# A statistical modeling method for estimating mortality and abundance of spawning salmon from a time series of counts

# **R. Glenn Szerlong and David E. Rundio**

**Abstract:** We present a statistical modeling method for estimating mortality and abundance of spawning salmon from time-series counts that eliminates the need for separate information about mortality. We model arrival and mortality using differential equations, where mortality can be constant or changing linearly, and estimate mortality and abundance from counts using maximum likelihood when multiple estimates of detection rate are available. We also develop an approximate likelihood to estimate mortality and abundance when only a single value for detection rate is available or to estimate only mortality when detection rates are entirely unknown. We demonstrate our approach using counts of coho salmon (*Oncorhynchus kisutch*) where mortality, abundance, and detection were determined from tagging at a weir. Our model for nonconstant mortality produced mortality estimates that closely matched the empirical data and were robust to variation in other parameters. It also provided a better fit to the stream counts and a closer abundance estimate to the weir count than the constant mortality model. Monte Carlo simulations indicated that the approximate likelihood provided reasonable estimates of mortality over most of the ranges of parameters explored, particularly under the nonconstant mortality model, and produced relatively unbiased abundance estimates using a single value for detection.

Résumé : Nous présentons une méthode de modélisation statistique pour estimer la mortalité et l'abondance de saumons en fraie à partir de séries chronologiques de dénombrements sans nécessiter de renseignements séparés sur la mortalité. Nous modélisons l'arrivée et la mortalité à l'aide d'équations différentielles dans lesquelles la mortalité peut être constante ou changer de façon linéaire et nous estimons la mortalité et l'abondance d'après les dénombrements en utilisant une méthode de vraisemblance maximale lorsqu'il existe des estimations multiples du taux de détection. Nous avons aussi mis au point une vraisemblance approximative pour estimer la mortalité et l'abondance lorsqu'une seule valeur du taux de détection est connue ou alors pour estimer la mortalité seule lorsque le taux de détection est complètement inconnu. Nous illustrons notre méthodologie en déterminant la mortalité, l'abondance et la détection chez des saumons coho (Oncorhynchus kisutch) à l'aide de dénombrements obtenus par marquage à un barrage. Notre modèle à mortalité non constante produit des estimations de la mortalité qui correspondent de près aux valeurs empiriques et qui sont robustes en cas de variation des autres variables. Ce modèle produit aussi un meilleur ajustement aux dénombrements dans le cours d'eau et une estimation plus exacte de l'abondance à partir des dénombrements au barrage que le modèle à mortalité constante. Des simulations de Monte Carlo indiquent que la vraisemblance approximative permet de faire des estimations adéquates de la mortalité sur presque toute l'étendue des variables examinées, particulièrement avec le modèle à mortalité non constante; elle génère des estimations relativement peu faussées de l'abondance lorsqu'on utilise une seule valeur de détection.

[Traduit par la Rédaction]

# Introduction

The "area under the curve" (AUC) method is a common technique for estimating abundance of spawning salmon in streams and rivers (Neilson and Geen 1981; English et al. 1992; Irvine et al. 1992). By this method, counts of live fish made during periodic surveys are converted to an estimate of total abundance by fitting a curve to the entire time series of

counts (usually after adjusting the counts for detection rates less than one), integrating the curve numerically to calculate a total number of fish-days, and then dividing this total fishdays by the mean individual survival time in days. This method, while conceptually simple, has several limitations in practice. First, in addition to the time-series counts, it requires data on detection and survival rates that often are difficult to obtain. Second, because the method is essentially

Received 27 October 2006. Accepted 3 July 2007. Published on the NRC Research Press Web site at cjfas.nrc.ca on 13 December 2007. J19619

**R.G. Szerlong<sup>1</sup> and D.E. Rundio.<sup>2</sup>** National Marine Fisheries Service, Southwest Fisheries Science Center, 110 Shaffer Road, Santa Cruz, CA 95060, USA.

<sup>1</sup>Present address: 2817 A East Arthur Place, Seattle, WA 98112, USA. <sup>2</sup>Corresponding author (e-mail: dave.rundio@noaa.gov).

deterministic, there is no clear way to account for missing counts or to include uncertainty of detection and survival estimation in the final abundance estimate.

New statistical modeling approaches have been developed recently that address some of the limitations of traditional AUC methods to improve accuracy, estimate uncertainty, and permit estimation with fewer data. Quinn and Gates (1997) developed a general likelihood approach for estimating escapement of pink salmon (Oncorhynchus gorbuscha) by treating spawner abundance as an unknown but estimable function of a stochastic arrival-and-death model. Various functions for the immigration and mortality dynamics then can be fitted to the data by, for example, maximum likelihood (ML), and the resulting abundance estimates compared for accuracy against the observed time-series counts. Errors of estimation, where they occur, might be thought of as arising either from unpredictable variability in the random immigration and mortality components of the model or else as a consequence of imperfect sampling. The decided-upon error structure of the data thus implies a particular likelihood model, which (if sufficient information is available for likelihood maximization) then serves as a basis for estimating missing counts, estimating precision, and selecting among sets of possibly competing arrival or death models.

The major limitation of the method of Quinn and Gates (1997) is that extensive data on immigration and mortality dynamics are required to fit the model, and several adaptations of their general approach have been developed subsequently for situations in which these data are limited or unavailable. The most common approach ties a parametric function for arrival times and auxiliary data on mean mortality into likelihood models for sampling errors. For example, Hilborn et al. (1999) reduced the need for additional timeseries data on arrival and mortality by modeling temporal autocorrelation in abundance counts by a probability density curve and by constraining the expected number of deaths to depend on the time of peak abundance and the mean survival time (i.e., longevity) of spawners (for an extension to steelhead (Oncorhynchus mykiss), see Korman et al. 2002). Estimates of longevity are obtained from prior or auxiliary data (in these examples, from weir studies) and then used to specify a Bayesian prior distribution (see Carlin and Louis 2000) for maximum-likelihood abundance and precision estimation. Su et al. (2001) modified the model of Hillborn et al. (1999) to account for declining longevity over the spawning season (e.g., Neilson and Geen 1981; English et al. 1992; Korman et al. 2002) and also developed a hierarchical Bayesian method to use historical information on peak abundance and longevity to reduce uncertainty of estimates from incomplete time-series counts (Adkison and Su 2001). These customizations, while clearly more flexible than the traditional AUC approach, all retain a strong dependence on the prior estimate for longevity. Consequently, large errors in the abundance estimate can arise when the true longevity departs sufficiently from the assumed prior. In addition, because of the restrictions imposed on the parametrical model, there appears to be no way to infer mortality dynamics directly from the abundance counts.

In this paper, we develop a method for estimating both mortality and abundance from a single time series of livespawner counts, eliminating the need for prior or auxiliary information on longevity. Our approach is model-dependent, requiring the assumptions that all salmon die after spawning (i.e., they are semelparous; cf. Korman et al. 2002) and that the expected abundance change in the stream can be described as an arrival-death process. Our method closely follows that of Manly (1974) and Zonneveld (1991) for estimating mortality and abundance of emergent insects from serial transect counts, but departs where we allow mortality rates to change over time (see also Su et al. (2001)). Like both traditional AUC and the newer likelihood models (Hilborn et al. 1999; Su et al. 2001), our method requires additional information about detection rates in order to estimate abundance. However, we develop a model approximation that permits estimation of mortality even for cases when detection rates are unknown.

We present our approach in several steps. First, we introduce our model for the expected arrival and death dynamics. For simplicity, we begin by assuming that the rate of mortality is constant and then relax this assumption to allow for mortality rates that change linearly with time. Next, we develop a likelihood model that admits errors due to counting and uncertainty in the visual detection of fish for estimating mortality and abundance when estimates of detection rate are available for every count in the time series. Then we develop a second form of the likelihood to accommodate some missing estimates of detection. In addition, because data on detection may be available only as single-point estimates or, more often, lacking altogether from spawner surveys, we present an approximate likelihood model to estimate mortality and abundance using a single value for detection rate or to estimate only mortality when detection rates are entirely unknown. We use simulations to assess bias in estimates from this approximation under different parameter values. Finally, we demonstrate the methods using data that were collected in a tag-and-survey study for adult coho salmon (Oncorhynchus kisutch) in northern California where abundance was known from a weir.

# Methods

#### Modeling arrival and death

For semelparous species such as coho salmon, the number of live fish that are present in a stream at time t(x(t)), and therefore susceptible to visual sampling, depends largely on two biological processes: (*i*) arrival to the stream or the sample area and (*ii*) mortality or permanent emigration (Quinn and Gates 1997). Where sampling covers the entire stream length, removals from the sample are due mostly to mortality, and in this case, the inverse of the mortality rate (days<sup>-1</sup>) is equivalent to the adult longevity (sometimes called "stream-residence time" (English et al. 1992) or "stream life" (Hilborn et al. 1999)). When only part of the stream is sampled, the longevity estimate will include both mortality and permanent emigration.

To model the counting process through time, we first define the rate of abundance change, dx(t)/dt, as the instantaneous difference between the number of new arrivals and new deaths (Manly 1974; Zonneveld 1991; Quinn and Gates 1997). For simplicity, we assume that the overall pattern of arrival can be modeled by a two-parameter gamma density curve having mode at time  $t = (r - 1)/\lambda$ , where  $r \ge 1$  and  $\lambda > 0$ . The advantage of modeling arrival by a parametric curve is that the density function is assured to integrate to unity, thus permitting the total escapement to be modeled despite intermittent observation (Hilborn et al. 1999; Su et al. 2001; Korman et al. 2002). We selected the gamma curve because the gamma density function is constrained to the positive numbers and because it can take a wide variety of shapes, including asymmetrical ones. Below we model mortality as constant or changing linearly with time. The expected rate of abundance change is then given by the proportion of total escapement arrived at time *t* minus the expected deaths at time *t*.

#### Constant mortality

With constant per-capita mortality ( $\theta$ ), the rate of abundance change is given by

(1) 
$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = E \frac{\lambda^r t^{r-1} \mathrm{e}^{-\lambda t}}{\Gamma(r)} - \Theta x(t), \quad t \ge 0$$

where E is the total escapement,  $\Gamma(r)$  is the normalizing gamma constant

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy$$

and  $\lambda$  and *r* are positive constants to be estimated from the time-series counts. The solution to eq. 1 for *E* is

$$E = \Theta \int_0^\infty x(t) dt$$

which is the total number of fish that died and is equivalent to the AUC solution for the total escapement. By noting that eq. 1 is a first-order linear differential equation, it can be analytically solved for x(t) (see, e.g., Boyce and DiPrima 2004). The solution for x(t), with initial condition x(0) = 0, yields the expected number-present curve:

(2) 
$$x(t) = Ea \int_0^{t(\lambda-\theta)} \frac{u^{r-1}e^{-u}}{\Gamma(r)} du, \quad t \ge 0$$

where  $a = [(\lambda/(\lambda - \theta)]^r e^{-\theta t}$  for  $\lambda > \theta$ . Under this model, abundance rises to a peak at  $t = (r - 1)/(\lambda - \theta)$ , then declines smoothly toward zero as *t* becomes large. For any time *t*, the expected number of live fish present and susceptible to sampling depends on four parameters: the total escapement (*E*), the mortality rate ( $\theta$ ), and the two arrival parameters of the gamma density function (*r* and  $\lambda$ ).

#### Nonconstant mortality

To accommodate potentially increasing (or decreasing) mortality with time, we assume linear change and substitute  $\theta$  in eq. 1 with  $\theta(t) = \theta_0 + \theta_1 t$ , where  $\theta_0 > 0$  is the initial mortality rate and  $\theta_1$  is the coefficient of mortality acceleration. The linear model, while potentially restrictive for the overall pattern of mortality, may be considered a first-order approximation to a more complicated rate function (e.g., Quinn and Gates 1997; Su et al. 2001) that otherwise might be too difficult to estimate precisely from a short time series of counts. The solution for x(t), with x(0) = 0, yields

(3) 
$$x(t) = E e^{-b} \int_0^t \frac{\lambda^r u^{r-1} e^{c-\lambda u}}{\Gamma(r)} du, \quad t \ge 0$$

where

$$b = \int_0^t \Theta(\mathbf{v}) d\mathbf{v} = \Theta_0 t + 0.5 \Theta_1 t^2$$

and

$$c = \int_0^u \theta(v) dv = \theta_0 u + 0.5 \theta_1 u^2, \quad u \ge 0$$

Note that for  $\theta_1 = 0$ , the mortality rate is constant and that this solution returns eq. 2 as a special case. Model selection for mortality dynamics may therefore proceed in the form of a statistical test on the value of the acceleration coefficient.

#### Admitting unpredictable variability

Random variation can be introduced to the model in one of three ways: (i) as process variability in arrival and death dynamics (Quinn and Gates 1997); (ii) as sampling variability due to counting errors and differences in detection rates during multiple visual surveys (Quinn and Gates 1997; Hilborn et al. 1999; Su et al. 2001); or (iii) as both process and sampling variability (Parken et al. 2003). Immigration and death are likely to vary as a consequence of biology and environmental conditions. However, this variation often is difficult to infer with intermittent observations because most irregular features of immigration and death are effectively removed (or smoothed) from the data as a result of counting fish that arrived or died over multiple surveys. For simplicity, we shall assume that process variability has a negligible effect on the data and that most of the observable variation in our counts arises from sampling.

#### A likelihood model

Our inferential model therefore takes the form of a sampling model for a series of observations from the realized number-present curve. During each of q survey occasions, the entire stream is sampled and  $n_u$  ( $u = t_0, t_1, ..., t_{q-1}$ ) individual fish are counted. We assume that each individual that is alive and present in the stream on occasion u is detected with frequency  $\pi_u$ , where the detection frequencies themselves are independently and identically distributed (iid) as beta random variables. The beta density function for detection is

(4) 
$$f(\pi_u) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi_u^{\alpha - 1} (1 - \pi_u)^{\beta - 1},$$

 $0 \le \pi_u \le 1$  and  $\alpha, \beta > 0$ 

with mean  $E(\pi_u) = \alpha/(\alpha + \beta)$ . Thus, the unconditional distribution of a count from the number-present curve  $(x_u \equiv x(u))$  is

(5) 
$$g(n_u) = \int_0^1 p(N_u = n_u | \pi_u) f(\pi_u) d\pi_u$$
$$= \begin{pmatrix} x_u \\ n_u \end{pmatrix} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + n_u)\Gamma(\beta + x_u - n_u)}{\Gamma(\alpha + \beta + x_u)}$$

with mean  $x_u \alpha / (\alpha + \beta)$ . Equation 5 can be recognized as a beta-binomial distribution (Carlin and Louis 2000), because

$$p(N_u = n_u | \pi_u) = \begin{pmatrix} x_u \\ n_u \end{pmatrix} \pi_u^{n_u} (1 - \pi_u)^{x_u - n_u}$$

20

assumptions to this point are that the distribution of detection frequencies is homogeneous with respect to the individual fish and that the distribution of counts follows a conditional binomial distribution.

The necessary data for statistical inference include the time-series counts  $n_{t_0}$ ,  $n_{t_1}$ ,..., $n_{t_{q-1}}$ , and the time series of detection frequencies, which we shall treat separately below as being either fully or partially known. From the data, we wish to estimate the  $x_1$ ,  $x_2$ ,..., or more precisely, by substituting eq. 2 or eq. 3 into to eq. 5, we wish to estimate the model parameters  $\{E, r, \lambda, \theta\}$  in the case of constant mortality (eq. 2) or  $\{E, r, \lambda, \theta_0, \theta_1\}$  in the case of nonconstant mortality (eq. 3).

If the full time series of detection frequencies is known (i.e.,  $\{\pi_{t_0} = p_{t_0}\}, \{\pi_{t_1} = p_{t_1}\}, ..., \{\pi_{t_{q-1}} = p_{t_{q-1}}\}$ ), the model likelihood  $L_1$  is most easily expressed for parameter estimation as a binomial model for  $x_u$ :

(6) 
$$L_1 = \left[ \prod_{u=0}^{q-1} \binom{x_u}{n_u} p_u^{n_u} (1-p_u)^{x_u-n_u} \right]$$

where u (u = 0, 1, ..., q - 1) denotes survey occasions. Estimation of the unknown parameters then should proceed by replacing  $x_u$  with eq. 2 or eq. 3, with the necessary constraint that the detection rate on the first survey be some value greater than zero, and numerically maximizing the binomial component of the likelihood directly at the time-series counts.

We may accommodate some missing values for detection by assuming that the distribution of detection frequencies is identically distributed with respect to the separate survey occasions. Mean detection undoubtedly will vary over time in response to variable stream conditions, but the assumption of a stationary distribution for detection seems reasonable if sampling is scheduled to coincide with consistently favorable counting conditions, such as during low stream flow and during peak daylight. Under these assumptions the model likelihood then can be written

(7) 
$$L_{2} = \left[\prod_{u=0}^{q-1} \binom{x_{u}}{n_{u}} p^{n_{u}} (1-p)^{x_{u}-n_{u}}\right] \times \left[\prod_{u=0}^{q-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_{u}^{\alpha-1} (1-p_{u})^{\beta-1}\right]$$

The advantage of this formulation for the likelihood is that the overall mean detection now can be estimated by the average of any two or more individual estimates of detection rate. Also, extra uncertainty is included in the model in the form of beta variability because full detection frequencies are unknown and must be estimated by the mean detection rate over time.

#### An approximate likelihood

The model likelihood can be recast in its most simple, albeit restrictive, form to estimate abundance and mortality if only a single-point estimate of detection is available or to estimate only mortality if detection estimates are entirely lacking. By assuming that detection rates are constant (i.e.,  $p_u = p$  for all u), the binomial distribution tends, for large samples, to a Poisson distribution with mean and variance equal

to  $\mu_u = px_u$ . This leads to an approximate likelihood (e.g., Zonneveld 1991):

(8) 
$$L_3 \simeq \prod_{u=0}^{q-1} \frac{(px_u)^{n_u} e^{-px_u}}{n_u!} = \prod_{u=0}^{q-1} \frac{\mu_u^{n_u} e^{-\mu_u}}{n_u!}$$

It should be noted that the Poisson distribution restricts the variance, which may mean that the widths of the confidence intervals (CIs) are somewhat underestimated (i.e., too narrow).

When detection rates are entirely unknown, abundance cannot be estimated by ML because p and E are completely confounded in the estimating equations. (This limitation is common to other ML methods as well, e.g., Hilborn et al. (1999) or Su et al. (2001).) However, the solutions to the ML equations  $\partial L_3/\partial \theta = 0$  or  $\partial L_3/(\partial \theta_0 \partial \theta_1) = (0,0)$  are not affected by the confounded parameters p and E, so an approximate ML estimate for mortality can be obtained despite having no information on the true level of escapement or the detection rate.

# A simulation study to examine bias of the approximate likelihood

We performed Monte Carlo simulations to examine bias of mortality and abundance estimation under our approximate likelihood model. Using eqs. 3 and 7, we simulated 200 time series of eight counts (the number from our tagand-survey study; see Demonstration of the approach) under each of 60 parameter combinations for mean detection, total escapement, and mortality acceleration  $\theta_1$ . We assumed that beta variability in detection under eq. 3 incorporated observation error into the simulated data. To apply the results to small populations, we restricted the range of total escapement for simulations to between 350 and 1400. Values of the remaining parameters were selected based on our analysis of data from the tag-and-survey study and are presented with the results in Figs. 1 and 2. To each simulated time series, we fitted eq. 3 by iterative maximization of the natural logarithm of the Poisson likelihood (eq. 8), thus obtaining a total of 200 × 60 individual estimates for  $\{E, r, \lambda, \theta_0, \theta_1\}$ . From each set of 200, we estimated the median mortality-rate estimate,  $M[\hat{\theta}(t)]$  and the asymptotic relative bias of the escapement estimate, ARB =  $(M[\tilde{E}] - E)/E$ .

#### Demonstration of the approach

#### Collection of empirical data

We collected empirical data on spawning coho salmon in the South Fork Noyo River, a small coastal river in northern California, to evaluate the performance of the models against empirical estimates of mortality and escapement. We conducted traditional spawner surveys to obtain counts of live adults for fitting the models and used counts of adult coho at a weir and mark–recapture surveys to estimate total escapement, longevity, and detection rates for comparison. We checked the weir daily during the entire coho spawning season, December 2003 to February 2004. Coho were measured, identified to sex, marked with an operculum punch and a uniquely numbered Floy tag (T-bar anchor tags; Floy Tag, Seattle, Wash.), and released upstream. We marked three groups of fish (100–240 each) with different-colored

**Fig. 1.** Median mortality-rate estimates,  $M[\hat{\theta}(t)]$ , from simulated time-series counts as a function of mean detection and total escapement  $(\pi, E)$ . The solid line in each plot indicates the basis mortality rate for simulations:  $\theta_0 + \theta_1(t)$ , with  $\theta_0 = 1/34 = 0.029$  and (a)  $\theta_1 = 0$  or (b)  $\theta_1 = 0.003$ . For clarity, some results are not shown.



tags on separate dates 1-2 weeks apart in December to estimate detection rates during subsequent spawner surveys. We counted live spawners and recovered tagged carcasses during approximately weekly surveys covering the entire habitat upstream of the weir, roughly 35 km of stream. We conducted surveys when stream flow and visibility were favorable until no new spawners or carcasses were found, for a total of eight surveys. For estimating longevity, we took the date of death as the midpoint between the survey when a carcass was found and the prior survey (to reduce overestimating stream life because of the intervals between surveys) and calculated longevity as the number of days between tagging of a fish at the weir and date of death. We estimated detection rate for each color-tagged group of fish as the proportion of tagged fish that was observed during the survey on the week after the group was tagged. We estimated escapement by adjusting the weir count using a Petersen estimate to account for unmarked carcasses that were found above the weir. To resolve variation in the longevity data, we fitted a multiple-regression model with longevity taken as the dependent variable and sex, fork length, arrival day, and sex-by-day interaction taken as independent variables.

#### Model fitting and testing

We fitted models for constant and nonconstant mortality to the time-series counts from spawner surveys using, for demonstration, both  $L_2$  (beta-binomial likelihood) and  $L_3$ (Poisson likelihood) to compare parameter estimates under the different likelihooods. Each model was fitted by iteratively maximizing the natural logarithm of the particular likelihood using the simplex search algorithm (Lagarias et al. 1998) in MATLAB (The MathWorks Inc. 2004). Integrals for calculating abundance were evaluated numerically at each iteration of the search procedure by adaptive Gauss– Lobatto quadrature (Gander and Gautschi 2000). For detection rate, we used 0.22 (mean from the tagging estimates) and also 0.50 and 0.70 for comparison because the empirical values probably underestimated actual detection (see Results below). We made pairwise comparisons between mortality models (constant vs. nonconstant) using  $G^2$  tests (generalized likelihood-ratio tests; Rice 1995) to compare quality of fits to the data.

Precision, bias, and correlations of the parameter estimates were estimated by parametric bootstrap (Bradley and Tibshirani 1993). For the data, we simulated 2000 time series of eight counts under eqs. 3 and 7, with parameters substituted by their ML estimates and selecting a detection rate that yielded an escapement estimate closest to the weir count for ease of comparison. Sampling distributions were generated by refitting the model to each simulated time series. Precision was estimated by the percentiles of the empirical sampling distribution and by the coefficient of variation (CV) = SD/median. In this case, CIs may be underestimated somewhat because of our choice of a fixed detection rate and our use of a Poisson model that does not accommodate overdispersion. Bias was estimated by ARB = (median – estimate)/estimate.

# Results

#### Simulations using the approximate likelihood

Median mortality-rate estimates from the Monte Carlo simulations for constant mortality ( $\theta_1 = 0$ ) showed an inverse relationship between estimates of  $\theta_0$  and  $\theta_1$ , indicating negative bias in  $\hat{\theta}_0$  with positive bias in  $\hat{\theta}_1$  (Fig. 1*a*). As a consequence of this relationship, estimates of mortality acceleration, which should equal zero, were instead positive, implying an increasing rate of mortality through time and resulting in a change from negative bias. The degree of bias appeared to diminish over the higher levels of detection and total escapement, and for  $\pi \approx 0.80$  and  $E \ge 1050$ , the mortality estimate did not differ much from its basis value of  $\theta_0 = 0.029$  (determined from the tag-and-survey study).

**Fig. 2.** Asymptotic relative bias (ARB) for estimates of total escapement from simulated time-series counts as a function of parameters  $\pi$  and  $\theta_1$ . Each solid circle represents 200 estimates from data simulated under a binomial probability distribution with values r = 2.68,  $\lambda = 0.16$ , and  $\theta_0 = 0.029$  and (*a*) E = 350 or (*b*) E = 1400. Contours were fitted using a linear-interpolation algorithm in MATLAB (The MathWorks 2004).



For nonconstant mortality ( $\theta_1 > 0$ ), estimates of  $\theta_0$  and  $\theta_1$ also were inversely related, but in contrast to the previous case, the degree of bias generally was insensitive to variability in mean detection and escapement (Fig. 1*b*). Although we set  $\theta_0 = 0.029$  for all simulations, estimates of  $\theta_0$  never deviated much from zero, suggesting, at least for these data, that the initial mortality was difficult to estimate. Estimates of  $\theta_1$  were higher than the basis value in all cases.

Despite the bias observed in the mortality-rate estimates (Fig. 1), escapement estimation appeared to be little affected by Poisson approximation (Fig. 2). This is because the total escapement can be considered a weighted sum of the mortality and the number present,  $E = \int_{0}^{\infty} \theta(t) x(t) dt$ , and therefore bias in the escapement estimates is incurred largely as a function of the estimated net mortality,  $\int_{0}^{\infty} \hat{\theta}(t) dt$ , rather than the estimated instantaneous mortality,  $\hat{\theta}(t)$ . For instance, where bias in the estimates of instantaneous mortality was negative and then positive and large (over low  $\theta_1$  and low  $\pi$ ; Fig. 2a), the net mortality was overestimated, and as a result, the escapement was estimated high (lower diagonals of Figs. 2a, 2b). Conversely, where bias in the estimates of instantaneous mortality were negative and small (over high  $\theta_1$ ; Fig. 2b), net mortality was underestimated, and as a result, escapement was estimated low (upper diagonals of Fig. 2). Also, because the bias was little affected by detection and escapement for  $\theta_1 > 0$  (Fig. 2*b*), the bias in the escapement estimates was generally unchanged over the basis values of  $\pi$ and E. For  $\theta_1 > 0$ , the amount of bias often was small (mostly between -20% and 20%; Fig. 2b).

#### Demonstration of the approach

#### **Empirical results**

We captured and marked 645 adult coho at the South Fork Noyo River weir (Fig. 3) and recovered 162 carcasses during spawner surveys. Fourteen of the carcasses were unmarked, resulting in a Peterson estimate of total escapement of 706 fish. Estimates of detection rate from the three groups of uniquely tagged fish were 0.24, 0.20, and 0.22, suggesting that detection frequencies were stationary over time and that eq. 7 was an appropriate likelihood for parameter estimation. However, the mean detection rate ( $\hat{p} = 0.22$ ) probably underestimated true detection as a result of unobserved mortality and tag loss (e.g., 35% of carcasses had operculum punches but were missing Floy tags). Multiple regression analysis revealed that longevity was related to arrival day (type III *F* tests, with  $\alpha = 0.05$ , b = -0.51, F = 31.09, P < 0.001,  $R^2 =$ 0.25, n = 92) but unrelated to sex or size, with no interactions. The final fitted regression for the mean longevity shows a steep decline from approximately 30 days at the start of the immigration period to approximately 10 days on the 40th day of the season (Fig. 4).

#### Performance and comparison of models

Estimates of per-capita mortality and total escapement, as well as the negative of the maximized log-likelihood  $(-\ln L_{max})$ for model selection and evaluation of model goodness-of-fit, are shown in Table 1. The model for nonconstant mortality  $(\theta_1 \neq 0)$  provided a significantly better fit to the counts than the model for constant mortality, regardless of the level of detection or whether the test statistic was taken as a ratio of beta-binomial or Poisson likelihoods (Table 1). However, comparison of fits between the model for nonconstant mortality and a general model, in which the  $x_t$  were taken directly as the counts themselves rather than being estimated by eq. 3, yielded  $G^2$  values of 30.0 for the Poisson and 28.1, 42.3, and 67.9 for the beta-binomial with p = 0.22, 0.5, and0.7, respectively. Considering that the critical value for the test is  $G_{\alpha,\nu}^2 = 6.25$  (from a chi-square distribution with  $\alpha =$ 0.1 and v = 3 df), the large values indicated a significant model lack-of-fit (when compared with the most general model) and that the data were better described with more pa**Fig. 3.** Daily counts of coho salmon at the South Fork Noyo River weir. Day 0 marks the time of the first upstream spawner survey.



rameters. Nonetheless, the nonconstant mortality model shows a reasonably good fit to the counts (Fig. 5), which is heartening in light of the irregular pattern of arrival that we observed at the weir (Fig. 3).

Most differences between the escapement estimates were due to differences in average detection and whether mortality was taken as constant or nonconstant rather than due to differences in the assumed likelihood distribution (Table 1). Estimates of escapement between likelihoods were virtually identical for models of constant mortality but were lower under the Poisson distribution, because of negative bias in estimates of net mortality (see Simulations using the approximate likelihood) for models of nonconstant mortality. The escapement estimate closest to the weir count was obtained by the nonconstant mortality model with p = 0.50, suggesting, if the model is correct, that our three mark-resighting surveys underestimated mean detection of tagged fish (as noted above) and that the average detection may indeed have been closer to 0.50. Escapement estimates by the constant mortality model were always about twice that estimated by the nonconstant mortality model and were always much higher (by a factor of 1.67-5.33) than the weir count.

Initial parameter estimates and the bootstrap results for estimating precision, bias, and correlations of the parameter estimates are shown in Table 2. On eight counts, estimates were generally accurate but imprecise, with ARB  $\leq |0.05|$ and CV  $\leq 0.5$ . Estimates of intrinsic mortality ( $\theta_0$ ) were highly imprecise, but the estimates were close enough to zero to have little practical importance. Note that the bootstrapped 90% CI for  $\hat{\theta}_1$  does not include zero, suggesting that our decision to include  $\theta_1$  in the model was correct. The 50th percentile of our model-based longevity estimate (the inverse of the mortality-rate estimate) was 7.3 days (90% CI: 4.1,10.1) at t = 40, which is similar to the empirical estimate based on tagging (Fig. 4). Correlations between  $\hat{r}$  and  $\hat{\lambda}$  and between  $\hat{\lambda}$  and  $\hat{\theta}_1$  were high, indicating that several combinations of parameter values could describe the arrival and

**Fig. 4.** Longevity of adult coho salmon with respect to arrival day at the South Fork Noyo River weir (linear regression, y = 30.03 - 0.51x, P < 0.001,  $R^2 = 0.25$ ). Longevity was estimated from the recovery of tagged carcasses during spawner surveys.



mortality pattern in the data equally well. The high correlation between  $\hat{E}$  and  $\hat{\theta}_1$  reflect the strong dependence of the escapement estimate on the estimate of mortality rate.

# Discussion

The AUC approach for abundance estimation lends naturally to statistical modeling because the escapement process can be partitioned straightforwardly into manageable components of arrival and death. Likelihood methods admit uncertainty in the underlying components because of process or sampling variability, as well as provide a general framework for statistical inference. In this paper, we relied on the setting of differential equations, put forth by Quinn and Gates (1997), to model the escapement process through time. Autocorrelation in arrival was managed by modeling immigration by parametric density function (Hilborn et al. 1999), and mortality was modeled, on a per-capita basis, as constant (Manly 1974; Zonneveld 1991) or changing linearly over time. The solution to the differential equation yielded our model for the expected number present, which we incorporated as the mean of a beta-binomial or Poisson probability model to allow for variability from sampling. Conditional on the model, we extracted information on deaths from the time-series counts by explicitly accounting for the effect of mortality on the location of the peak and on the rate of change in the fitted number-present curve. Our results from applying the model to count data on coho salmon from South Fork Noyo River gave accurate estimates of mortality compared with empirical estimates based on tagging and produced reasonable estimates of abundance within the uncertainty surrounding the detection rate.

The major improvement of our approach over previous AUC and likelihood methods is the ability to estimate mortality directly from time-series counts of spawners, eliminating the need to collect additional data on mortality rates.

Model	р	–ln L <sub>max</sub>	$\hat{E}$	$\hat{\theta}_0$	$\hat{\theta}_1$	$G^2$	P value <sup>a</sup>
Beta-binomial							
θ	0.22	38.19	3764	0.189			
θ	0.50	46.59	1655	0.186			
θ	0.70	61.82	1182	0.182			
$\theta(t)$	0.22	32.94	1710	$1.73 \times 10^{-8}$	$3.52 \times 10^{-3}$	10.50	$1.20 \times 10^{-3}$
$\Theta(t)^c$	0.50	38.72	727	$1.84 \times 10^{-9}$	$3.37 \times 10^{-3}$	15.73	7.31×10 <sup>-5</sup>
$\theta(t)$	0.70	49.76	522	b	$3.32 \times 10^{-3}$	24.10	9.13×10 <sup>-7</sup>
Poisson							
θ	0.22	34.98	3765	0.191			
θ	0.50	34.98	1656	0.191			
θ	0.70	34.98	1183	0.191			
$\theta(t)$	0.22	30.81	1593	b	$3.31 \times 10^{-3}$	8.35	$3.86 \times 10^{-3}$
$\theta(t)$	0.50	30.81	701	b	$3.31 \times 10^{-3}$	8.35	$3.86 \times 10^{-3}$
$\theta(t)$	0.70	30.81	501	b	$3.31 \times 10^{-3}$	8.35	$3.86 \times 10^{-3}$

**Table 1.** Estimates of parameters for models of constant mortality  $\theta$  and nonconstant mortality  $\theta(t)$  fit to counts of adult coho from the South Fork Noyo River in 2003–2004.

Note: Parameters were estimated using a beta-binomial or Poisson likelihood and under different lev-

els of mean detection p. Escapement (E), estimated at the weir, was 706.

<sup>a</sup>Probability value for  $G^2 = 2\ln L_{\max,\theta} - 2\ln L_{\max,\theta(t)}$  from a chi-square distribution with v = 1 df.  ${}^b\hat{\theta}_0 < 1 \times 10^{-9}$ .

**Fig. 5.** Estimates of coho salmon abundance (lines; scaled by detection rate  $\hat{p} = 0.22$ ) and the number of live fish counted during periodic walking surveys of the South Fork Noyo River (circles). Error bars for counts are  $\pm 2$  SE, estimated from a beta-binomial distribution. Day 0 marks the time of the first survey. The model-based abundance estimates account for either a constant (dotted line) or increasing (solid line) rate of in-stream mortality.



This should reduce the time and cost where tagging (e.g., English et al. 1992) or other techniques (e.g., video; Shardlow 2004) otherwise would be used to estimate mortality and should improve the accuracy of estimates when estimating mortality is not practical and average values from the literature would be used instead. Furthermore, our method allows modeling of decreasing longevity over the spawning season, which improves the accuracy of escapement estimates compared with using constant values (our study; Su et al. 2001). The nonconstant mortality (i.e., decreasing lon-

gevity) model better fit the data than the constant mortality model under both likelihoods (beta-binomial and Poisson) and provided an abundance estimate much lower and closer to the weir count, and thus the form of the mortality model had a much greater effect on the abundance estimate than the form of the likelihood. The mortality estimates also were largely insensitive to the detection rate, especially under the nonconstant model.

In our paper, we worked from general forms of the models to simpler forms by imposing constraints based on logical assumptions, and the various forms of the models allow flexibility in applying the method under different data scenarios. As mentioned earlier, the nonconstant mortality model generally appears to be more suitable than the constant mortality form, although selection of the mortality form can be tested for individual data sets. The different forms of the likelihood cater to different levels of information on detection rate. If detection rates are estimated for every count in the time series (such as by using distance sampling; see below),  $L_1$  (eq. 6) can be used.  $L_2$  (eq. 7) is useful if some detection estimates are missing from the time series but enough data is available to calculate the mean or if enough prior information is available (or alternatively might be taken from the literature) to specify parameters for the beta distribution for detection. The approximate likelihood  $L_3$ (eq. 8) is useful for estimating abundance and mortality if only a single-point estimate of detection is available or for estimating just mortality in the absence of any detection rate information. The Monte Carlo simulations and the similarity of results between fitting the counts in the demonstration using the beta-binomial versus the Poisson likelihood suggest that the approximate (Poisson) likelihood offers a relatively unbiased and simple model that may be used in these cases.

Although our model offers several refinements over past methods, it shares two limitations common to all AUC approaches. First, our model requires the initial-condition constraint, x(0) = 0. This constraint was necessary to obtain the

(a) Initial pa	arameter estima	ites.				
		Percentile				
Parameter	Estimate	5	50	95	CV	ARB
Ε	727.3	533.8	741.6	1379.7	0.488	0.020
r	2.588	2.122	2.708	4.099	0.252	0.047
λ	0.149	0.105	0.148	0.291	0.444	-0.010
$\theta_0$	$1.84 \times 10^{-9}$	a	a	a	1.077	-0.050
$\theta_1$	$3.37 \times 10^{-3}$	$2.48 \times 10^{-3}$	$3.42 \times 10^{-3}$	$6.14 \times 10^{-3}$	0.465	0.014
(b) Correlati	on coefficients					
	$\hat{E}$	ŕ	λ	$\hat{\theta}_0$		
ŕ	-0.04					
λ	-0.39	0.91				
$\hat{\boldsymbol{\Theta}}_0$	0.03	0.03	0.04			
$\hat{\theta}_1$	0.99	-0.06	-0.40	0.03		

**Table 2.** Initial parameter estimates and bootstrap results from simulations of the South Fork Noyo River coho counts using the nonconstant mortality model (eqs. 3 and 7), with p = 0.5.

**Note:** A time-series of eight counts was simulated 2000 times. ARB, asymptotic relative bias.  ${}^{a}\theta_{0} < 1 \times 10^{-9}$ .

solution (eq. 2 or eq. 3) for the expected rate of abundance change (eq. 1) and follows because the gamma density at t =0 is equal to zero. The analogous constraint in AUC estimation is that the first and last counts must equal zero (English et al. 1992), but with the parametric approach only the first count must be so constrained. If fish are present on the first survey, the day of first arrival to the sample area must be guessed or estimated and the time series incremented accordingly. Otherwise, the first count in the series will be disregarded in estimation (i.e., assumed to be zero), thus leading to an error in the mortality and escapement estimates. The use of historical data may help to specify a Bayesian prior distribution on arrival time in the spirit of Su et al. (2001), but the easiest solution is to schedule surveys to occur either before or soon after the start of arrival. Unfortunately, such precise scheduling may not always be realistic.

The second limitation is the requirement for information on detection rate to estimate escapement. Because visibility rates for surveys often are poorly resolved, estimates of salmon abundance may be particularly susceptible to this form of uncertainty. The standard method of detection estimation is to take a two-stage sampling approach and "correct" visual counts by correlation with a set of independent abundance estimates obtained using more rigorous methods of enumeration (e.g., photography (Neilson and Geen 1981); electrofishing or snorkeling (Irvine et al. 1992); walking versus arial counts (Hilborn et al. 1999); mark and recapture (Lady and Skalski 1998; Manske and Schwarz 2000; Korman et al. 2002)). We took a similar approach, using a series of tag and resight surveys to collect data on detection. However, this approach was shown to be unreliable in practice because constraints on surveyor effort prevented more than three attempts at data collection, some level of tag loss occurred, and sampling conditions delayed the resighting surveys, resulting, presumably, in underestimates of the mean detection rate. Therefore, despite expending considerable effort and using a weir for tagging, we could generate only a small number of data to infer detection, which in addition to compromising accurate estimation of the mean also made it impossible for us to rigorously evaluate assumptions of detection homogeneity and stationarity.

In conclusion, although our method reduces one of the data limitations for estimating salmon spawner abundance from standard stream counts by estimating longevity directly from the counts, a robust method of determining detection still is sorely needed to improve the accuracy of escapement estimates. One method that deserves serious consideration is distance sampling, in particular, line-transect sampling (Buckland et al. 1993). Line-transect sampling has found wide use in visual surveys of marine mammals (Laake et al. 1997; Gelatt and Siniff 1999; Pollard et al. 2002) and reef fishes (Watson and Quinn 1997; Kulbicki 1998; Samoilys and Carlos 2000) but has not been used much in streams (however, see Ensign et al. 1995). The theoretical basis for the method is that the probability of detection within a surveyed area declines from unity as a predictable function of distance between the object of interest and the observer. A large statistical literature exists for inferring detection functions from counts, data on distances, and any potentially important covariates (e.g., stream clarity, depth, habitat variables). We believe that by incorporating such methods, for which our models are easily modified, the model-based approaches would be largely improved, thus yielding a more mature methodology of count-based escapement estimation.

### Acknowledgements

Masami Fujiwara provided extremely helpful advice and suggestions on the models and manuscript and Scott Harris and Morgan Knechtle of the California Department of Fish and Game provided critical support for the demonstration study. Michael Mohr, Eric Bjorkstedt, Brian Spence, and two anonymous reviewers provided helpful comments that improved the manuscript.

# References

- Adkison, M.D., and Su, Z.M. 2001. A comparison of salmon escapement estimates using a hierarchical bayesian approach versus separate maximum likelihood estimation of each year's return. Can. J. Fish. Aquat. Sci. 58: 1663–1671.
- Boyce, W.E.E., and DiPrima, R.C. 2004. Elementary differential equations. 8th ed. John Wiley & Sons, New York.
- Bradley, E., and Tibshirani, R. 1993. An introduction to the bootstrap. Monographs on statistics and applied probability 57. Chapman & Hall, New York.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., and Laake, J. 1993. Distance sampling: estimating abundance of biological populations. 1st ed. Chapman & Hall, London, UK.
- Carlin, B.P., and Louis, T.A. 2000. Bayes and empirical Bayes methods for data analysis. 2nd ed. Chapman & Hall, New York.
- English, K.K., Bocking, R.C., and Irvine, J.R. 1992. A robust procedure for estimating salmon escapement based on the area-underthe-curve method. Can. J. Fish. Aquat. Sci. 49: 1982–1989.
- Ensign, W.E., Angermeir, P.L., and Dolloff, C.A. 1995. Use of line transect methods to estimate abundance of benthic stream fishes. Can. J. Fish. Aquat. Sci. 52: 213–222.
- Gander, W., and Gautschi, W. 2000. Adaptive quadrature revisited. BIT, **40**: 84–101.
- Gelatt, T., and Sini, D. 1999. Line transect survey of crabeater seals in the Amundsen–Bellingshausen seas, 1994. Wildl. Soc. Bull. 27: 330–336.
- Hilborn, R., Bue, B.G., and Sharr, S. 1999. Estimating spawning escapements from periodic counts: a comparison of methods. Can. J. Fish. Aquat. Sci. 56: 888–896.
- Irvine, J.R., Bocking, R.C., English, K.K., and Labelle, M. 1992. Estimating coho salmon (*Oncorhynchus kisutch*) spawning escapements by conducting visual surveys in areas selected using stratified random and stratified index sampling designs. Can. J. Fish. Aquat. Sci. **49**: 1972–1981.
- Korman, J., Ahrens, R.N.M., Higgins, P.S., and Walters, C.J. 2002. Effects of observer efficiency, arrival timing, and survey life on estimates of escapement for steelhead trout (*Oncorhynchus mykiss*) derived from repeat mark–recapture experiments. Can. J. Fish. Aquat. Sci. 59: 1116–1131.
- Kulbicki, M. 1998. How the acquired behaviour of commercial reef fishes may influence the results obtained from visual censuses. J. Exp. Mar. Biol. Ecol. 222: 11–30.

- Laake, J., Calambokidis, J., Osmek, S., and Rugh, D. 1997. Probability of detecting harbor porpoise from aerial surveys: estimating g(0). J. Wildl. Manag. 61: 63–75.
- Lady, J.M., and Skalski, J.R. 1998. Estimators of stream residence time of Pacific salmon (*Oncorhynchus* spp.) based on release– recapture data. Can. J. Fish. Aquat. Sci. 55: 2580–2587.
- Lagarias, J., Reeds, J.A., Wright, M.H., and Wright, P.E. 1998. Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM J. Optimization, 9: 112–147.
- Manly, B.F.J. 1974. Estimation of stage-specific survival rates and other parameters for insect populations developing through several stages. Oecologia, 15: 277–285.
- Manske, M., and Schwarz, C.J. 2000. Estimates of stream residence time and escapement based on capture–recapture data. Can. J. Fish. Aquat. Sci. 57: 241–246.
- Neilson, J.D., and Geen, G.H. 1981. Enumeration of spawning salmon from spawner residence time and aerial counts. Trans. Am. Fish. Soc. 110: 554–556.
- Parken, C.K., Bailey, R.E., and Irvine, J.R. 2003. Incorporating uncertainty into area-under-the-curve and peak count salmon escapement estimation. N. Am. J. Fish. Manag. 23: 78–90.
- Pollard, J., Palka, D., and Buckland, S. 2002. Adaptive line transect sampling. Biometrics, 58: 862–870.
- Quinn, T.J., and Gates, R. 1997. Estimation of salmon escapement: models with entry, mortality and stochasticity. Nat. Resour. Model. 10: 217–250.
- Rice, J.A. 1995. Mathematical statistics and data analysis. 2nd ed. Wadsworth, Belmont, Calif.
- Samoilys, M., and Carlos, G. 2000. Determining methods of underwater visual census for estimating the abundance of coral reef fishes. Environ. Biol. Fishes, 57: 289–304.
- Shardlow, T. 2004. Using time-lapsed video to estimate survey life for area-under-the-curve methods of escapement estimation. N. Am. J. Fish. Manag. 24: 1413–1420.
- Su, Z.M., Adkison, M.D., and Van Alen, B.W. 2001. A hierarchical bayesian model for estimating historical salmon escapement and escapement timing. Can. J. Fish. Aquat. Sci. 58: 1648–1662.
- The MathWorks Inc. 2004. MATLAB. The MathWorks, Inc., Natick, Massachusetts.
- Watson, R., and Quinn, T. 1997. Performance of transect and point count underwater visual census methods. Ecol. Model. 104: 103–112.
- Zonneveld, C. 1991. Estimating death rates from transect counts. Ecol. Entomol. **16**: 115–121.