielsen, 1984) and through changes in the growth conditions, e.g. food availability and temperature, with time (Szalai et al., 2003). Because of its obvious importance, there have been many attempts to model

growth; however, incorporating these important sources of variation presents many challenges.

There are a number of approaches to modeling growth. Models based on simple mass-balance equations such as the von Bertalanffy growth function (VBGF) have been remarkably effective in capturing the basic ontogenetic growth trajectory of many species (von Bertalanffy, 1938; Quinn and Deriso, 1999; West et al., 2001). However, in their simplest form, these models assume no individual or temporal variation. Ignoring individual variation can lead to biased estimates of growth. For example, studies have found that ignoring individual variability when fitting the VBGF can lead to substantial bias in parameter estimates (Sainsbury, 1980; Eveson et al., 2007) leading researchers to allow for individual differences when fitting the VBGF to mark-recapture data (Francis, 1988; Laslett et al., 2002). A Bayesian hierarchical framework offers a tractable and informative approach to account for individual variability in growth in a population (Clark, 2003; Clark et al., 2005). For example, accounting for individual heterogeneity in capture probabilities and survival can be difficult in mark-recapture studies when maximizing the likelihood, but is more easily accomplished by using Monte Carlo integration to handle the random effects (Zheng et al., 2007). Hence, a Bayesian framework offers a viable

* Corresponding author. Present address: Department of Wildlife Ecology, Nutting Hall, University of Maine, Orono, ME 04469-5755, United States.
Tel.: +1 207 581 2870; fax: +1 207 581 2858.

E-mail address: douglas.sigourney@maine.edu (D.B. Sigourney).

alternative to maximum likelihood approaches when estimating growth parameters for the VBGF (He and Bence, 2007).

In addition to individual variability, environmental conditions change through time producing temporal variability in growth dynamics. For example, organisms living in temperate environments will often exhibit strong seasonality in their growth dynamics (Baba et al., 2004; Bacon et al., 2005; Strothotte et al., 2005). This seasonality can reflect a mixture of fluctuations in temperature and food availability as well as other factors that affect the growth process (Vollestad et al., 2002; Letcher and Gries, 2003). Although time-dependent solutions of the VBGF have appeared in the literature for decades (Levin and Goodyear, 1980), there have been comparatively few attempts to account for temporal variability when fitting VBGF models to field data (but see Szalai et al., 2003; Eveson et al., 2004). Previous approaches for incorporating temporal variability in growth typically used parametric forms to capture the seasonal rise and fall in growth rates (Jones et al., 2002; Eveson et al., 2004). For example, Eveson et al. (2004) embedded a sinusoidal function into the VBGF to describe seasonal growth in southern bluefin tuna. Although this approach could capture some of the temporal variability, they note that the sinusoidal function lacks flexibility. Similarly, Jones et al. (2002) developed a model based on dynamic energy budget theory that included a seasonality function to capture the rise and fall in growth. However, these approaches assume the chosen model is a good approximation to the unknown temporal driver. An alternative approach is to use a non-parametric function that does not assume a rigid relationship between time and growth. Non-parametric functions offer a promising method to infer ecological relationships when the underlying function is unknown (Cook, 1998; Munch et al., 2005; Dahlgren et al., 2011). When embedded in a Bayesian framework, auxiliary information such as seasonal changes in prey availability can be incorporated to inform this dynamical process.

A Bayesian state-space framework (BSSF) offers a comprehensive probabilistic approach to modeling that can deal with many sources of uncertainty (Clark, 2005). For example, the state-space approach deals explicitly with measurement and process errors (de Valpine and Hastings, 2002). In a growth modeling context, if measurement error is ignored it can lead to substantial bias in model predictions and parameter estimates (Clark et al., 2007). The Bayesian context also allows for the use of prior information to help inform the growth process (McCarthy and Masters, 2005). Finally, Markov chain-Monte Carlo (MCMC) techniques commonly used to estimate parameters in a BSSF can efficiently deal with hierarchical models. With the increase in computer power and availability of software packages to perform MCMC integration, the application of these techniques has become increasingly common in the ecological literature (Cressie et al., 2009; de Valpine, 2009). Because of these advantages, a number of recent studies have adopted a Bayesian approach to estimating parameters for the VBGF (Zhang et al., 2009; Bal et al., 2011).

In this paper we present a simple linear growth model that can be derived from the VBGF. We apply our model to mark-recapture data of age-1 juvenile Atlantic salmon (*Salmo salar*). To capture the various components of variation inherent in our dataset, we adopt a BSSF that allows us to simultaneously account for both measurement error and process error in the growth dynamics. Our modeling approach is similar to one proposed by Clark et al. (2007) to model incremental growth in individual trees; however, we include a flexible non-parametric growth function that allows for temporal variation in the growth dynamics. A number of studies have indicated the importance of food availability in determining growth rates of salmonids in the wild (Nislow et al., 1998; Bacon et al., 2005), and therefore, we include auxiliary information in the form of invertebrate drift samples to inform the temporal growth dynamics, highlighting the advantage of the BSSF.

2. Methods

2.1. Data collection

We applied our model to a mark-recapture dataset of juvenile Atlantic salmon from the Westbrook River located in Whately, MA, USA. Details of the study site and data collection can be found elsewhere in the literature (Letcher and Gries, 2003) so we will provide only relevant details here. To capture individuals, electrofishing was conducted in an upstream direction using a two pass removal approach. Sampling occurred on four occasions throughout the course of the year from March through December. Each sampling occasion took approximately ten days. Captured salmon were placed in buckets before tagging. Each fish that was large enough for tagging (>6.0 cm and >2.0 g) was placed in a solution of MS-222 that was buffered with sodium bicarbonate. Once a fish was sufficiently anaesthetized, it was scanned for a PIT tag, measured for total length and fork length and weighed. If a PIT tag was not present, the fish was tagged by making a small incision just below the pectoral fins and inserting a tag into the peritoneal cavity. In addition, all fish are lightly squeezed around the abdomen to check for milt, indicating a mature male. All fish were returned to the section of capture. To fit our model we used data from the 2000 cohort that included 137 individuals with a total of 365 observations of length.

In addition, in between each sampling occasion a drift sample of invertebrates was conducted during three different days at three different locations within the study site (see Grader and Letcher, 2006 for more details). Drift samples were conducted by placing a net in the brook for 20 min and collecting the drift in a cod end. Flow velocities were measured to estimate the volume of water passing through the net. For the purpose of our growth model we used total invertebrate biomass by volume as a measure of food availability during that specific growth interval. Values ranged from 0.78 mg/m³ in the spring to 0.06 mg/m³ in summer. Because three replicates were made during each sampling period we could calculate a mean and variance for each sample. We used this information to help inform the shape of the temporal growth function described below.

2.2. Model development

Similar to other mark-recapture studies of stream salmonids (Vollestad et al., 2002; Bacon et al., 2005), we observed considerable variation in both seasonal and individual growth within the study site (Fig. 1). Therefore, to estimate parameters and make an inference about seasonal growth potential we adopt a hierarchical state-space framework wherein we allow for individual variation in growth while explicitly modeling observation and process uncertainty (Fig. 2). The observation model accounts for the fact that the observed lengths may have associated measurement errors while the process model allows for the uncertainty associated with individual growth dynamics. In our specific data example, error in the measurements of length of recaptured Atlantic salmon is quite low and could probably be safely ignored. However, we retain the state-space framework to facilitate application of the model to a variety of other datasets where uncertainty in measurements could be considerably higher, e.g. back-calculated growth trajectories from otoliths or scales.

The basic structure of the state-space framework includes specifying a probability distribution for the observation model that connects the observed data to the model predictions and specifying a process model which describes growth dynamics (Clark, 2005). Finally, to complete the Bayesian model specification, prior distributions are specified for all parameters. See Clark (2005) for an excellent introduction to Bayesian state-space modeling. To model

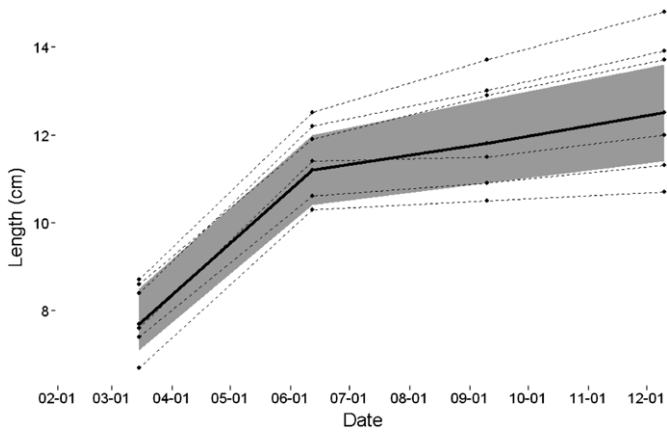


Fig. 1. Growth trajectories in length (cm) of age 1 juvenile Atlantic salmon over the course of a season (March to December). Mean size (—) ± 1 SD (grey region) and individual growth trajectories (···) of six individuals are shown. Black circles (●) represent observed lengths (cm).

temporal variability in growth, we add another layer to this structure, by specifying a prior on seasonal growth potential. Finally, to account for individual variation in growth, we add one more layer to the model describing individual variation in growth potential. An overview of the basic model structure is given by

Observation model : $P(y_{i,t}|L_{i,t}, \theta)$

Process model : $P(L_{i,t}|L_{i,t-1}, F(t), \alpha_i, \theta)$

Prior on seasonal growth potential : $P(F(t)|\theta)$

Prior on individual growth : $P(\alpha_i|\theta)$

Priors on parameters/hyperparameters : $P(\theta)$

where $y_{i,t}$ is the observed length of individual i at time t , $L_{i,t}$ is its actual length, $F(t)$ is the unknown function describing seasonal growth potential, α_i represents individual variation in growth potential, and θ is a vector representing all of the remaining model parameters. In the paragraphs that follow, we describe specific forms for each of the model components.

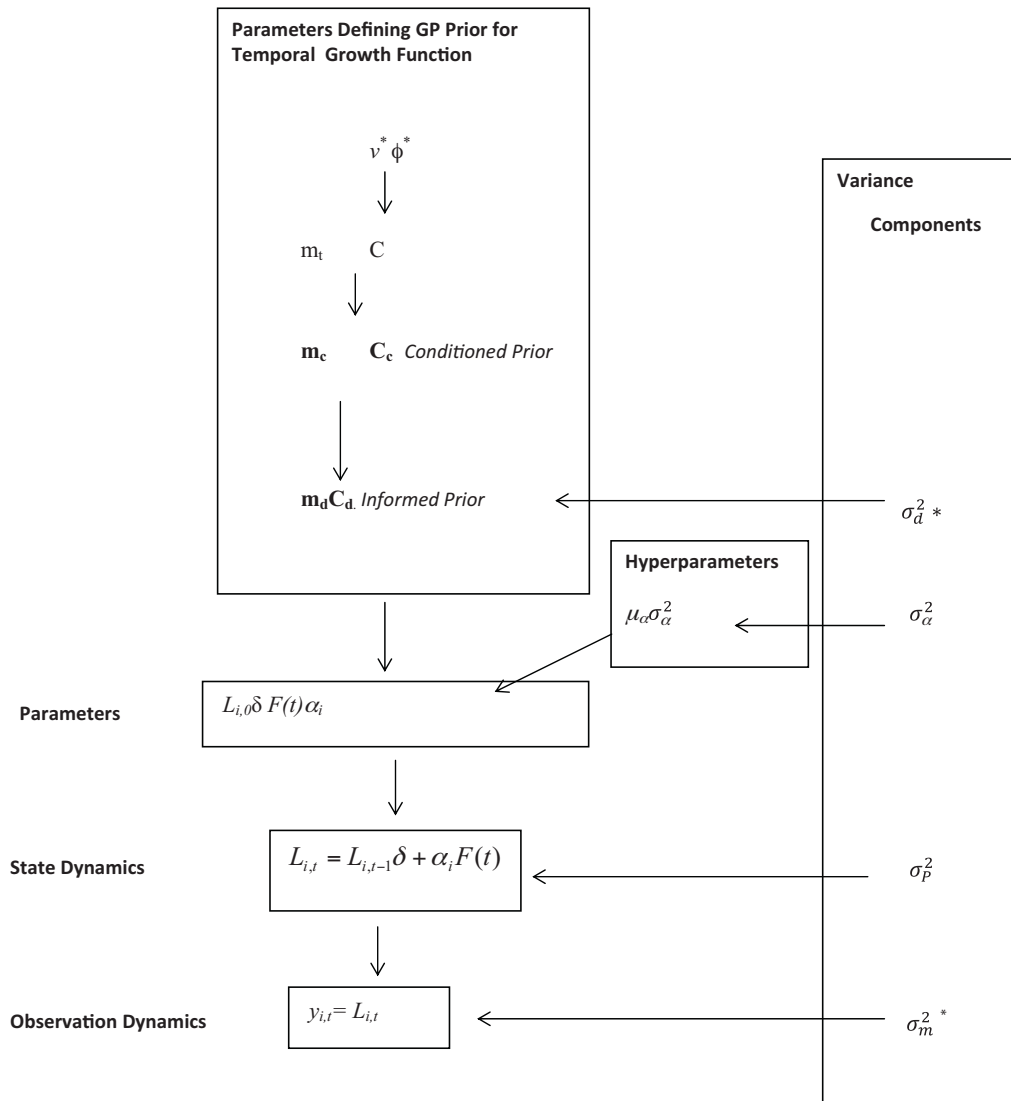


Fig. 2. Parameters defining GP prior for temporal growth function.

For the observation model, we assume normal, zero-mean measurement errors. Specifically, the observed length of the i th individual at time t , $y_{i,t}$, is normally distributed with mean $L_{i,t}$ and measurement variance σ_m^2

$$y_{i,t} \sim N(L_{i,t}, \sigma_m^2) \quad (1)$$

Development of the process model is somewhat more involved and began by specifying a continuous time model for individual growth including seasonal variation in growth potential. However, since our data were collected at nearly fixed time intervals (90 days \pm 5 days), we integrated the continuous time model to obtain a discrete time representation (see Appendix A for details). Because individual growth dynamics are likely to be driven by stochastic variation in feeding rates, activity levels, and local environmental conditions, the resulting linear model is only an approximation to an individual's growth history. We therefore allow for process uncertainty in the growth dynamics. Specifically, the process model for growth in length is

$$L_{i,t} = L_{i,t-1}\delta + \alpha_i F(t) + \varepsilon_{i,t} \quad (2)$$

where $\varepsilon_{i,t}$ represents process errors and are assumed to be normally distributed with variance σ_p^2 . Since the model specifies the dynamics of length (L) recursively, and the initial 'true' length of an individual ($L_{i,0}$) is unknown, we specify a uniform prior for initial length as $L_{i,0} \sim U(0,500)$.

One goal of our modeling approach was to model individual variation in growth using a hierarchical structure. We therefore treat the α_i 's as random effects arising from a normal distribution with mean μ_α and variance σ_α^2 . That is,

$$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2) \quad (3)$$

2.3. Modeling seasonal growth potential

There are a number of parametric functions that could be assumed for $F(t)$ to capture the seasonal growth observed in our data. However, we endeavored to develop a model with wide applicability that could capture any number of unknown functions while making limited assumptions about the shape of those functions. Therefore, to make an inference about temporal fluctuations in environmental growth potential, we specify a Gaussian process (GP) prior (Neal, 1999) for $F(t)$ and update it given the observed growth time series. A GP is a continuous stochastic process which generalizes the multivariate normal distribution and analogously is specified in terms of mean and covariance functions, m and C respectively. The mean function, $m(t)$ gives the average value of $F(t)$ evaluated at time t . The covariance between realizations of $F(t)$ at two points in time, say t and t' is given by $C(t,t') = \text{Cov}(F(t), F(t'))$. The covariance function controls the shape of $F(t)$ by determining both the variance in $F(t)$ at a single point in time and the wiggleness of $F(t)$ through time. Although the GP prior makes distributional assumptions, it can be considered a nonparametric approach in the sense that the parameters do not restrict the shape of the regression model (see Munch et al., 2005 for a detailed explanation).

The distribution of $F(t)$ evaluated at a discrete, finite set of times is multivariate normal. Specifically, on a discrete set of times, $\mathbf{t} = \{t_1, t_2, \dots, t_n\}$, the prior distribution for $F(t)$ at those times, $F(\mathbf{t}) = \{F(t_1), F(t_2), \dots, F(t_n)\}$ is given by the multivariate normal $\text{MVN}(\mathbf{m}, \mathbf{C})$ where the vector $\mathbf{m} = \{m(t_1), \dots, m(t_n)\}^T$ is the mean function evaluated at the times in \mathbf{t} and the covariance matrix \mathbf{C} is the covariance function C evaluated at all pairs of times, i.e. the j th and k th element of \mathbf{C} is given by $\mathbf{C}_{\{j,k\}} = C(t_j, t_k)$. We set the prior mean function constant at 1 for all t . That is $m(t) = 1$ and

consequently \mathbf{m} is a vector of 1's. We assume a squared-exponential covariance function $C(t,t')$, given by

$$C(t, t') = \nu \exp(-\phi(t - t')^2) \quad (4)$$

where ν is the prior variance in the growth potential and ϕ represents the rate at which the covariance decays with the time difference. This particular choice of covariance function asserts that, in the absence of other information, variability in growth potential is smooth with roughly constant curvature over the year. More in-depth description of covariance functions and their roles in GP-based inference can be found in Rasmussen and Williams (2006).

Importantly, the model described so far is unidentifiable in that the average level of $F(t)$ is confounded with the mean of the individual growth parameters, μ_α . To address this problem we can constrain either the distribution for the α_i 's or constrain $F(t)$ in some way. Using standard results for GPs (Rasmussen and Williams, 2006), we modified the prior such that realizations of $F(t)$ are constrained to integrate to 1 over the course of a year. We denote the mean and covariance functions of this constrained prior distribution by $m_c(t)$ and $C_c(t,t')$. The corresponding constrained mean vector and covariance matrix are denoted \mathbf{m}_c and \mathbf{C}_c .

Note that this prior specification for $F(t)$ is still relatively uninformative and asserts that we have no information on seasonal fluctuations in growth potential. However, we certainly know that growth potential varies through time because of changes in food availability and temperature (Nislow et al., 1998). We therefore used our invertebrate drift samples to update the prior for $F(t)$, inducing temporal variation in our prior on growth potential. To do so, we standardized the drift density such that the mean is 1. We then assume that, given $F(t_j)$, the observed, normalized drift density (d_j) at time t_j follows a normal distribution centered on $F(t_j)$ with observation variance σ_d^2 . That is, $d_j \sim N(F(t_j), \sigma_d^2)$. Gathering the data for each sampling time, $\mathbf{t} = \{t_1, t_2, \dots, t_n\}$ into a vector of observed drift densities $\mathbf{d} = \{d_1, d_2, \dots, d_n\}^T$, we then use Bayes' rule to update the mean and covariance for $F(t)$ given the information in \mathbf{d} , and denote the mean and covariance conditioned on the drift densities by \mathbf{C}_d and \mathbf{m}_d , i.e.

$$F(\mathbf{t}) | \mathbf{d}, \sigma_d^2 \sim \text{MVN}(\mathbf{m}_d, \mathbf{C}_d) \quad (5)$$

where

$$\mathbf{C}_d = \left(\mathbf{C}_c^{-1} + \frac{1}{\sigma_d^2} \mathbf{I}_n \right)^{-1} \quad (6)$$

and

$$\mathbf{m}_d = \mathbf{C}_d \left(\mathbf{C}_c^{-1} \mathbf{m}_c + \frac{1}{\sigma_d^2} \mathbf{d} \right) \quad (7)$$

Here, \mathbf{I}_n denotes the $n \times n$ identity matrix. This inclusion of auxiliary data on prey availability reflects anticipated variation in $F(t)$. Critically, this approach to incorporating temporal variation still retains significant flexibility; the seasonal fluctuation in $F(t)$ that we infer from the growth data is not restricted to following this prior and can depart significantly from it when there is clear information in the growth data.

Eqs. (1)–(7) specify the observation model, a hierarchical model for individual growth potential, the process model for individual growth dynamics, and a GP prior for seasonal variation in growth potential. To complete the model, we specify priors for the parameters and hyperparameters. Prior distributions for δ , the hyperparameters μ_α , σ_α^2 , the process error variance, σ_p^2 , and the initial, latent states, $L_{i,0}$, are listed in Table 1. For the remaining, unknown parameters (ϕ , ν , σ_d^2), we fixed their values in the prior. These parameters can be estimated (see Munch et al., 2005 for an example), but doing so greatly extends the time to convergence.

Table 1
Definitions of parameters and corresponding prior distribution.

Parameter	Definition	Prior
<i>Growth parameters</i>		
δ	Derived parameter that is proportional to net energy loss	$N(0.95,1000)$
μ_α	Hyperparameter for the mean of the random effects	$U(0,1000)$
$F(t)$		See below
<i>Parameters for the prior of $F(t)$</i>		
ν	Magnitude of the prior variance in $F(t)$	Fixed to 10
ϕ	Rate of decay in the covariance of $F(t)$	Fixed to 0.5
$m(t)$	Mean growth potential at time t	
\mathbf{m}_c	Conditioned prior mean	
\mathbf{C}_c	Conditioned covariance	
\mathbf{m}_d	Informed prior mean	
\mathbf{C}_d	Informed prior covariance	
<i>Variance components</i>		
σ_p^2	Process error variance	$IG(2,0.01)$
$\sigma_i(\uparrow 2)$	Hyperparameter determining the variance in the random effects	$IG(2,0.1)$
σ_m^2	Measurement error variance	Fixed to 0.1
σ_d^2	Variance in observed food data	Fixed to 0.3
<i>State variables</i>		
$y_{i,t}$	Observed length of fish i at time t	
$L_{i,0}$	Latent state representing the true, initial length of fish i	$U(0,500)$

Moreover, posterior distributions for $F(t)$ are relatively insensitive to choice of ϕ , and ν provided that ν is reasonably large. Therefore, to speed convergence, we fixed the values of ϕ and ν to 0.5, and 10 respectively in the prior (see Table 1). In keeping with other applications of state-space methods (Clark et al., 2007), the measurement variances for length and drift density were estimated directly from repeated measurements. Specifically, we used a value of 0.1 cm² for σ_m^2 based on 10 repeated measurements of 10 individuals and a value of 0.3 for σ_d^2 which was the highest variance observed over all food samples. For each fixed parameter, we conducted a sensitivity analysis by increasing the value of the parameter by 100%, re-running the model, and summarizing the percent change in the posterior distributions of the estimated parameters.

We used a Gibbs sampler (Gelman et al., 2004) to draw samples from the posterior and obtain parameters estimates. All calculations were carried out in software written by DS in Matlab (The Mathworks, Natick, MA). We ran the sampler for 1 million iterations keeping one in every 100 samples to reduce autocorrelation. To assess convergence, we ran three separate chains. Visual examination of trace plots indicated adequate mixing. In addition, convergence of marginal, posterior distributions was further confirmed by Gelman–Rubin statistics (Gelman et al., 2004) that were less than 1.02 for all parameters except for σ_p^2 which had a value of 1.14.

3. Results

3.1. Parameter estimates

Posterior medians and Bayesian credible intervals for the growth parameters and variances are provided in Table 2. Narrow credible intervals indicate substantial learning in the parameters

Table 2
Posterior median values and 95% Bayesian credible intervals for parameters of the growth model.

Parameter	Posterior median	Credible interval	
		2.5%	97.5%
δ	0.99	0.92	1.04
μ_α	4.54	3.29	6.80
$\sigma_i(\uparrow 2)$	0.12	0.04	0.33
σ_p^2	0.007	0.002	0.020

of interest (Table 2). The inclusion of individual random effects (α_i) in the model seemed to capture an important component of variation as evidenced both by histograms of posterior median estimates (Fig. 3) and posterior estimates of the variance component, σ_α^2 , with Bayesian credible intervals that did not cover zero (Table 2). The process error variance, σ_p^2 , was substantially lower than the random effects error variance, σ_α^2 , and the assumed measurement error variance, σ_m^2 . The small variance in process error is not surprising as the other sources of variation allow substantial flexibility in the model.

Sensitivity analysis verified that marginal posterior distributions of growth parameters were relatively insensitive to choices made in the prior (Fig. 4). Graphical comparisons of posterior distributions of fixed values (Figs. 4a and b, respectively). Of all the fixed parameters, posterior distributions of δ and μ_α and the median estimates of latent states, $L_{i,t}$, were most sensitive to assumptions of measurement error. However, we found that the latent states varied by less than 4%, and the majority of estimates differed by less than 0.5%, indicating that overall estimates of growth were robust to assumptions on the magnitude of σ_m^2 .

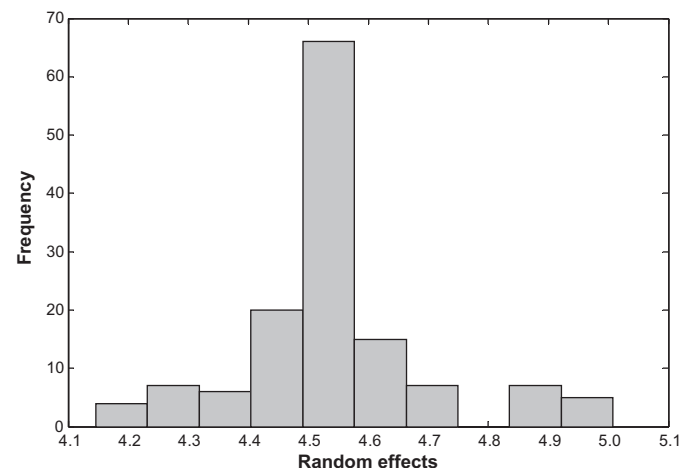


Fig. 3. Posterior median estimates of the random effects (α_i).

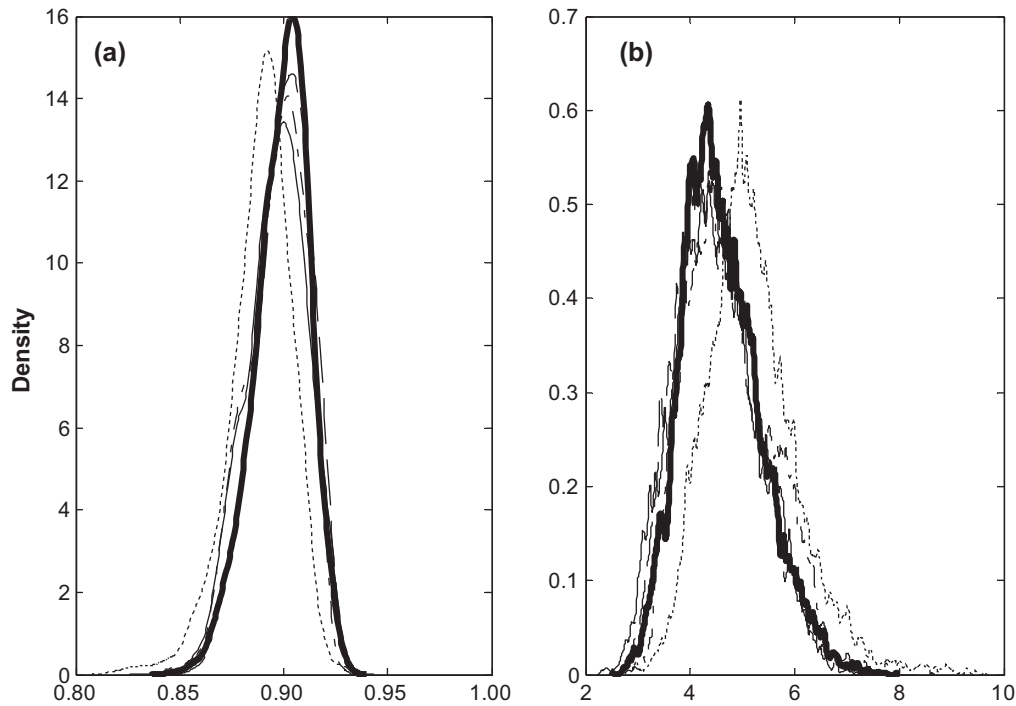


Fig. 4. Results of a sensitivity analysis on the posterior distributions of (a) δ and (b) μ_α after perturbing parameters that had values fixed in the prior (ϕ , v , σ_d^2 , and σ_m^2). The original posterior distributions using the values listed in Table 1 are shown by the solid line (—), along with resulting posterior distributions obtained after increasing σ_m^2 (---), σ_d^2 (---), v (---) and ϕ (---) by 100%.

3.2. Seasonal growth potential

A comparison of prior distributions to the posterior distributions of $F(t)$ suggest considerable learning regardless of whether or not food data were available (Fig. 5). The uninformed prior (Fig. 5a) and the constrained prior (Fig. 5b) were both relatively flat with high prior uncertainty. The observed invertebrate drift data did inform the shape of the prior though there was still considerable uncertainty in $F(t)$ (Fig. 5c). The estimated posterior distribution generally

followed the shape of the informed prior but with considerably greater precision (Fig. 5d). To assess the importance of incorporating food availability into the prior we re-computed the posterior without using the invertebrate data, i.e. using only the constrained prior shown in Fig. 5b. The posterior distribution of $F(t)$ using the uninformed prior (Fig. 5e) was nearly identical to posterior distribution using the constrained prior (Fig. 5d). Thus, information on the food environment was not critical for estimating $F(t)$ suggesting that the observed growth trajectories contained considerable

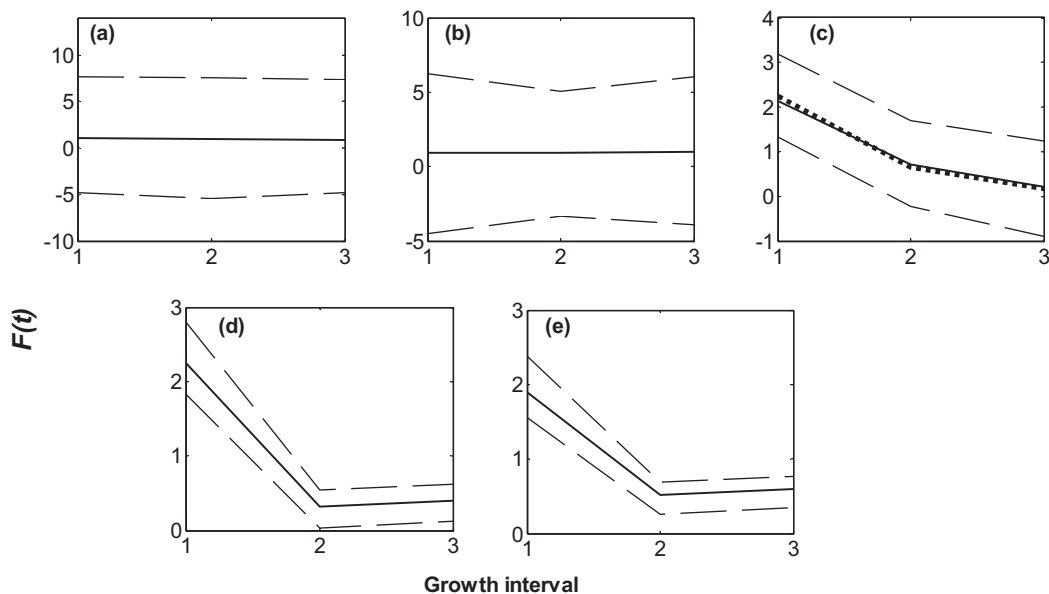


Fig. 5. Median estimates and 95% credible intervals of 10,000 draws from different distributions of the seasonal growth potential, $F(t)$, including the (a) the unconstrained prior distribution (b) the constrained prior distribution (c) the informed prior distribution updated with information on invertebrate drift collected in the study site (d) the posterior distribution estimated using the informed prior and (e) the posterior distribution estimated using the uninformed, constrained prior. Posterior medians (—) and 95% Bayesian credible intervals (---) are shown. Standardized values of the observed invertebrate drift (...) are plotted in panel c.

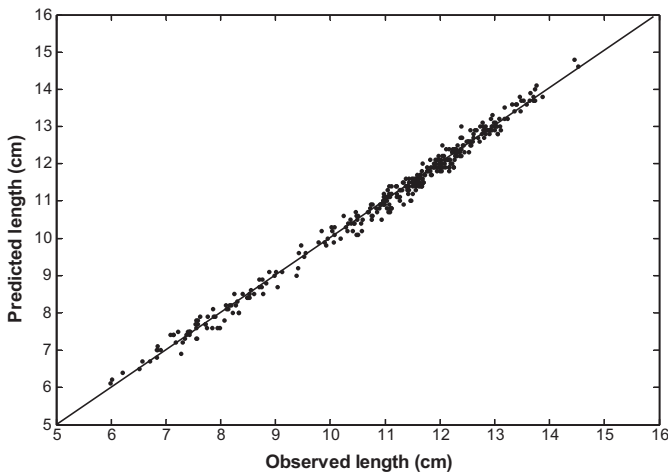


Fig. 6. Observed length ($y_{i,t}$) vs median posterior estimates of “true” length ($L_{i,t}$) from the growth model ($r^2 = 0.98$).

information on seasonal changes in growth potential. Alternately, the similarity between the informed prior and the posterior indicates strong correspondence between food availability and growth potential. This type of prior information could therefore be of use in inferring seasonal growth potential when growth data are less informative.

3.3. Growth trajectories

We found close agreement between observed lengths and model estimates of length (Fig. 6). The close agreement between observed lengths, $y_{i,t}$, and the estimated “true” lengths, $L_{i,t}$, is not all that surprising as the individual variation and seasonal parameters add a high degree of flexibility. Posterior estimates of individual length trajectories suggested the model was able to capture the seasonal shape of individual growth trajectories as well as

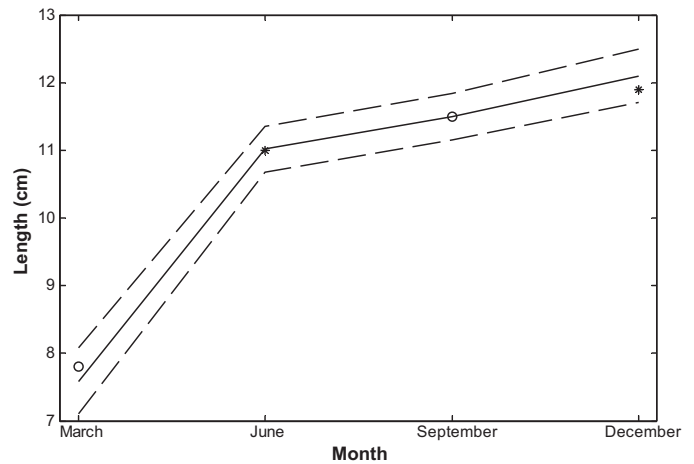


Fig. 8. Posterior median estimates (—) and 95% Bayesian credible intervals (--) of the true lengths ($L_{i,t}$) of an individual salmon where two observations were deleted. Open circles (○) represent observed lengths and asterisks (*) represent deleted observations.

the variability among individuals (Fig. 7a–d). The model estimated the growth trajectory quite well for individuals with no missing observations as evidenced by close agreement between estimates and observations as well as narrow credible intervals around estimates of length (Fig. 7a). For individuals with missing observations, uncertainty in unobserved length tended to increase if there were missing observations at the beginning of the growth trajectory (Fig. 7b). However, information about an individual’s length at the end and beginning of the growth seemed to help pin down uncertainty where observations were missing (Figs. 7c and d).

The model’s ability to predict lengths for individuals with missing observations was well demonstrated by its ability to predict observations for an individual whose observations had been removed from the dataset (Fig. 8).

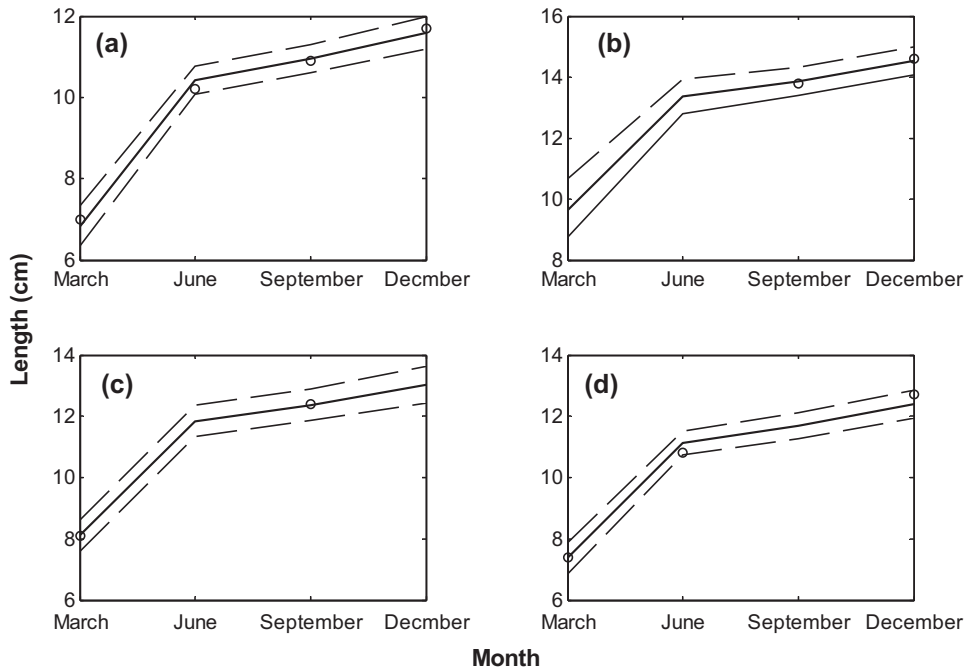


Fig. 7. Posterior median estimates of length (cm) (solid lines) and 95% Bayesian credible intervals (dashed lines) for individual salmon with different numbers of repeated observations. Observations of length are indicated with open circles (○).

4. Discussion

Using a comprehensive growth modeling framework, we were able to account for multiple levels of variation inherent in a mark-recapture data set of juvenile Atlantic salmon. The BSSF used here allowed us to take a relatively simple model and accurately predict the complex seasonal growth trajectories in a cohort of fish with separate life histories. Without this framework it would be difficult to accomplish this goal. Our model assumes that fish grow according to the VBGF, but also accounts for process error that may arise from model misspecification. In addition, we simultaneously address both temporal and individual variation in the growth dynamics, both of which are intrinsic characteristics of Atlantic salmon growth.

Atlantic salmon are known to exhibit individual variation in growth due to their territorial behavior and differences in metabolism (Metcalf, 1998). Consequently, a model of Atlantic salmon growth needs to account for this variation. From our data, it was clear that there was significant individual variation. Mixed models that account for individual variation offer many advantages over fixed effects models (Weisberg et al., 2010). Our modeling approach is not the first to include random effects in a growth modeling context, however, most attempts have used likelihood approaches which involve complex integrals that can be computationally expensive (Sainsbury, 1980; Wang et al., 1995; Eveson et al., 2007). For this reason a number of researchers are opting to use Bayesian approaches as an alternative (Zhang et al., 2009). In a Bayesian context, the random effects can be integrated over using MCMC techniques. The linear dynamics of this model allow for efficient Gibbs algorithms to be derived where we can sample both the fixed maintenance parameter and the individual parameters from a conditional distribution. Thus, our modeling approach can efficiently deal with the issue of individual variation in a comprehensive probabilistic framework.

The temporal dynamics of temperature, flow and food availability result in highly seasonal growth dynamics of stream salmonids in northern streams (Letcher and Gries, 2003; Vollestad et al., 2002; Bacon et al., 2005). There have been a number of attempts to include temporal variability in growth models by using environmental variables (Millar and Myers, 1990; Millar et al., 1999; Jones et al., 2002). However, as Szalai et al. (2003) point out, these approaches either use a linear statistical model to relate environmental variation to growth or assume some functional relationship between the growth and the environmental variable(s) of interest. Complex biological interactions such as cannibalism, pulses in prey abundance or variation in the timing of reproduction can lead to variable seasonal growth trajectories that are not easily summarized by one functional form (Robins et al., 2006; Armstrong et al., 2010). Using a GP prior, we included a temporal function that is not limited by an assumed functional form, and therefore, allow the data more freedom to inform the shape of the function. For example, the GP prior does not assume growth increases or decreases in any particular way over the course of a year, and therefore, the relationship between season and growth can take on a number of functional forms. In this sense, our model is not restricted to animals that only exhibit fast spring growth and can be applied to fit any number of complicated growth trajectories. The advantage is no increase in parameters is needed to model more complex growth trajectories with more temporal variability. Though rare in ecological modeling, flexible Bayesian nonparametric modeling approaches using GP priors may offer a viable alternative to parametric approaches in modeling complicated temporal dynamics (Munch et al., 2005; Vanhatalo et al., 2012).

One advantage of our approach to modeling temporal variation in growth is it allows for the use of prior information. In particular, we asserted a priori that growth potential should track seasonal

fluctuations in drift density. This correlation has been observed in the study site (see Grader and Letcher, 2006). A popular alternative approach would have been to include drift density as a covariate in the model (Bacon et al., 2005). Doing so invariably requires some assumptions about how drift density interacts with other environmental variables in driving growth, even when flexible functional forms are used. For instance, including an additive drift density term to the discrete time growth model (Eq. (2)) directly would produce substantially different results from including an additive drift density term to the continuous time growth model (Eq. (A1)) and then integrating, which generates an exponentially weighted moving average of the basis functions used. Moreover, either addition would assume that there was no interaction between food and temperature. In our approach, drift densities are included only in the prior for $F(t)$. This method does not constrain the posterior for $F(t)$ to follow the observed drift density exactly while simultaneously smoothing out variation in drift density via (Eq. (7)). The value in doing so is that the inferred $F(t)$ reflects realized seasonal variation in growth potential rather than an assumed functional form and implicitly integrates all of the environmental variables that drive growth through time.

Although measurement error was small and could possibly be ignored without biasing the growth dynamics, we found it necessary to provide a BSSF that simultaneously addresses both the error in the measurement of lengths and process uncertainty in the assumed growth dynamics. First of all, we aimed to present a general framework that could be applied to the growing number of longitudinal studies where measurement error may be more substantial. For example, many studies use otoliths or scales to reconstruct the growth history of individuals (Li et al., 2008; McCarthy et al., 2008). Another promising approach to measuring individual size is through the use of underwater video (Harvey et al., 2003). These techniques, however, are all subject to non-negligible measurement error. Secondly, it is often easy to gather prior information on the magnitude and variability in measurement error through repeated measurements. In the BSSF, this information can be used to develop informative priors that help pin down the measurement uncertainty and inform how much of the remaining uncertainty is due to error in the growth model assumptions. Application of the general state-space method has proven successful in overcoming the problem of imperfect detection in a variety of situations including analyzing population dynamics (Meyer and Millar, 1999; de Valpine, 2003; Lindley, 2003), survival estimation (Gimenez et al., 2007), animal movements (Barraquand and Benhamou, 2008; Jonsen et al., 2005), and growth (Fujiwara et al., 2005). By allowing the use of prior information, the BSSF offers a comprehensive statistical method to address the vexing problem of observation uncertainty (Clark et al., 2007).

There are a number of potential uses but also limitations of this modeling approach. Its ability to estimate size for missing observations and uncertainty around those observations could be of great utility in studies that attempt to estimate size-dependent processes from mark-recapture data. For example, maturity, emigration and survival are all important demographic processes that are often dependent on size. These processes can be efficiently modeled with mark-recapture data; however, time varying covariates such as size can present challenges (Gimenez et al., 2008). Our growth modeling approach could be embedded in a detailed survival analysis that tries to relate size to survival, maturation and movement. On the other hand, we need to be cautious in using this approach to make inferences about fish bioenergetics. For example, while it seems that this approach would be useful to infer temporal and individual variation in consumption, the mass-scaling assumptions of the VBGF can result in biased estimates of consumption (Essington et al., 2001).

In summary we have presented a simple model for growth in length that follows VBGF dynamics and simultaneously accounts for temporal and individual variation in growth. There have been a number of sophisticated modeling approaches applied to the VBGF that have dealt with individual variation (Sainsbury, 1980; Wang et al., 1995) or temporal variation (Szalai et al., 2003), but we are not aware of any attempts to simultaneously include both in a BSSF. Using the statistical approach described herein, we developed a growth modeling framework that could (i) efficiently deal with inherent individual variation, (ii) include temporal variation in growth without assuming a parametric function, (iii) explicitly separate measurement error from process error and (iv) allow for auxiliary information to be used in a probabilistic framework to inform the growth dynamics. By explicitly addressing multiple sources of variation inherent in longitudinal studies, this approach should be readily applicable to a great number of datasets.

Acknowledgments

The authors would like to thank the many people who helped sample the study site, in particular Todd Dubreuil and Gabe Gries. Ole Shelton provided useful comments on earlier versions of the paper.

Appendix A.

To derive a discrete time model for growth, we assume that growth of individuals in length, L , is well approximated by a time-varying VBGF model

$$\frac{dL_{i,t}}{dt} = G_i(t) - \lambda L_{i,t} \quad (\text{A1})$$

where $G_i(t)$ represents the i th individual's energy gain rate, or growth potential, and $\lambda L_{i,t}$ is the rate of energy loss. In keeping with much physiological data (Kerr, 1971; Kitchell et al., 1977), this form asserts that most of the individual level variability in growth results from variation in rates of energy gain rather than size-specific energy losses.

To allow for persistent differences in individual growth, we decompose the net rate of energy gain into individual and seasonal components as

$$G_i(t) = \alpha_i f(t) \quad (\text{A2})$$

where $f(t)$ represents seasonal variation in environmental growth potential common to all individuals and α_i represents the multiple of this growth potential realized by individual i . Our approach to modeling $f(t)$ is described further below.

A more flexible alternative, which we defer to future analyses, would be to use a hierarchical model for $G_i(t)$ directly.

Now that we have specified a continuous time model, we can integrate Eq. (A1) to derive the following discrete time model

$$L_{i,t} = L_{i,t-1} \delta + \alpha_i F(t) \quad (\text{A3})$$

where

$$\delta = e^{-\lambda}$$

and

$$F(t) = e^{-\lambda t} \int_{t-1}^t e^{\lambda s} f(s) ds$$

References

Arendt, J.D., 1997. Adaptive intrinsic growth rates: an integration across taxa. *Quarterly Review of Biology* 72, 149–177.

- Armstrong, J.B., Schindler, D.E., Omori, K.L., Ruff, C.P., Quinn, T.P., 2010. Thermal heterogeneity mediates the effects of pulsed subsidies across a landscape. *Ecology* 91, 1445–1454.
- Baba, K., Kawajiri, T., Kuwahara, Y., Nakao, S., 2004. An environmentally based growth model that uses finite difference calculus with maximum likelihood method: its application to the brackish water bivalve *Corbicula japonica* in Lake Abashiri, Japan. *Fisheries Bulletin* 102, 14–24.
- Bacon, P.J., Gurney, W.S.C., Jones, W., McLaren, I.S., Youngson, A.F., 2005. Seasonal growth patterns of wild juvenile fish: partitioning variation among explanatory variables based on individual growth trajectories of Atlantic salmon (*Salmo salar*) parr. *Journal of Animal Ecology* 74, 1–11.
- Bal, G., Rivot, E., Prévost, E., Pious, C., Baglinière, J.L., 2011. Effect of water temperature and density of juvenile salmonids on growth of young-of-the-year Atlantic salmon *Salmo salar*. *Journal of Fish Biology* 78, 1002–1022.
- Barraquand, F., Benhamou, S., 2008. Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts. *Ecology* 89, 3336–3348.
- Clark, J.S., 2003. Uncertainty and variability in demography and population growth: a hierarchical approach. *Ecology* 84, 1370–1381.
- Clark, J.S., 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8, 2–14.
- Clark, J.S., Ferraz, G.A., Ogue, N., Hays, H., DiCostanzo, J., 2005. Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. *Ecology* 86, 2232–2244.
- Clark, J.S., Wolosin, M., Dietze, M., Ibanez, I., LaDeau, S., Welsh, M., Kloeppel, B., 2007. Tree growth inference and prediction from diameter censuses and ring widths. *Ecological Applications* 17, 1942–1953.
- Cook, R.M., 1998. A sustainability criterion for the exploitation of North Sea cod. *ICES Journal of Marine Science* 55, 1061–1070.
- Cressie, N.C., Calder, A., Clark, J.S., Ver Hoef, J.M., Wilke, C.K., 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications* 19, 553–570.
- Dahlgren, J.P., Garcia, M.B., Ehlren, J., 2011. Nonlinear relationships between vital rates and state variables in demographic models. *Ecology* 92, 1181–1187.
- DeAngelis, D.I., Rose, K.A., Crowder, L.B., Marshall, E.A., Lika, D., 1993. Fish cohort dynamics: application of complementary modeling approaches. *American Naturalist* 142, 604–622.
- de Valpine, P., Hastings, A., 2002. Fitting population models incorporating process noise and observation error. *Ecological Monograph* 72, 57–76.
- de Valpine, P., 2003. Better inferences from population-dynamics experiments using Monte Carlo state-space likelihood methods. *Ecology* 84, 3064–3077.
- de Valpine, P., 2009. Shared challenges and common ground for Bayesian and classical analysis of hierarchical statistical models. *Ecological Applications* 19, 584–588.
- Ebenman, B., Persson, L., 1988. *Size-structured Populations: Ecology and Evolution*. Springer, Berlin.
- Essington, T.E., Kitchell, J.F., Walters, C.J., 2001. The von Bertalanffy growth function, bioenergetics, and the consumption rate of fishes. *Canadian Journal of Fisheries and Aquatic Sciences* 58, 2129–2138.
- Eveson, J.P., Laslett, G.M., Polacheck, T., 2004. An integrated model for growth incorporating tag-recapture, length-frequency, and direct aging data. *Canadian Journal of Fisheries and Aquatic Sciences* 61, 292–306.
- Eveson, J.P., Polacheck, T., Laslett, G.M., 2007. Consequences of assuming an incorrect error structure in von Bertalanffy growth models: a simulation study. *Canadian Journal of Fisheries and Aquatic Sciences* 61, 292–306.
- Francis, R., 1988. Maximum likelihood estimation of growth and growth variability from tagging data. *New Zealand Journal of Marine and Freshwater Research* 22, 43–51.
- Fujiwara, M., Kendall, B.E., Nisbet, R.M., 2004. Growth autocorrelation and animal size variation. *Ecology Letters* 7, 106–113.
- Fujiwara, M., Kendall, B.E., Nisbet, R.M., Bennett, W.A., 2005. Analysis of size trajectory data using an energetic-based growth model. *Ecology* 86, 1441–1451.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC.
- Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J.P., Pradel, R., 2007. State-space modelling of data on marked individuals. *Ecological Modelling* 206, 431–438.
- Gimenez, O., Viallefont, A., Charmantier, A., Pradel, R., Cam, E., Brown, C.R., Anderson, M.D., Brown, M.B., Covas, R., Gaillard, J.M., 2008. The risk of flawed inference in evolutionary studies when detectability is less than one. *American Naturalist* 172, 441–448.
- Grader, M., Letcher, B.H., 2006. Diel and seasonal variation in food habits of Atlantic salmon parr in a small stream. *Journal of Freshwater Ecology* 21, 503–517.
- Harvey, E., Cappel, M., Shortis, M., Robson, S., Buchanan, J., Speare, P., 2003. The accuracy and precision of underwater measurements of length and maximum body depth of southern bluefin tuna (*Thunnus maccoyii*) with a stereo-video camera system. *Fisheries Research* 63, 315–326.
- He, J.X., Bence, J.R., 2007. Modeling annual growth variation using a hierarchical Bayesian approach and the von Bertalanffy growth function with application to lake trout in southern Lake Huron. *Transactions of American Fisheries Society* 136, 318–330.
- Jones, W., Gurney, W.S.C., Speirs, D.C., Bacon, P.J., Youngson, A.F., 2002. Seasonal patterns of growth, expenditure and assimilation in juvenile Atlantic salmon. *Journal of Animal Ecology* 71, 916–924.
- Jonsen, I.D., Flenning, J.M., Myers, R.A., 2005. Robust state-space modeling of animal movement data. *Ecology* 86, 2874–2880.

- Kerr, S.R., 1971. Analysis of laboratory experiments on growth efficiency of fishes. *Journal of Fisheries Research Board of Canada* 28, 801–808.
- Kitchell, J.F., Stewart, D.J., Weininger, D., 1977. Applications of a bioenergetics model to yellow perch (*Perca flavescens*) and Walleye (*Stizostedion vitreum vitreum*). *Journal of Fisheries Research Board of Canada* 34, 1922–1935.
- Laslett, G.M., Everson, J.P., Polacheck, T., 2002. A flexible maximum likelihood approach to fitting growth curves to tag-recapture data. *Canadian Journal of Fisheries and Aquatic Sciences* 59, 976–9867.
- Letcher, B.H., Gries, G., 2003. Effects of life history variation on size and growth in stream-dwelling Atlantic salmon. *Journal of Fish Biology* 62, 97–114.
- Levin, S.A., Goodyear, C.P., 1980. Analysis of an age-structured fishery model. *Journal of Mathematical Biology* 9, 245–274.
- Li, L., Hoie, H., Geffen, A.J., Heegaard, E., Skadal, J., Folkvord, A., 2008. Back-calculation of previous fish size using individually tagged and marked Atlantic cod (*Gadus morhua*). *Canadian Journal of Fisheries and Aquatic Sciences* 65, 2496–2508.
- Lindley, S.T., 2003. Estimation of population growth and extinction parameters from noisy data. *Ecological Applications* 13, 806–813.
- McCarthy, J.L., Friedland, K.D., Hansen, L.P., 2008. Monthly indices of the post-smolt growth of Atlantic salmon from the Drammen River, Norway. *Journal of Fish Biology* 72, 1572–1588.
- McCarthy, M.A., Masters, P., 2005. Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology* 42, 1012–1019.
- Metcalfe, N.B., 1998. The interaction between behavior and physiology in determining life history patterns in Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* 55, 93–103.
- Meyer, R., Millar, R.B., 1999. Bayesian stock assessment using a state-space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences* 56, 37–52.
- Millar, R.B. and Myers, R.A., 1990. Modeling environmentally induced change in growth for Atlantic Canada cod stocks. ICES CM 1990/G:24.
- Millar, R.B., McArdle, B.H., Harley, S.J., 1999. Modeling the size of snapper (*Pagrus auratus*) using temperature modified growth curves. *Canadian Journal of Fisheries and Aquatic Sciences* 56, 1278–1284.
- Munch, S.B., Kottas, A., Mangel, M., 2005. Bayesian nonparametric analysis of stock–recruitment relationships. *Canadian Journal of Fisheries and Aquatic Sciences* 62, 1808–1821.
- Neal, R.M., 1999. Regression and classification using Gaussian process priors. In: Bernardo, J.M., Berger, J.O., David, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics 6*. Clarendon Press, Oxford, UK.
- Nislow, K.H., Folt, C., Seandel, M., 1998. Food and foraging behavior in relation to microhabitat use and survival of age-0 Atlantic salmon. *Canadian Journal of Fisheries and Aquatic Sciences* 55, 116–127.
- Peacor, S.D., Schiesari, L., Werner, E.E., 2007. Mechanisms of nonlethal predator effects on cohort size variation: ecological and evolutionary implications. *Ecology* 88, 1536–1547.
- Peters, R.H., 1983. *The Ecological Implications of Body Size*. Cambridge University Press, Cambridge, UK.
- Pfister, C.A., Peacor, S.D., 2003. Variable performance of individuals: the role of population density and endogenously formed landscape heterogeneity. *Journal of Animal Ecology* 72, 725–735.
- Quinn, T.J., Deriso, R.B., 1999. *Quantitative Fish Dynamics*. Oxford University Press, New York.
- Rasmussen, C.E., Williams, C., 2006. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, USA.
- Robins, J., Mayer, D., Staunton-Smith, J., Halliday, I., Sawynok, B., Sellin, M., 2006. Variable growth rates of the tropical estuarine fish barramundi *Lates calcarifer* (Bloch) under different freshwater flow conditions. *Journal of Fish Biology* 69, 379–391.
- Sainsbury, K.J., 1980. Effect of individual variability on the von Bertalanffy growth equation. *Canadian Journal of Fisheries and Aquatic Sciences* 37, 241–247.
- Schmidt-Nielsen, K., 1984. *Scaling: Why is Animal Size so Important?* Cambridge University Press, Cambridge, UK.
- Strothotte, E., Chaput, G.J., Rosenthal, H., 2005. Seasonal growth of wild Atlantic salmon juveniles and implications on age at smoltification. *Journal of Fish Biology* 67, 1585–1602.
- Szalai, E.B., Fleischer, G.W., Bence, J.R., 2003. Modeling time-varying growth using a generalized von Bertalanffy model with application to bloater (*Coregonus hoyi*) growth dynamics in Lake Michigan. *Canadian Journal of Fisheries and Aquatic Sciences* 60, 55–66.
- Vanhatalo, J., Veneranta, L., Hudd, R., 2012. Species distribution with Gaussian processes: a case study with the youngest stages of sea spawning whitefish (*Coregonus lavaretus* L. s.l.) larvae. *Ecological Modelling* 228, 49–58.
- Vollestad, L.A., Olsen, E.M., Forseth, T., 2002. Growth-rate variation in brown trout in small neighbouring streams: evidence for density-dependence? *Journal of Fish Biology* 61, 1513–1527.
- von Bertalanffy, L., 1938. A quantitative theory of organic growth. *Human Biology* 10, 181–213.
- Wang, Y.G., Thomas, M.R., Somers, I.R., 1995. A maximum likelihood approach for estimating growth from tag-recapture data. *Canadian Journal of Fisheries and Aquatic Sciences* 52, 252–259.
- Weisberg, S., Spangler, G., Richmond, L.S., 2010. Mixed effects models for fish growth. *Canadian Journal of Fisheries and Aquatic Sciences* 67, 269–277.
- West, G.B., Brown, J.H., Enquist, B.J., 2001. A general model for ontogenetic growth. *Nature* 413, 628–631.
- Zhang, Z., Lessard, J., Campbell, A., 2009. Use of Bayesian hierarchical models to estimate northern abalone, *Haliotis kamtschatkana*, growth parameters from tag-recapture data. *Fisheries Research* 95, 289–295.
- Zheng, C.Z., Ovaskainen, O., Saastamoinen, M., Hanski, I., 2007. Age-dependent survival analyzed with Bayesian models of mark-recapture data. *Ecology* 88, 1970–1976.