

# Bayesian analysis of size-dependent overwinter mortality from size-frequency distributions

STEPHANIE M. CARLSON,<sup>1</sup> ATHANASIOS KOTTAS, AND MARC MANGEL

Center for Stock Assessment Research, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, California 95064 USA

**Abstract.** Understanding the relationship between body size and mortality is an important problem in ecology. We introduce a novel Bayesian method that can be used to quantify this relationship when the only data available are size-frequency distributions of unmarked individuals measured at two successive time periods. The inverse Gaussian distribution provides a parametric form for the statistical model development, and we use Markov chain Monte Carlo methods to evaluate posterior distributions. We illustrate the method using data on threespine stickleback (*Gasterosteus aculeatus*) collected before and after the winter season in an Alaskan lake. Our method allows us to compare the intensity of size-biased mortality in different years. We discuss generalizations that include more complicated relationships between size and survival as well as time-series modeling.

**Key words:** inverse Gaussian distribution; Markov chain Monte Carlo; Metropolis-Hastings algorithm; threespine stickleback.

## INTRODUCTION

Varying environmental conditions contribute to variation in population dynamics. For example, mortality rates are influenced by seasonal variation in environmental conditions such as floods (e.g., Chapman and Kramer 1991) and droughts (e.g., Elliott et al. 1997). Moreover, this mortality is often nonrandom with respect to individual phenotypes (observable traits), providing evidence for natural selection (e.g., Kingsolver et al. 2001). The consequence of selection is that individuals displaying certain trait values may be more or less likely to survive a given perturbation. The relationship between mortality and body size, in particular, has been studied extensively in wild populations (e.g., Sogard 1997, Kingsolver and Pfennig 2007, Carlson et al. 2008). Here, we use a novel Bayesian method to quantify this relationship when the only data available are size-frequency distributions of unmarked individuals measured at two successive time periods.

There are several empirical approaches for assessing size-biased mortality in wild populations. The various approaches fall into two general categories: (1) those that compare body size of individuals measured *before* and *after* an episode of selection (e.g., pre-winter body size of one group of individuals is compared to the post-winter body size of another group of individuals); and (2) those that compare body size measured at a *common* time

between two groups (e.g., pre-winter body size is compared between a group of individuals alive before the winter and the subset of individuals known to have survived the winter). The various approaches are each associated with unique strengths and weaknesses, which influence the uncertainties associated with estimating size-biased mortality. The four most common methods include:

1. *Cross-sectional studies of unmarked individuals.*—Collecting data on unmarked individuals at multiple sampling events is exceedingly common in ecological studies. These data are only sometimes used to assess size-biased mortality (e.g., Toney and Coble 1979). Because one group of individuals is measured for body size *before* and a second group is measured *after* an episode of selection, one source of uncertainty is the extent that the post-selection trait values are influenced by *growth*. A second source of uncertainty is the extent that post-selection trait values are influenced by *size-biased movement* (both immigration and emigration).

2. *Cross-sectional studies of batch-marked individuals.*—By batch-marking a group of individuals during the first sampling event, it is possible to include only the *subset* of recaptured (i.e., marked) individuals in analyses. Because the two groups are measured for body size before and after an episode of selection, the potential for *growth* to influence the post-selection trait distribution is an issue with this method as well. However, by batch-marking organisms, analyses are no longer biased by size-biased immigration because only marked individuals (i.e., individuals present in the study area at an earlier time) are included in analyses. *Size-biased emigration*, however, remains an issue.

3. *Longitudinal studies tracking the fates of individually-marked organisms.*—By tracking the fates of indi-

Manuscript received 17 February 2009; revised 8 July 2009; accepted 10 July 2009. Corresponding Editor: A. M. Ellison.

<sup>1</sup> Present address: Department of Environmental Science, Policy, and Management, University of California, Berkeley, California 94720 USA.

E-mail: smcarlson@berkeley.edu

viduals through time, it is possible to compare the size of individuals that are recaptured in the future (i.e., survived the episode of selection) vs. the size of the *subset* of individuals that were never recaptured (i.e., died or emigrated). These data are not influenced by growth because body size is measured at a common time (i.e., *before* the episode of selection for both groups). By including only marked individuals in analyses, size-biased immigration will not influence the post-selection trait distribution although disentangling *size-biased emigration* from size-biased mortality remains problematic (e.g., Letcher et al. 2005).

4. *Back-calculating size in fishes.*—An additional method exists for fishes, which entails back-calculating body size at an earlier time based on a modeled relationship between body size and a fish chronometric structure (e.g., otoliths [ear stones], opercula, scales). Using this approach, fish chronometric structures are sampled from two groups of individuals—one before and one after an episode of selection—and their body size at some *common* earlier time is back-calculated (e.g., size-at-emergence). The relationship between body size and fish chronometric structures is often assumed to be linear although this may be a poor assumption (e.g., Wright et al. 1990). One new source of uncertainty with this approach, therefore, is the error associated with estimating body size from other body parts. As with the longitudinal approach focusing on individuals, growth is no longer an issue because size is back-calculated to a *common* earlier time and then compared at that earlier time (e.g., Good et al. 2001, Johnston et al. 2005). However, *size-biased movement* between the two sampling occasions could influence the post-selection trait distribution.

In addition to these issues, there are also economic and ethical considerations. For instance, the cost and effort required to tag and track individually marked organisms greatly surpasses that required to collect data on unmarked individuals at successive time periods. The back-calculation method, which is often used to study size-biased mortality in fishes, may often require sacrificing fish to, for example, extract otoliths or opercula.

To determine the relative frequency of each of the above approaches for studying size-biased mortality, we performed a literature review. We sampled 50 papers identified with Google Scholar using the search terms “fish” and one of the following terms: “size-dependent mortality,” “size-dependent survival,” or “size-selective mortality.” From each paper, we determined which of the four approaches had been used. We focused our review on fishes to include the fourth category (i.e., back-calculation approach). We found that the most common approaches focused on a comparison of unmarked individuals or the back-calculation approach (Fig. 1). Despite the prevalence of studies focusing on unmarked individuals sampled at successive time intervals, existing approaches for quantifying the relationship between body size and mortality from these data are unsatisfactory.

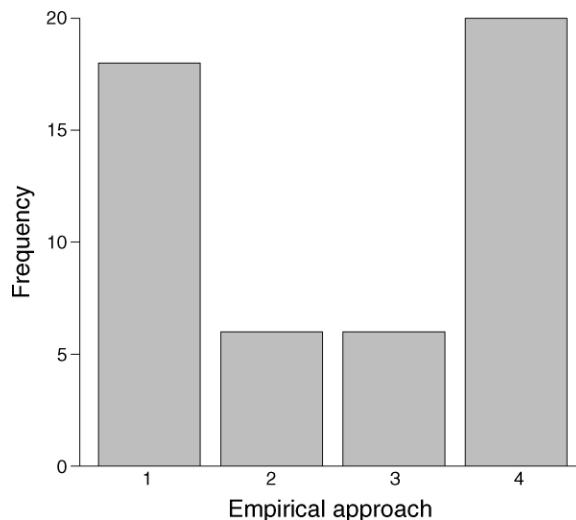


FIG. 1. Results of a literature review in which we determined the frequency of four approaches for studying size-biased mortality in fishes from a sample of 50 peer-reviewed papers related to this topic. The bars in the graph correspond, from left to right, to (1) cross-sectional studies of unmarked individuals; (2) cross-sectional studies of batch-marked individuals; (3) longitudinal studies tracking the fates of individually marked organisms; and (4) studies back-calculating size. The 50 papers included in our review were selected by searching the following terms on Google Scholar: “fish” in conjunction with “size-dependent mortality,” “size-dependent survival,” or “size-selective mortality.”

The statistical analysis of size-dependent mortality based on pairs of length-frequency distributions typically proceeds as a comparison of means via a simple ANOVA, which is a test of linear (directional) selection. A shift towards larger average size combined with a decrease in variance is often interpreted as evidence of positive linear selection (Brodie et al. 1995), that is, selection favoring relatively large individuals. While this approach can reveal whether there has been a significant shift in mean body size after an episode of selection, it may miss features of nonlinearity in the relationship between size and survival.

Here, we introduce an approach to estimate the survival function based on pairs of length-frequency distributions that allows for more flexible fitness functions than the standard approach described above. This approach can be applied to data from unmarked individuals with one major caveat. In particular, when data are based on groups of individuals measured at successive time periods (i.e., cross-sectional studies), our approach is appropriate when growth is negligible over the interval of interest. Previous research has focused on separating size-biased growth from size-biased mortality when both growth and mortality are occurring over the interval of interest (e.g., Post and Evans 1989, Munch et al. 2003). No growth during the winter season is a common assumption for populations residing at high latitudes and/or altitudes (e.g., Elliott 1994): an assumption that has been empirically validated in several systems. For example,

previous work on threespine stickleback (*Gasterosteus aculeatus*) suggests that growth ceases during winter when temperatures are low (Allen and Wootton 1982). No overwinter growth was observed for a high-latitude population of Atlantic silverside (*Menidia menidia*), although overwinter growth was documented for a low-latitude population (Munch et al. 2003). Similarly, in many locations in the southern ocean, Antarctic krill (*Euphausia superba*) do not grow at all during the Austral winter (May–November; Kawaguchi et al. 2006, 2007) and size-frequency distributions are commonly used in the study of krill (e.g., Quetin and Ross 2003).

We use a Bayesian modeling approach to estimate the mortality function from a pair of length-frequency distributions, which has several benefits. First, an analysis based on length–frequency distributions can be applied to data collected on unmarked individuals whereas other approaches, e.g., a logistic regression of survival on body size or Cormack-Jolly-Seber modeling with size as a covariate (see Carlson et al. 2008 for an example of these two methods), require data collected on marked individuals. Second, the approach is based on a structured parametric model so that values for the parameters of the survival function can be extracted and compared among studies. Third, the method allows us to generalize the structure of the survival function to allow for flexible fitness functions (see *Discussion*). Finally, the Bayesian approach yields full and exact inference and, in particular, it provides a framework for capturing the uncertainty associated with the parameter estimates, and thus also with the mortality function.

#### STUDY SPECIES AND DATA COLLECTION

The data concern threespine stickleback in Lake Aleknagik, Alaska, USA. Lake Aleknagik is an oligotrophic lake characterized by a short growing season in the summer followed by ice cover from approximately December through May (Hartman and Burgner 1972). Resident threespine stickleback and anadromous sockeye salmon (*Oncorhynchus nerka*) dominate the fish community in the lake. Stickleback breed in the near-shore area of the lake in June and July after the lake has become ice free (Rogers 1977). Most individuals in this system mature at age-3 although some large individuals mature at age-2 (age-0 corresponds to the first year of life, age-1 to the second year, and so on). By mid-July, most of the age-1 and older fish have migrated from the near-shore environment to the pelagic zone of the lake while the age-0 fish remain in the near-shore environment through August.

The Fisheries Research Institute at the University of Washington has been sampling the size distributions of the Lake Aleknagik fish community at the beginning and end of the growing season (i.e., post- and pre-winter data, respectively) each year since 1960 (Schindler et al. 2005). The pre-winter (i.e., end of the growing season) data are collected in the pelagic zone at night using a tow-net during the first week of September. Temperatures drop quickly in the fall until the lake freezes over, and so little

growth is presumed to occur after the September sampling (see also Schindler et al. 2005). Fall sampling proceeds by towing a  $3 \times 3$  m square net at the surface between two boats traveling approximately 3 km/h for five minutes at each of nine sampling sites around the lake, and these data are pooled. All captured fish are identified to species, enumerated, and a subset are measured for body length (fork length, mm). The post-winter sampling occurs in late June shortly after the lake becomes ice free. At this period, fish are still aggregated in the near-shore areas of the lake. Sampling is conducted using a 30-m beach seine deployed from a boat and manually pulled to shore. This procedure is repeated at 10 near-shore sites distributed around the lake and the data are then pooled. Again, all captured fish are identified to species, enumerated, and a subset are measured for body length. For further details on the tow-net and beach seine sampling, we refer readers to Robins et al. (2005) and Schindler et al. (2005).

We use data collected from three winters (1994–1996) to illustrate our approach to estimating the size dependence of mortality from pairs of length-frequency distributions.

#### METHODS

Our methodology is based on data on length measurements before (denoted by subscript b) and after (denoted by subscript a) winter for a specific year. We let  $y_{bi}$ ,  $i = 1, \dots, n_b$ , and  $y_{aj}$ ,  $j = 1, \dots, n_a$ , denote the pre-winter and post-winter observations, respectively, and let  $f_b(y)$  and  $f_a(y)$  (where  $y > 0$ ) denote the pre-winter and post-winter length densities, respectively. In general, these will be unrelated subsamples of a larger population so that  $n_a$  and  $n_b$  need not be equal and their relationship with total population size will be unknown.

#### *Probability models for the mortality function and length distributions*

During a non-breeding season, in a population that lacks structure, numbers will decline according to  $dN/dt = -mN$ , where  $N(t)$  is the number of individuals at time  $t$  and  $m$  is the mortality rate. In this case, the number of individuals declines exponentially in time. Size is one of the most fundamental metrics that structures populations and, in general, the rate of mortality will depend upon size. Also, in general, since sizes change due to growth concomitant with mortality, one must consider the number of individuals of a given size at a certain time. This density satisfies the Von Foerster partial differential equation (e.g., Murray 2002:37 and following pages). A situation of intermediate complexity, and which occurs commonly when organisms cease growth over the winter, is that in which mortality is structured by size but size does not change during the interval of interest. We may then characterize the population by  $N_y(t)$ , the number of individuals of size class  $y$  at time  $t$  and assume that  $dN_y/dt = -m(y)N_y$ , where the rate of mortality now depends upon size. The solution of this equation from  $t = 0$  to  $t = T$  is  $N_y(t) = N_y(0) \exp[-T \times m(y)]$  from which we

define  $M(y) = T \times m(y)$ , which we call the size-dependent mortality function. Note that  $M(y)$  is dimensionless. In this paper, we develop methods to analyze mortality functions that have the following structure:

$$M(y) = \beta + \alpha y^{-1} \quad y > 0 \tag{1}$$

where  $\beta > 0$  and  $\alpha > 0$ . The parameter  $\beta$  (which is unitless) characterizes the intensity of size-independent mortality, and  $\alpha$  (which has units of length) characterizes the intensity of size-dependent mortality. To illustrate the method, we assume that larger individuals are in general more likely to survive. Survival over the winter is thus  $S(y) = \exp\{-M(y)\} = \exp(-\beta - \alpha y^{-1})$ , for  $y > 0$ .

The post-winter length density is determined by the pre-winter length density and the survival function through

$$f_a(y) = \frac{S(y)f_b(y)}{\int_0^\infty S(u)f_b(u)du} \quad y > 0 \tag{2}$$

the integral in the denominator being a normalizing constant such that  $f_a(y)$  is a proper density function. Hence, the expression in Eq. 2 is an example of a weighted distribution (see *Discussion* for details on the related literature). In particular, the post-winter length distribution arises as a weighted version of the pre-winter length distribution with weights determined by the survival function.

Therefore, based on the model formulation in Eqs. 1 and 2, the intercept of the mortality function can not be estimated using data on length only. Since  $\beta$  cancels out of the numerator and denominator in Eq. 2, we are unable to estimate it from two frequency distributions. In order to estimate  $\beta$  one would require additional information on absolute population size before and after the winter. Thus, we define the survival function as

$$S(y) = \exp(-\alpha y^{-1}) \quad y > 0 \tag{3}$$

to include only the size-dependent portion of the mortality function.

To complete the model formulation, we need to specify the distributional form for the pre-winter length density. Note that, depending on the choice of  $f_b(y)$ , the required normalizing constant,  $\int_0^\infty S(u)f_b(u)du$ , for  $f_a(y)$  may not be available in closed form, which complicates the corresponding computational technique for Bayesian inference. Hence, the choice of  $f_b(y)$  needs to balance model flexibility with computational feasibility. We use the inverse Gaussian distribution (e.g., Seshadri 1999) as a parametric form for the pre-winter length distribution that achieves this desired balance. Under the inverse Gaussian distribution, the pre-winter length density is given by

$$f_b(y) = \left(\frac{\phi\mu}{2\pi y^3}\right)^{1/2} \exp\left(\phi - \frac{\phi\mu}{2y} - \frac{\phi y}{2\mu}\right) \quad y > 0 \quad \phi > 0 \quad \mu > 0. \tag{4}$$

where  $\mu$  is the mean of the inverse Gaussian distribution, and  $\phi$  is a shape parameter. Moreover, the variance is given by  $\mu^2\phi^{-1}$ , and thus, the coefficient of variation is  $\phi^{-1/2}$ .

Working with the survival function in Eq. 3 and the pre-winter density in Eq. 4, Eq. 2 yields a post-winter length density  $f_a(y)$  that corresponds again to an inverse Gaussian distribution with updated parameters that depend on  $\mu$ ,  $\phi$ , and  $\alpha$ . Specifically, the mean of the post-winter inverse Gaussian distribution is given by

$$\tilde{\mu} = \mu \left(1 + \frac{2\alpha}{\mu\phi}\right)^{1/2} \tag{5}$$

and the shape parameter by

$$\tilde{\phi} = \phi \left(1 + \frac{2\alpha}{\mu\phi}\right)^{1/2}. \tag{6}$$

We note that  $\tilde{\mu} > \mu$  and  $\tilde{\mu}^2\tilde{\phi}^{-1} = \mu\phi^{-1}\tilde{\mu} > \mu^2\phi^{-1}$ ; thus, both the mean and the variance of the post-winter length distribution are increased relative to the pre-winter length distribution. In the discussion, we show how the modeling framework can be extended if this is not a plausible assumption for a particular data set (see *Discussion*).

### Bayesian model

The likelihood for the full data vector  $\mathbf{y} = (\{y_{bi}: i = 1, \dots, n_b\}, \{y_{aj}: j = 1, \dots, n_a\})$  is built from independent product forms for the pre-winter and post-winter data components:

$$L(\mu, \phi, \alpha; \mathbf{y}) = \left[\prod_{i=1}^{n_b} f_b(y_{bi}; \mu, \phi)\right] \times \left[\prod_{j=1}^{n_a} f_a(y_{aj}; \mu, \phi, \alpha)\right]. \tag{7}$$

Ignoring terms that do not depend on the parameters, we have

$$\begin{aligned} L(\mu, \phi, \alpha; \mathbf{y}) &\propto h(\mu, \phi, \alpha; \mathbf{y}) \\ &= (\phi\mu)^{n_b/2} (\phi\mu + 2\alpha)^{n_a/2} \\ &\times \exp\left\{\phi n_b - \frac{\phi\mu}{2} \left(\sum_{i=1}^{n_b} y_{bi}^{-1}\right) - \frac{\phi}{2\mu} \left(\sum_{i=1}^{n_b} y_{bi}\right)\right\} \\ &\times \exp\left\{n_a\phi \left(1 + \frac{2\alpha}{\mu\phi}\right)^{1/2} - \frac{\phi\mu + 2\alpha}{2} \left(\sum_{j=1}^{n_a} y_{aj}^{-1}\right) \right. \\ &\quad \left. - \frac{\phi}{2\mu} \left(\sum_{j=1}^{n_a} y_{aj}\right)\right\}. \end{aligned} \tag{8}$$

The full Bayesian model is completed with prior distributions for  $\mu$ ,  $\phi$ , and  $\alpha$ . We work with independent gamma distributions for these parameters. For example

(Mangel 2006), the prior density for  $\mu$  is

$$\pi(\mu) = \frac{b^a \mu^{a-1} \exp(-b\mu)}{\Gamma(a)} \propto \mu^{a-1} \exp(-b\mu) \quad (9)$$

with analogous expressions for the prior densities  $\pi(\phi)$  and  $\pi(\alpha)$ . Here, the prior hyperparameters  $a > 0$  and  $b > 0$  are fixed, and  $\Gamma(a)$  denotes the Gamma function with parameter  $a$ .

The role of parameter  $\mu$  as the mean of the pre-winter length distribution facilitates specification of its prior hyperparameters. In particular, a prior guess at the center of the pre-winter length data can be used to specify the prior mean for  $\mu$  (given by  $ab^{-1}$  in Eq. 9), with the rate parameter,  $b$ , chosen to provide a dispersed prior (note that the variance of Eq. 9 is given by  $ab^{-2}$ ). Moreover, a prior guess at the range of the pre-winter length data can be converted to a proxy for the variance,  $\mu^2\phi^{-1}$ , of the pre-winter length distribution. Hence, having specified the prior for  $\mu$ , this information can be used to set the prior mean for  $\phi$ , again, with a rate parameter chosen to yield a sufficiently dispersed gamma prior for  $\phi$ . Finally, although direct prior information for  $\alpha$  may not be available, its prior can be specified based on the expression for the mean or shape parameter of the post-winter length distribution (Eqs. 5 and 6, respectively). For instance, having chosen the priors for  $\mu$  and  $\phi$ , a prior guess at the center of the post-winter length data yields the prior mean for  $\alpha$ , using Eq. 5.

Given the data, the posterior distribution for the model parameters can be expressed as

$$p(\mu, \phi, \alpha | \mathbf{y}) = C^{-1} h(\mu, \phi, \alpha; \mathbf{y}) \pi(\mu) \pi(\phi) \pi(\alpha) \quad (10)$$

where  $h(\mu, \phi, \alpha; \mathbf{y})$  is defined in Eq. 8, and  $C$  is the normalizing constant for the posterior given by

$$C = \int_0^\infty \int_0^\infty \int_0^\infty h(\mu, \phi, \alpha; \mathbf{y}) \pi(\mu) \pi(\phi) \pi(\alpha) d\mu d\phi d\alpha. \quad (11)$$

This integral is not available in closed form, and we therefore use a Markov chain Monte Carlo (MCMC) posterior simulation algorithm (Clark 2007) for inference as detailed in the next section.

*Markov chain Monte Carlo method for posterior inference*

We used a Metropolis-Hastings algorithm (see, e.g., Robert and Casella 2004) to sample from the posterior distribution in Eq. 10. (We have included R code to implement the algorithm in the Supplement.) The key aspect of the Metropolis-Hastings posterior simulation method is the proposal distribution, which is used to propose new values for parameters ( $\mu, \phi, \alpha$ ) that are then accepted or rejected in a stochastic fashion. To facilitate the choice of the proposal distribution, we work on the logarithmic scale for all parameters and use

a trivariate normal proposal distribution. Therefore, the proposal density on the original scale is given by

$$q(\mu, \phi, \alpha; \mathbf{m}, \mathbf{D}) = \mu^{-1} \phi^{-1} \alpha^{-1} \mathcal{N}_3([\log(\mu), \log(\phi), \log(\alpha)]; \mathbf{m}, \mathbf{D}) \quad (12)$$

where  $\mathcal{N}_3(\cdot; \mathbf{m}, \mathbf{D})$  denotes the trivariate normal density with mean vector  $\mathbf{m}$  and covariance matrix  $\mathbf{D}$ .

We let  $(\mu^{(t)}, \phi^{(t)}, \alpha^{(t)})$  denote the current state (iteration) of the Metropolis-Hastings MCMC algorithm. Then to move to the next state  $(\mu^{(t+1)}, \phi^{(t+1)}, \alpha^{(t+1)})$ , we implement the following steps:

1) Draw the proposed set of values  $(\mu^*, \phi^*, \alpha^*)$  from the proposal distribution with density given by Eq. 12, with mean vector  $\mathbf{m} \equiv \mathbf{m}^{(t)} = (\log(\mu^{(t)}), \log(\phi^{(t)}), \log(\alpha^{(t)}))$ , and covariance matrix  $\mathbf{D}$  developed as discussed below.

2) Set  $(\mu^{(t+1)}, \phi^{(t+1)}, \alpha^{(t+1)}) = (\mu^*, \phi^*, \alpha^*)$  with probability  $p$ , and  $(\mu^{(t+1)}, \phi^{(t+1)}, \alpha^{(t+1)}) = (\mu^{(t)}, \phi^{(t)}, \alpha^{(t)})$  with probability  $1 - p$ , where the acceptance probability is

$$p = \min \left\{ 1, \frac{\tilde{p}(\mu^*, \phi^*, \alpha^* | \mathbf{y})}{\tilde{p}(\mu^{(t)}, \phi^{(t)}, \alpha^{(t)} | \mathbf{y})} \times \frac{\mu^* \phi^* \alpha^*}{\mu^{(t)} \phi^{(t)} \alpha^{(t)}} \right\} \quad (13)$$

where  $\tilde{p}(\mu, \phi, \alpha | \mathbf{y}) = h(\mu, \phi, \alpha; \mathbf{y}) \pi(\mu) \pi(\phi) \pi(\alpha)$  is the unnormalized posterior density in Eq. 10.

The covariance matrix,  $\mathbf{D}$ , of the proposal distribution is key to the performance of the algorithm. To develop it, we started with a diagonal covariance matrix (that is, with independent normal proposal distributions for the parameters on the logarithmic scale), and experimented with various sets of values for the diagonal elements in order to obtain a good acceptance rate. The resulting samples for parameters ( $\mu, \phi, \alpha$ ) provided an initial approximation to the posterior distribution. Next, we used the empirical covariance matrix from these initial samples (on the logarithmic scale) to estimate the proposal covariance matrix,  $\mathbf{D}$ , and then ran the algorithm with this improved covariance matrix to obtain the final results. This approach resulted in an efficient implementation of the Metropolis-Hastings algorithm with fast convergence and good mixing. For all the results discussed in the next section, the acceptance rate ranged from 41% to 46%. To obtain the final sets of draws from the posterior distribution, we used a (conservative) burn-in period of 10 000 iterations, and a thinning rate of 10 iterations to eliminate autocorrelations in the posterior samples.

We let  $\{(\mu_\ell, \phi_\ell, \alpha_\ell): \ell = 1, \dots, L\}$  denote the  $L$  samples from the posterior distribution  $p(\mu, \phi, \alpha | \mathbf{y})$ , obtained after burn-in and thinning. (For the results reported in the next section, the posterior sample size was  $L = 9000$ .) These posterior samples can be used directly for inference for the three model parameters (as in Fig. 2). Importantly, they can also be used to estimate the survival function as well as the pre-winter and post-

winter length densities. For instance, for any specified value of body length, e.g.,  $y_0$ , the samples  $\{\exp(-\alpha y_0^{-1}): \ell = 1, \dots, L\}$  provide the posterior distribution of survival at length  $y_0$ . Repeating over a grid of length values within an interval of interest, we obtain the entire posterior distribution for the survival function (up to the chosen grid). Using the posterior mean and two percentiles for each grid point, we can plot posterior point and interval estimates for the survival function, and using an analogous procedure, also for the pre-winter and post-winter length densities. An example of this type of inference is reported in Fig. 3.

RESULTS

We first consider the pre-winter and post-winter threespine stickleback length data, discussed in *Study species and data collection*, from the year 1996. We followed the approaches to prior specification and posterior inference detailed in *Methods*. In particular, to determine the influence of prior hyperparameter values, we performed a prior sensitivity analysis using two distinct prior choices. The first prior choice corresponds to the following prior hyperparameter values:  $\mu$  (mean = 46, rate = 40/46);  $\phi$  (mean = 114, rate = 20/114); and  $\alpha$  (mean = 500, rate = 5/500). The second prior choice was based on the following prior hyperparameter values:  $\mu$  (mean = 50, rate = 500/50);  $\phi$  (mean = 150, rate = 20/150); and  $\alpha$  (mean = 900, rate = 15/900). The resulting posterior inference was very robust for all three parameters (Fig. 2). Indeed, the two posterior distributions for  $\mu$  were practically indistinguishable despite a substantial difference in the corresponding priors (Fig. 2A). Similarly, little difference was detected between the posterior distributions for  $\phi$  (Fig. 2B) and  $\alpha$  (Fig. 2C).

Next, we used our first choice of prior hyperparameter values to obtain posterior point estimates (given by posterior means) and 90% interval estimates for the pre-winter (Fig. 3A) and post-winter (Fig. 3B) length densities. Moreover, we generated posterior point and 90% interval estimates for survival as a function of body length, which revealed a marked survival advantage for longer fish (Fig. 3C).

Because  $\alpha$  is the key parameter in determining the relationship between survival and body length, we next repeated the analysis for two additional winters (1994 and 1995) to explore among-year variation in the strength and form of this relationship. Detailed inference results for these additional years are provided in the Appendix (Figs. A1 and A2). A comparison of the year-specific posterior distributions for  $\alpha$  suggests substantial variation among years in the relationship between body size and survival (Fig. 4). Non-overlapping posterior distributions suggest that the relationship between survival and body length during the winter of 1994 differed from that in the winters of 1995 and 1996 (Fig. 4; also, compare Figs. 3C, A1F, A2F).

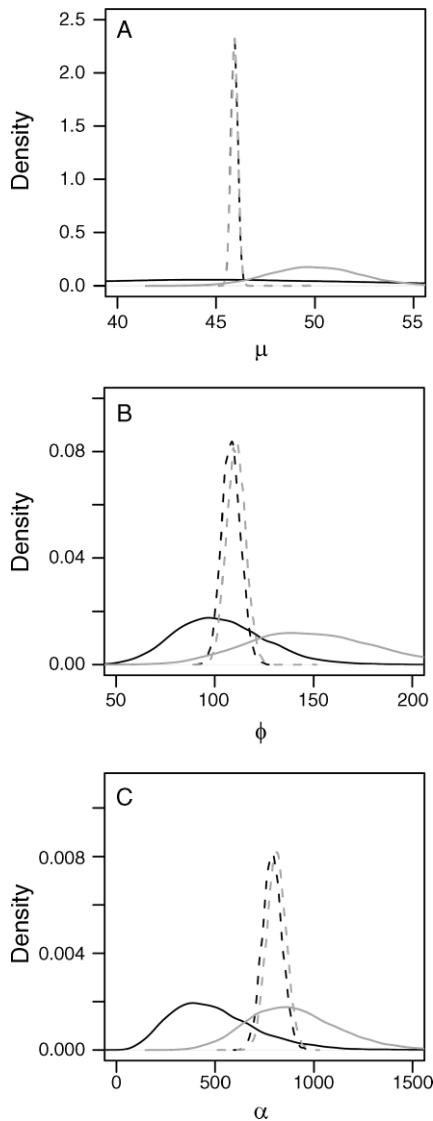


FIG. 2. Prior and posterior distributions for (A)  $\mu$ , (B)  $\phi$ , and (C)  $\alpha$ , where  $\mu$  and  $\phi$  represent the mean and shape parameters, respectively, of the pre-winter inverse Gaussian length distribution, and where  $\alpha$  represents a parameter describing the intensity of size-dependent mortality. Here  $\mu$  has units of length (on the original scale of the data),  $\phi$  is unitless, and  $\alpha$  has units of length, but at a scale driven by Eqs. 1 or 3. We used gamma priors for all three parameters. For each parameter, we show two prior choices (solid lines) and the resulting posterior distributions (dashed lines). The first prior choice (represented in black), corresponds to the following prior hyperparameter values:  $\mu$ , mean = 46, rate = 40/46;  $\phi$ , mean = 114, rate = 20/114; and  $\alpha$ , mean = 500, rate = 5/500. The second prior choice (represented in gray), corresponds to the following prior hyperparameter values:  $\mu$ , mean = 50, rate = 500/50;  $\phi$ , mean = 150, rate = 20/150; and  $\alpha$ , mean = 900, rate = 15/900.

DISCUSSION

Empirical approaches for assessing size-biased mortality in nature span the range from cross-sectional studies in which the size distribution of unmarked

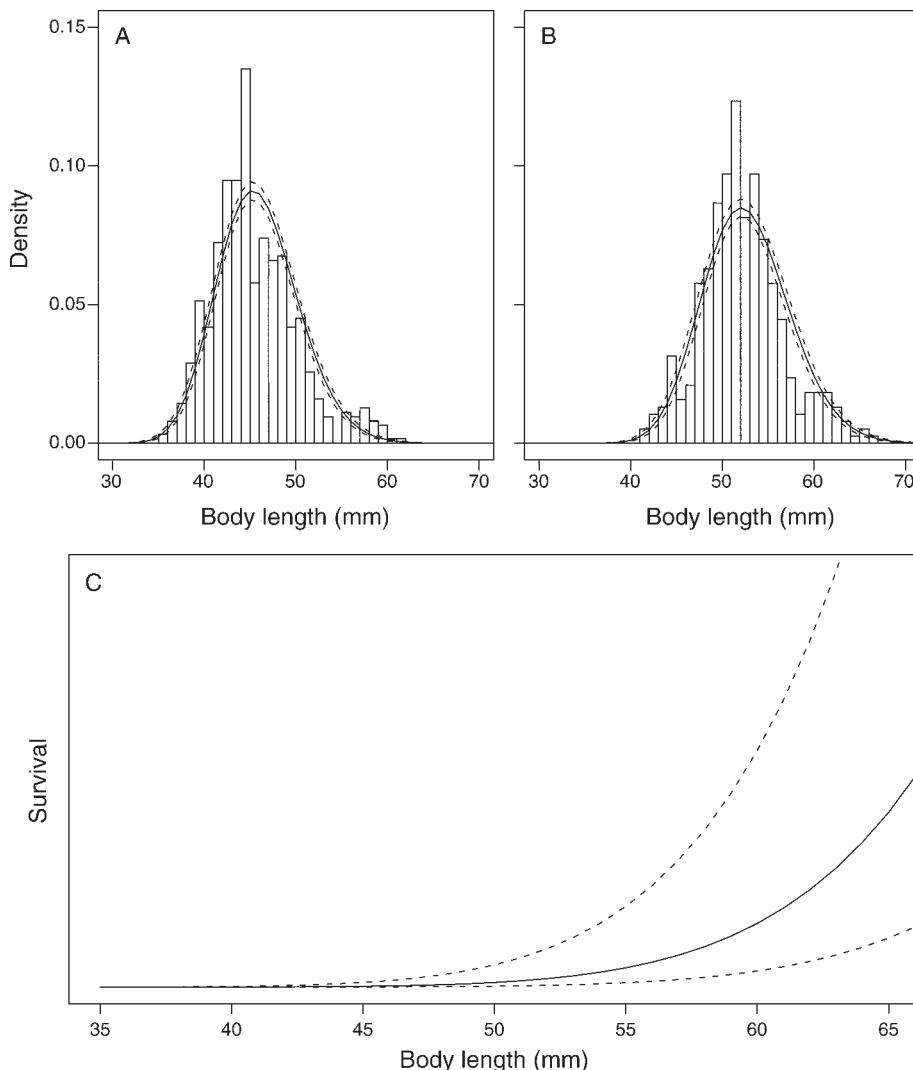


FIG. 3. Posterior mean estimates (solid lines) and 90% interval estimates (dashed lines) for the (A) pre-winter length density, (B) post-winter length density, and (C) survival as a function of body length. The pre-winter and post-winter length observations are included in panels A and B, respectively. Results are based on the first prior choice (see Fig. 2 legend).

individuals is characterized at successive time intervals (e.g., Toney and Coble 1979) to longitudinal studies in which the fates of marked and measured individuals are tracked through time (e.g., Carlson et al. 2008). Our review of these methods in the *Introduction* makes explicit the nature of the uncertainties associated with each method, as this allows researchers to make sensible decisions regarding their approach in the field. In short, longitudinal studies tracking the fates of marked individuals are preferred because estimates are not confounded by size-biased immigration or growth, but the tagging and tracking of individually marked organisms requires considerable effort and can be cost-prohibitive. Consequently, far more data have been collected from unmarked individuals sampled at successive time intervals. Despite this, the current approaches for quantifying the relationship between body size and

mortality from such data are unsatisfactory. We have developed a novel Bayesian modeling approach to estimate the size dependence of mortality from pairs of size-frequency histograms.

Our results show that this parametric Bayesian approach can provide new and important information about survival when the only information available is size distributions of unmarked individuals. To illustrate the method, we focused on a relatively simple mortality function in which the rate of mortality declines with body size. However, our method would also apply for a more complicated mortality rate of the form:  $M(y) = \beta + \alpha_1 y^{-1} + \alpha_2 y$ ,  $y > 0$ , with  $\alpha_1 \geq 0$  and  $\alpha_2 \geq 0$ . The two parameters  $\alpha_1$  and  $\alpha_2$  now determine the relative strength of mortality, which is minimized at length  $\sqrt{\alpha_2/\alpha_1}$ . Thus, we can account for mortality declining with size ( $\alpha_2$  very small or 0), increasing with size ( $\alpha_1$

very small or 0), or the intermediate case with a size that minimizes mortality within the range of the data (i.e., stabilizing or disruptive selection). For example, if a preliminary analysis of a data set suggested that mortality increased with size we might have a prior that supports values of  $\alpha_1$  close to 0. With this more general mortality function, the inverse Gaussian distribution shows its real strength because of the conjugacy of the inverse Gaussian and the more general mortality rate. Although the inverse Gaussian distribution has been used in a range of applications (Seshardi 1999), the application that we report here is novel.

As discussed briefly in *Methods*, the formulation in Eq. 2 for the post-winter length distribution is an example of the class of weighted distributions (e.g., Patil 2002). In particular, the weighted version of a non-negative random variable with density  $f(y)$  has density given by  $f^w(y) = w(y)f(y)/\{\int w(u)f(u) du\}$ , where the nonnegative weight (recording) function  $w(y)$  controls the probability that a realization  $y$  under  $f(y)$  is recorded. Resource selection models (e.g., Thomas et al. 2006, Johnson et al. 2008) provide a particular application area where resources are selected by animals from a distribution of available resources (with density  $f(y)$ ) which is transformed to the “use” distribution (with density  $f^w(y)$ ) through the resource selection function  $w(y)$ . Length-biased (size-biased) distributions arise as a special case of the class of weighted distributions with  $w(y) = y$ . Moreover, length-biased versions of the inverse Gaussian distribution are studied in Gupta and Akman (1995) and Sansgiry and Akman (2001). Finally, there is a fairly rich statistical literature on selection models, where the function  $w(y)$  is, typically, assumed to be non-decreasing and bounded (see e.g., Lee and Berger 2001).

Note that we analyzed the before/after winter size-frequency distributions for each of three years as separate entities. However, it is reasonable to expect that the parameters of the mortality rate function are at least weakly correlated in time. Our approach can be extended to incorporate hierarchical or time series structure in the model for the pre-winter and post-winter size-frequency distributions. A generic hierarchical modeling approach would involve year-specific mortality parameters  $\alpha_t$  (where  $t$  indexes the years) arising from a random effects distribution. For added flexibility, the hierarchical formulation could be extended to the parameters  $\mu_t$  and  $\phi_t$  of the inverse Gaussian pre-winter length distribution. If specific forms of temporal dependence are plausible, more structured dynamic priors for the  $\alpha_t$  and/or the  $(\mu_t, \phi_t)$  could be formulated.

In conclusion, we have presented a novel Bayesian parametric approach for analyzing size-dependent mortality based on pairs of size-frequency histograms. This approach has great appeal because (1) it is flexible and can be applied to data collected from unmarked individuals; (2) it allows an assessment of the uncertain-

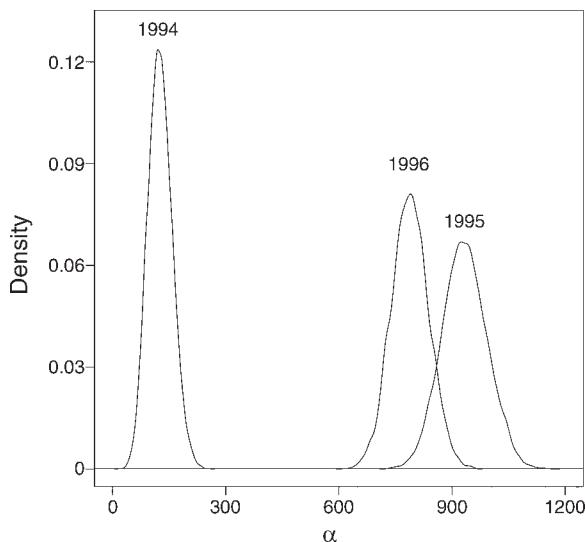


FIG. 4. Posterior densities for  $\alpha$  based on stickleback length-frequency data collected over three winters (1994, 1995, 1996).

ties associated with this method; and (3) can be generalized to consider more complicated mortality functions as well as time-series modeling for the data.

#### ACKNOWLEDGMENTS

We thank the Alaska Salmon Program at the University of Washington for graciously sharing their long-term stickleback data set with us. We also thank Kate Richerson for performing the literature review needed to create Fig. 1, and two reviewers for helpful comments. S. Carlson was supported by an NSF Postdoctoral Research Fellowship in Biological Informatics (DBI-0630626) and by the Center for Stock Assessment Research; A. Kottas was supported in part by NSF grant DEB-0727543; M. Mangel was partially supported by the Lenfest Ocean program.

#### LITERATURE CITED

- Allen, J. R. M., and R. J. Wootton. 1982. Age, growth and rate of food consumption in an upland population of the three-spined stickleback, *Gasterosteus aculeatus* L. *Journal of Fish Biology* 21:95–105.
- Brodie, E. D., A. J. Moore, and F. J. Janzen. 1995. Visualizing and quantifying natural selection. *Trends in Ecology and Evolution* 10:313–318.
- Carlson, S. M., E. M. Olsen, and L. A. Vøllestad. 2008. Seasonal mortality and the effect of body size: A review and an empirical test using individual data on brown trout. *Functional Ecology* 22:663–673.
- Chapman, L. J., and D. L. Kramer. 1991. The consequences of flooding for the dispersal and fate of poeciliid fish in an intermittent tropical stream. *Oecologia* 87:299–306.
- Clark, J. S. 2007. *Models for ecological data*. Princeton University Press, Princeton, New Jersey, USA.
- Elliott, J. M. 1994. *Quantitative ecology and the brown trout*. Oxford University Press, New York, New York, USA.
- Elliott, J. M., M. A. Hurley, and J. A. Elliott. 1997. Variable effects of droughts on the density of a sea-trout *Salmo trutta* population over 30 years. *Journal of Applied Ecology* 34: 1229–1238.
- Good, S. P., J. J. Dodson, M. G. Meekan, and D. A. J. Ryan. 2001. Annual variation in size-selective mortality of Atlantic

- salmon (*Salmo salar*) fry. Canadian Journal of Fisheries and Aquatic Sciences 58:1187–1195.
- Gupta, R. C., and H. O. Akman. 1995. On the reliability studies of a weighted inverse Gaussian model. Journal of Statistical Planning and Inference 48:69–83.
- Hartman, W. L., and R. L. Burgner. 1972. Limnology and fish ecology of sockeye salmon nursery lakes of the world. Journal of the Fisheries Research Board of Canada 29:699–715.
- Johnson, D. S., D. L. Thomas, J. M. Ver Hoef, and A. Christ. 2008. A general framework for the analysis of animal resource selection from telemetry data. Biometrics 64:968–976.
- Johnston, P., N. E. Bergeron, and J. J. Dodson. 2005. Assessment of winter size-selective mortality of young-of-the-year Atlantic salmon (*Salmo salar*) using otolith microstructure analysis. Ecology of Freshwater Fish 14:168–176.
- Kawaguchi, S., S. G. Candy, R. King, M. Naganobu, and S. Nicol. 2006. Modeling growth of Antarctic krill I. Growth trends with sex, length, season, and region. Marine Ecology Progress Series 306:1–15.
- Kawaguchi, S., L. A. Finley, S. Jarman, S. G. Candy, R. M. Ross, L. B. Quetin, V. Siegel, W. Trivelpiece, M. Naganobu, and S. Nicol. 2007. Male krill grow fast and die young. Marine Ecology Progress Series 345:199–210.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. American Naturalist 157:245–261.
- Kingsolver, J. G., and D. W. Pfennig. 2007. Patterns and power of phenotypic selection in nature. BioScience 57:561–572.
- Lee, J., and J. O. Berger. 2001. Semiparametric Bayesian analysis of selection models. Journal of the American Statistical Association 96:1397–1409.
- Letcher, B. H., G. E. Horton, T. L. Dubreuil, and M. J. O'Donnell. 2005. A field test of the extent of bias in selection estimates after accounting for emigration. Evolutionary Ecology Research 7:643–650.
- Mangel, M. 2006. The theoretical biologist's toolbox. Cambridge University Press, Cambridge, UK.
- Munch, S. B., M. Mangel, and D. O. Conover. 2003. Quantifying natural selection on body size from field data: winter mortality in *Menidia menidia*. Ecology 84:2168–2177.
- Murray, J. D. 2002. Mathematical biology. I. An introduction. Third edition. Springer, New York, New York, USA.
- Patil, G. P. 2002. Weighted distributions. Pages 2369–2377 in A. H. El-Shaarawi and W. W. Piegorsch, editors. Encyclopedia of environmetrics. Volume 4. Wiley, Chichester, UK.
- Post, J. R., and D. O. Evans. 1989. Size-dependent over-winter mortality of young-of-the-year yellow perch (*Perca flavescens*): laboratory, in situ enclosure, and field experiments. Canadian Journal of Fisheries and Aquatic Sciences 46:1958–1968.
- Quetin, L. B., and R. M. Ross. 2003. Episodic recruitment in the Antarctic krill *Euphausia superba* in the Palmer LTER study region. Marine Ecology Progress Series 259:185–200.
- Robert, C. P., and G. Casella. 2004. Monte Carlo statistical methods. Second edition. Springer, New York, New York, USA.
- Robins, J. B., C. A. Abrey, T. P. Quinn, and D. E. Rogers. 2005. Lacustrine growth of juvenile pink salmon and a comparison with sympatric sockeye salmon. Journal of Fish Biology 66:1671–1680.
- Rogers, D. E. 1977. Collection and analysis of biological data from the Wood River lake system, Nushagak District, Bristol Bay, Alaska: biology of the threespine stickleback in the Wood River lakes. Fisheries Research Institute, University of Washington, Seattle, Washington, USA.
- Sangry, P. S., and O. Akman. 2001. Reliability estimation via length-biased transformation. Communications in Statistics, Theory and Methods 30:2473–2479.
- Schindler, D. E., D. E. Rogers, M. D. Scheuerell, and C. A. Abrey. 2005. Effects of changing climate on zooplankton and juvenile sockeye salmon growth in southwestern Alaska. Ecology 86:198–209.
- Seshadri, V. 1999. The inverse Gaussian distribution, statistical theory and applications. Springer, New York, New York, USA.
- Sogard, S. M. 1997. Size-selective mortality in the juvenile stage of teleost fishes: a review. Bulletin of Marine Science 60:1129–1157.
- Thomas, D. L., D. Johnson, and B. Griffith. 2006. A Bayesian random effects discrete-choice model for resource selection: population-level selection inference. Journal of Wildlife Management 70:404–412.
- Toneys, M. L., and D. W. Coble. 1979. Size-related, first winter mortality of freshwater fishes. Transactions of the American Fisheries Society 108:415–419.
- Wright, P. J., N. B. Metcalfe, and J. E. Thorpe. 1990. Otolith and somatic growth rates in Atlantic salmon parr (*Salmo salar* L.): evidence against coupling. Journal of Fish Biology 36:241–249.

#### APPENDIX

Supplemental figures showing the results from the winters of 1994 and 1995 (*Ecological Archives* E091-071-A1).

#### SUPPLEMENT

R code to implement the Metropolis-Hastings algorithm (*Ecological Archives* E091-071-S1).