Ecology, 83(9), 2002, pp. 2379–2386

VALIDATING POPULATION VIABILITY ANALYSIS FOR CORRUPTED DATA SETS

ELIZABETH E. HOLMES[1,3] AND WILLIAM F. FAGAN[2]

[1]*REUT Division, Cumulative Risk Initiative, Northwest Fisheries Science Center, 2725 Montlake Blvd. E, Seattle, Washington 98112 USA*
[2]*Department of Biology, Arizona State University, Tempe, Arizona 85287-1501 USA*

Abstract. Diffusion approximation (DA) methods provide a powerful tool for population viability analysis (PVA) using simple time series of population counts. These methods have a strong theoretical foundation based on stochastic age-structured models, but their application to data with high sampling error or age-structure cycles has been problematic. Recently, a new method was developed for estimating DA parameters from highly corrupted time series. We conducted an extensive cross-validation of this new method using 189 long-term time series of salmon counts with very high sampling error and nonstable age-structure fluctuations. Parameters were estimated from one segment of a time series, and a subsequent segment was used to evaluate the predictions regarding the risk of crossing population thresholds. We also tested the theoretical distributions of the estimated parameters. The distribution of parameter estimates is an essential aspect of a PVA because it allows one to calculate confidence levels for risk metrics. This study is the first data-based cross-validation of these theoretical distributions. Our cross-validation analyses found that, when parameterization methods designed for corrupted data sets are used, DA predictions are very robust even for problematic data. Estimates of the probability of crossing population thresholds were unbiased, and the estimated parameters closely followed the expected theoretical distributions.

Key words: Dennis method; Dennis-Holmes method; diffusion approximation; extinction; model validation; population viability analysis; salmon; sampling error.

INTRODUCTION

Population viability analysis (PVA) has become a standard tool in conservation biology (Boyce 1992). Conservation organizations such as The Nature Conservancy use it to rank the quality of sites, the IUCN (International Union for the Conservation of Nature) uses it to establish the degree of risk faced by species, and federal agencies use it to assist management decisions regarding threatened and endangered species. In spite of its widespread use, there is vigorous debate in the academic literature regarding the merit of PVA models. Opinions range from the argument that PVA is a poor idea because confidence intervals surrounding risk metrics are too large (Fieberg and Ellner 2000) and sampling error makes parameterization error-prone (Ludwig 1999), to the belief that PVA can be used to establish relative risk even though absolute estimates are tenuous (Fagan et al. 2001), to the contention that PVA is supported by data and sufficiently accurate for risk assessments (Brook et al. 2000). Missing in this debate have been rigorous validation studies with large and long-term data sets. Brook et al. (2000) presented the first such validation study and examined detailed age-structured PVAs. This type of PVA, however, requires detailed population data; unfortunately, such data are seldom available. Instead, simple population counts are often the only available data for species of conservation concern. Although PVA methods for count data exist, cross-validations of these methods are lacking.

In this paper, we examine diffusion approximation (DA) methods for count-based viability analysis using a data set of 189 time series from western North American salmon, many from populations that are currently listed as endangered or threatened under the U.S. Endangered Species Act. Although DA methods have been used in a variety of conservation settings (Nicholls et al. 1996, Gerber et al. 1999, NMFS 2000), they are known to be sensitive to sampling error and other non-environmental variability in the data. Salmon time series suffer from such problems to an extreme degree. The data are characterized by high observation errors, and the life history of salmon makes them prone to severe age-structure oscillations. Such problems hide the underlying stochastic process. The standard methods for estimating DA parameters are designed for low non-environmental noise (Dennis et al. 1991) and fail in this situation.

A new DA method was recently developed (Holmes 2001) to handle these types of data problems by partitioning the variability of a population time series into ''non-process'' error, such as observation errors or cy-

Manuscript received 11 June 2001; revised 25 September 2001; accepted 13 November 2001; final version received 12 February 2002.

[3] E-mail: eli.holmes@noaa.gov

cles linked to age-structure perturbations, vs. ''process error,'' the environmental variability driving the long-term statistical distributions of population trajectories. Here, we cross-validate the new method using time series of salmon. Our large number of long time series allows us to cross-validate not only the bias in risk metrics (as did Brook et al. 2000), but also the statistical distributions of the estimated parameters. The statistical distributions of parameter estimates are perhaps the most critical aspect of a PVA because they allow one to calculate the uncertainty in one's risk estimates. Point estimates of risk metrics, such as the probability of extinction in x years, are by themselves of limited value, because even a simple comparison of risk between populations is meaningless without knowledge of the statistical distribution of the estimated risk metric. One strength of DA methods is that these distributions can be calculated. However, these calculations require numerous simplifying assumptions. Our study presents the first empirical cross-validation of these calculated distributions and, consequently, the theory underlying DA methods for PVAs.

METHODS

We assembled a data set of 147 chinook salmon and 42 steelhead time series of yearly spawner indices from databases maintained by the U.S. National Marine Fisheries Service and the Pacific States Marine Fisheries Commission (summarized in Appendix A with raw data in Supplement 2). The data are from Evolutionarily Significant Units (ESUs) in Washington, Oregon, and California, USA, and consist of egg-bed counts, dam counts, carcass counts, peak live counts, or total live estimates. Each time series was divided into 20-, 30-, or 40-yr overlapping segments (depending on the analysis), with the segments separated by five years; e.g., a 1960–1999 time series would be divided into the 30-yr segments: 1960–1989, 1965–1994, and 1970–1999. To limit overrepresentation of long time series, we allowed a maximum of 10 randomly chosen segments from each time series. To limit overrepresentation by two ESUs with a disproportionate number of time series, only one segment (randomly chosen) was used from each time series in the Snake River spring/summer chinook ESU, and only three were used from each series in the Oregon Coast chinook ESU. These restrictions applied to all analyses except the σ^2 estimates, which required a larger sample size. We also did a separate comparative analysis focused on a smaller geographic scale, using all time series in the Snake River spring/summmer chinook ESU in the Columbia River basin.

Each segment was divided into a parameterization period followed by an evaluation period. Parameter distributions and risk levels were predicted from the parameterization period, and then the data in the evaluation period were used to test these predictions. We did two basic analyses. First we cross-validated the parameter distributions estimated from the parameterization period; this tests the distributions used to calculate confidence intervals for DA risk metrics. Second, we asked, ''Do diffusion approximations properly estimate the probability of crossing population thresholds?'' This cross-validation addresses whether DAs are a reasonable tool for analyzing the risks of decline evident in the actual salmon population trajectories.

Estimating population viability metrics from corrupted counts

DA methods for viability analysis arose from density-independent, stochastic, age-structured models. Such population processes can be approximated by: $N_{t+1} = N_t\exp(\mu + \varepsilon_p)$, where $\varepsilon_p \sim$Normal$(0, \sigma_p)$ (Tuljapurkar 1989, Dennis et al. 1991). This model is a stochastic process where the annual population growth rate is a lognormally distributed random variable. The median annual growth rate is μ. The stochasticity in the annual growth rate is determined by the process error term, ε_p, which is normally distributed with variance σ_p^2. A diffusion approximation of this process gives the statistical distribution of the ratio of population size at time t vs. the population size at time τ later: $\ln(N_{t+\tau}/N_t)$ is distributed Normal$(\mu\tau, \sigma_p\sqrt{\tau})$. From this distribution, risk metrics such as mean long-term growth rates, probabilities of decline or extinction, and the mean time to extinction can be calculated (Dennis et al. 1991). Dennis et al. discuss methods for estimating μ and σ_p^2 using a time series of counts. These methods work well when the variability due to non-process error (e.g., sampling error or strong age-structure cycles) is low (see the petrel example in Holmes 2001). However, when the data are characterized by high non-process error, as are salmon data (Hilborn et al. 1999), the standard methods result in severe overestimates of σ_p^2, leading to poor estimation of risk metrics (Holmes 2001).

To deal with such problems, an alternative parameterization method was developed (Holmes 2001). We refer to viability analysis using this method as the Dennis-Holmes method, wherein estimation of model parameters follows Holmes (2001) and calculation of the risk metrics from the parameters follows Dennis et al. (1991). This method seeks to estimate μ and σ_p^2 from a time series representing highly corrupted observations, O_t, of the true population size, N_t:

$$N_{t+1} = N_t\exp(\mu + \varepsilon_p) \quad \text{where } \varepsilon_p \sim \text{Normal}(0, \sigma_p)$$
$$O_t = N_t\exp(\varepsilon_{np}) \quad \text{where } \varepsilon_{np} \sim f(\beta, \sigma_{np}). \qquad (1)$$

The parameter ε_{np} represents the level of non-process error that corrupts the observations of the true population size. It has some unknown distribution with mean β and variance σ_{np}^2. This noise makes the underlying environmental variability (σ_{np}^2) impossible to observe directly. The log of Eq. 1 is known as a linear state–space model. Such models are extensively studied in

the engineering literature, and Expectation–Maximization (EM) algorithms using Kalman filters have been developed to estimate the parameters from noisy data (Shumway and Stoffer 1982, Ghahramani and Hinton 1996), but to accurately estimate ε_p, these methods require information about the non-process error, particularly the bias, β. Such information is often not available for ecological data.

The method by Holmes (2001) adopts another approach designed for DA models used for population processes and does not require information about the non-process error. It takes advantage of the contrasting effects of process error (the environmental variability) vs. non-process error (e.g., sampling error) on the variance between $O_{t+\tau}$ and O_t, namely $\text{var}(\ln(O_{t+\tau}/O_t)) = \sigma_p^2\tau + \sigma_{np}^2$. This suggests that the slope of $\text{var}(\ln(O_{t+\tau}/O_t))$ vs. τ could recover the process error term in the face of high corruption. Unfortunately, this regression has problems for short time series because negative slopes (i.e., negative variance estimates) are frequent. The method circumvents this problem by noting that a short sum of L sequential O_t's ($R_t = \Sigma_{i=1}^{L} O_{i+i-1}$) retains the variance vs. τ relationship, but filters out the noise. The σ_p^2 estimate, termed $\hat{\sigma}_{slp}^2$, is the slope of a regression of $\text{var}(\ln(R_{t+\tau}/R_t))$ vs. τ with the intercept free. Simulations indicate that $L = 3$ to 5 is a good compromise between loss of information due to high filtering and errors due to low filtering (see Holmes 2001; also see Appendix B). For all of our analyses, $L = 4$ and max $\tau = 4$.

Numerical simulations indicate that $\hat{\sigma}_{slp}^2$ has approximately a χ^2 distribution:

$$\frac{\text{df}_{slp}\hat{\sigma}_{slp}^2}{\sigma_{slp}^2} \sim \chi^2_{\text{df}_{slp}}. \tag{2}$$

For a time series of length n, $\text{df}_{slp} = 0.333 + 0.212n - 0.387L$ for $n > 15$, gives a good estimate of the degrees of freedom. See Appendix B for a discussion and derivation of the χ^2 distribution and the numerical estimation of the formula for df_{slp}. Note, $\hat{\sigma}_{slp}^2$ is a biased estimator of σ_p^2. Appendix B shows the bias for simple lognormal observation error, and Holmes (2001) shows the biases using stochastic matrix models. In general, the bias will be poorly known, but the cross-validation results indicate that the level is not so severe as to significantly affect the predictions.

Estimation of μ from the corrupted time series does not generally suffer from bias, but does suffer from loss of precision. Use of running sums (the R_t's) helps to reduce this problem: $\hat{\mu}_R$ is the sample mean of $\ln(R_{t+1}/R_t)$. For σ_{np} small (e.g., <1) and L small, the distribution of this estimate is:

$$\hat{\mu}_R \sim \text{Normal}(\mu, \sigma_{\mu,R}) \quad \text{where}$$

$$\sigma_{\mu,R}^2 = \frac{1}{(n-L)^2}\left(\frac{2}{L}\sigma_{np}^2 + (n-L)\sigma_p^2\right). \tag{3}$$

As the time series length, n, increases, the variance of $\hat{\mu}_R$ goes to $\sigma_p^2/(n - L)$. This suggests that we could estimate the distribution of $\hat{\mu}_R$ from the data by using our estimate of σ_p^2, i.e., from $\hat{\sigma}_{slp}^2$:

$$\frac{\hat{\mu}_R - \mu}{\sqrt{\hat{\sigma}_{slp}^2/(n-L)}} \sim \frac{1}{\sqrt{\gamma}} t_{\text{df}_{slp}}$$

where

$$\gamma = \frac{\sigma_{slp}^2}{\left(\frac{2}{L(n-L)}\sigma_{np}^2 + \sigma_p^2\right)}. \tag{4}$$

Although γ is unknown, its range is not large (see Appendix B). For the salmon data sets, the observed mean γ was 0.7–1.2. Note that for corrupted time series, $\text{var}(\hat{\mu}_R) \neq \text{var}(\ln(R_{t+1}/R_t))$!

Derivations for Eqs. 2–4 are in Appendix B. The distributions of the estimated parameters (Eqs. 2–4) are approximate and involve a variety of simplifying assumptions. One main goal of this cross-validation is to test whether these approximate distributions are supported by data. This is critical because these distributions are used to calculate confidence intervals for risk metrics. Supplement 1 has S-PLUS code for estimating $\hat{\mu}_R$, $\hat{\sigma}_{slp}^2$, $\hat{\sigma}_{\mu,R}^2$, and $\hat{\sigma}_{np}^2$ from a time series, and S-PLUS code for estimating risk metrics and confidence intervals.

Cross-validating parameter distributions using time series

Our first cross-validation tested whether the $\hat{\mu}_R$ estimates from the data are consistent with the theoretical distribution of $\hat{\mu}_R$ (Eq. 4). To do this, we derived a t distribution governing the difference between $\hat{\mu}_R$ from the parameterization and evaluation periods ($\hat{\mu}_{R,p} - \hat{\mu}_{R,e}$):

$$\frac{(\hat{\mu}_{R,p} - \hat{\mu}_{R,e})}{\sqrt{\frac{\text{df}_{slp}\hat{\sigma}_{slp,p}^2 + \text{df}_{slp}\hat{\sigma}_{slp,e}^2}{2\text{df}_{slp}}\left(\frac{1}{n_p - L} + \frac{1}{n_e - L}\right)}} \sim \frac{1}{\sqrt{\gamma}} t_{2\text{df}_{slp}}. \tag{5}$$

The t statistic (the right-hand side of Eq. 5) was designed so that it has the same t distribution regardless of μ or σ_p^2 (See Appendix C). In this way, the t statistics from all the segments and time series could be combined and tested for their conformity to a single t distribution (the left-hand side of Eq. 5). It is not possible to simply compare $\hat{\mu}_R$'s to some distribution because each time series represents a different population with a different underlying distribution of annual growth rates driving its stochastic population process (i.e., the μ's and σ_p^2's are different). For this analysis, we used 15-yr parameterization (p) and evaluation (e) periods (to derive the t distribution, the periods must be the same). With $n = 15$, $\text{df}_{slp} \approx 1.96$.

For the second cross-validation, we examined whether the ratios of $\hat{\sigma}_{slp,e}^2$ from the evaluation period to

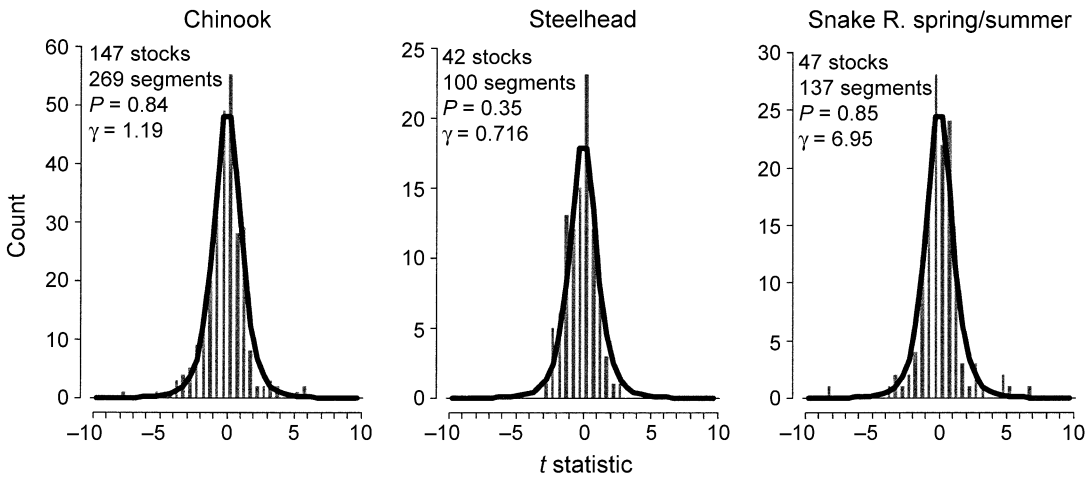

FIG. 1. Histogram of the predicted vs. observed distribution of $\hat{\mu}_R$, the estimate of median annual population growth rate. The solid lines show the expected (theoretical) t distribution governing the difference between $\hat{\mu}_R$ measured in the parameterization vs. evaluation periods. The bars show the observed distribution of t statistics measured from the actual time series. The P values are from a Kolmogorov-Smirnov goodness-of-fit test to a t distribution with 2×1.96 degrees of freedom: $t_{3.92}/\sqrt{\gamma}$, where γ for each species was estimated from the data. For the chinook and steelhead analyses, the set of segments used was variable because of random subsampling to reduce overrepresentation by long time series and by ESUs (Evolutionarily Significant Units) with many time series. The analysis was repeated 100 times with different random samples of segments. For the chinook analysis, the 5% and 95% quantiles for the P values and γ's were (0.50 and 0.93) and (1.07 and 1.27), respectively. For the steelhead analysis, the 5% and 95% quantiles for the P values and γ's were (0.24 and 0.58) and (0.71 and 0.77), respectively.

$\hat{\sigma}^2_{\text{slp,p}}$ from the parameterization period were consistent with the expected distribution of $\hat{\sigma}^2_{\text{slp}}$ (Eq. 2). If so, $(\hat{\sigma}^2_{\text{slp,e}}/\hat{\sigma}^2_{\text{slp,p}}) \sim F(\text{df}_{\text{slp}}, \text{df}_{\text{slp}})$. We examined three paired lengths of parameterization and evaluation periods (10 yr, 10 yr, $\text{df}_{\text{slp}} \approx 1.4$), (15 yr, 15 yr, $\text{df}_{\text{slp}} \approx 1.96$), and (20 yr, 20 yr, $\text{df}_{\text{slp}} \approx 3.0$). This allowed us to compare the observed $\hat{\sigma}^2_{\text{slp}}$ ratios to three different expected F distributions corresponding to the different df_{slp} values. To estimate F distributions with low degrees of freedom, we needed a large sample size, and therefore we pooled the chinook and steelhead data and did not subsample the Snake River spring/summer chinook and Oregon Coast chinook ESUs. This analysis studied the distribution of $\hat{\sigma}^2_{\text{slp}}$; the next analysis explored the degree and effect of bias between $\hat{\sigma}^2_{\text{slp}}$ and σ^2_{p}.

Cross-validating the probability of crossing population thresholds

The DA estimate of the probability that an observed trajectory will decline from O_{start} at the beginning of an evaluation period to at or below xO_{start} at the end of an evaluation period is the following:

$$\Pr(O_{\text{end}} \leq xO_{\text{start}}) = 1 - \Phi\left(\frac{\ln(x) + \hat{\mu}_R\tau_e}{\sqrt{2\sigma^2_{\text{np}} + \hat{\sigma}^2_{\text{slp}}\tau_e}}\right)$$

$$\text{assuming } \varepsilon_{\text{np}} \sim \text{Normal}(0, \sigma_{\text{np}}) \qquad (6)$$

where $\Phi(\cdot)$ is the cumulative distribution of the unit normal and τ_e is the length of the evaluation period (Dennis et al. 1991). We used a metric pertaining to the observed trajectory because the true trajectory is hidden. A point estimate of σ^2_{np}, $\hat{\sigma}^2_{\text{np}} = (\text{var}(\ln(N_{t+1}/N_t)) - \hat{\sigma}^2_{\text{slp}})/2$, was used for this calculation (see Appendix B). $\Pr(O_{\text{end}} \leq xO_{\text{start}})$ is much less sensitive to σ^2_{np} than other metrics, such as the probability that the time to first crossing is less than τ_e, and this makes it especially useful for validating bias in σ^2_{p} estimates.

We compared the observed fraction of evaluation periods experiencing a given decline to the expected fraction. The expected fraction is the average $\Pr(O_{\text{end}} \leq xO_{\text{start}})$ calculated over all segments. Differences between the expected and observed fractions may either indicate that the underlying DA approach is simply a poor approximation of the real trajectories, or may indicate persistent bias in the estimated parameters. For example, under- or overestimation of μ leads to under- or overestimation of the probability of crossing thresholds, whereas overestimation of σ^2_{p} leads to underestimation of the probability of hitting $x > 1$ thresholds combined with overestimation of the probability of hitting $x < 1$ thresholds.

RESULTS

Fig. 1 shows the observed and expected distribution of $\hat{\mu}_{R,\text{p}} - \hat{\mu}_{R,\text{e}}$. The close agreement between the observations and predictions supports that on average $\hat{\mu}_R$ is an unbiased estimator of the long-term rate of growth/decline and that the theoretical distribution of $\hat{\mu}_R$ is correct. To examine whether density dependence led to changes in μ, we examined the association between $\hat{\mu}_{R,\text{p}} - \hat{\mu}_{R,\text{e}}$ and the overall rate of growth/decline within a segment. The observed mean t statistics for segments increasing at >5% per year ($n = 42$), fluctuating between 2.5% and −2.5% annual growth ($n =$

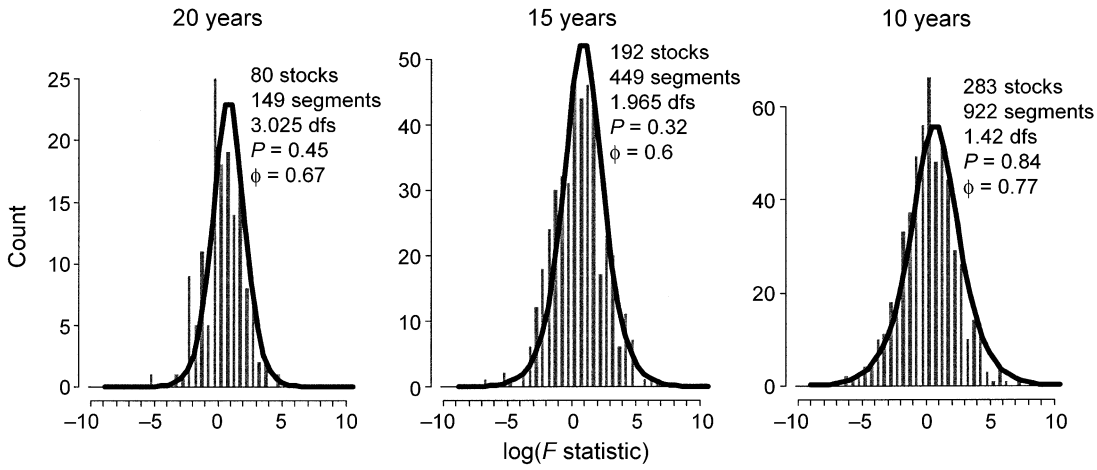

FIG. 2. Predicted vs. observed distribution of the process error estimate, $\hat{\sigma}^2_{slp}$. The solid lines show the theoretical F distribution of the ratio between $\hat{\sigma}^2_{slp}$ measured in the parameterization vs. evaluation periods (slp = slope). The histogram shows the observed distribution of log-transformed F statistics measured from the actual time series. The P values are from a Kolmogorov-Smirnov goodness-of-fit test of the observed F statistics to $\phi F(df_{slp}, df_{slp})$, where df_{slp} is ≈1.42, 1.96, and 3.0 for the 10-, 15-, and 20-yr analyses, respectively. For this analysis, segments from both chinook and steelhead time series were evaluated together. The analysis with 10-yr parameterization and evaluation periods (right-hand graph) required only 20 yr of data (rather than 30–40 yr); thus, we were able to use an additional 91 shorter time series of data for this specific analysis.

128), or decreasing >5% (n = 78) per year was 0.59 (P = 0.001), −0.34 (P = 0.18), and −0.06 (P = 0.80), respectively. The P values are for a test of whether the t statistics come from a t distribution with mean 0; i.e., whether $\hat{\mu}_{R,p} = \hat{\mu}_{R,e}$. Thus, for fluctuating or rapidly declining segments, there was no significant difference between the median annual growth rates in the parameterization and evaluation periods. However, for segments that exhibited rapid increases in population size, the mean rate of growth within the parameterization periods (with smaller population size) was significantly greater than the mean rate within the evaluation periods (with larger population size).

Fig. 2 shows the observed vs. expected distribution of F statistics for the $\hat{\sigma}^2_{slp,e}/\hat{\sigma}^2_{slp,p}$ ratios. In the plots, the F statistics were log-transformed to make visual comparison easier. The observed distributions were very close to (a constant, ϕ) × $F(df_{slp}, df_{slp})$ with the expected degrees of freedom. Fit was determined by Kolomogorov-Smirnov goodness-of-fit tests. This indicated that $\hat{\sigma}^2_{slp}$ had the expected χ^2 distribution, but the constant, ϕ, indicated a consistent bias between $\hat{\sigma}^2_{slp,p}$ and $\hat{\sigma}^2_{slp,e}$. Specifically, $\hat{\sigma}^2_{slp,e}$ was greater than $\hat{\sigma}^2_{slp,p}$ by a factor of 1.3, 1.7, and 1.5 for the 10-, 15-, and 20-yr analyses, respectively. Closer examination showed that the bias occurred only in the rapidly declining stocks characterized by an evaluation period with very low count numbers (e.g., 0–10 egg nests counted in a census) following a parameterization period with higher count numbers. We suspect that the bias occurred because of increased sampling error when counts are very low. When this happens, large percentage errors are common, because doubling or even tripling of the count represents a small difference in absolute numbers. Higher sampling error increases the expected value of $\hat{\sigma}^2_{slp}$ and would lead to the bias that we observed.

Fig. 3 shows the observed and expected fraction of segments declining to or below different threshold levels in 10 or 20 yr. This analysis used 20-yr parameterization periods. The solid gray lines show the expected fraction when parameters were calculated with the Dennis-Holmes method. We found close agreement between the observed and expected fractions for the 10-yr projections. For the 20-yr projections, the predictions began to diverge. The divergence is characteristic of the biases that occur due to variability (rather than bias) in μ estimates. However, some of the divergence also occurred because the set of time series with 40+ years needed for the 20-yr analysis included a larger fraction of increasing time series than the set of time series with 30+ years for the 10-yr analysis.

We noted previously that μ declined in the rapidly increasing segments. This should lead to underestimation of the risk of crossing thresholds, and indeed, for the rapidly increasing segments, we saw precisely this pattern. We also observed changes in $\hat{\sigma}^2_{slp}$, but we suspect that this was due to changes in the level of sampling error rather than temporal changes in σ^2_p. Indeed, the close correspondence between the observed and expected probabilities of decline indicated that the correspondence between $\hat{\sigma}^2_{slp}$ and σ^2_p was close enough to permit unbiased predictions. For comparison, the expected fractions using σ^2_p estimated with standard parameterization methods (Dennis et al. 1991) are shown with the dashed gray lines. For highly corrupted data, these methods overestimate σ^2_p, and the shift be-

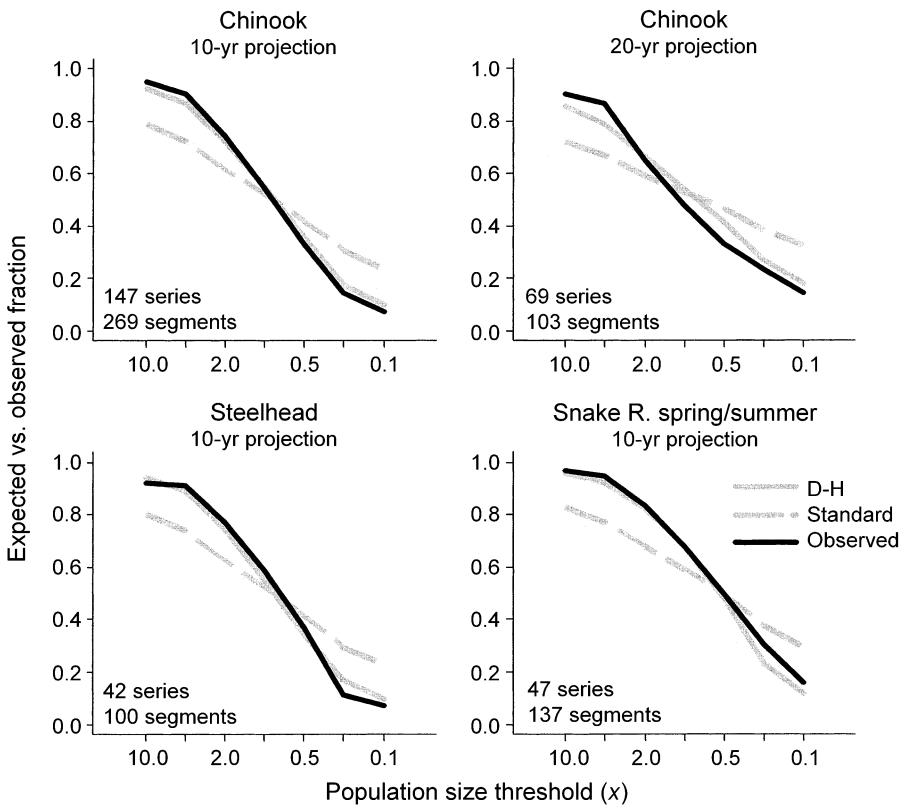

FIG. 3. Observed vs. predicted fraction of segments that reach population thresholds in 10 or 20 years. Segments were evaluated by determining whether the count at the end of the evaluation period, O_{end}, was less than or equal to xO_{start}, where O_{start} is the count at the start of the evaluation period. The x-axis (note log scale) is the factor x in xO_{start}. The 20-yr projection was only done for the chinook data set for which there were sufficient numbers of long time series. The predicted fractions were calculated using estimates of σ_p^2 from the ''standard'' methods of Dennis et al. (1991) vs. the Dennis-Holmes (''D-H'') method.

tween the observed and predicted lines is characteristic of high overestimation of σ_p^2. However, note that if data corruption is very low, the standard methods should give unbiased and less variable σ_p^2 estimates.

DISCUSSION

There are many reasons to expect viability assessments to fail (Coulson et al. 2001). A common concern is that population processes are not sufficiently stationary, meaning that the parameters describing the process change through time. For salmon, this might happen because reproduction is density dependent (as is normally assumed) or because environmental variability is autocorrelated. However, our analysis found that, despite an assumption of no density dependence and a host of other simplifying assumptions (such as low demographic stochasticity, stationarity, and low auto-correlation), DA methods worked remarkably well for describing the statistical distribution of fluctuating or declining population trajectories. Only for rapidly increasing populations did we see evidence of shifts in annual growth rates that were sufficient to cause overestimation of the risk of crossing thresholds. It should not be entirely surprising that DAs worked well for all but the rapidly increasing stocks; the asymptotic behavior of generalized stochastic age-structured population models (with survivorships and fecundities drawn from any of a variety of statistical distributions) is described by a DA model. Serious problems would not be expected unless there were serious violations of the stationarity assumptions. Our results, however, are striking because we concentrated exclusively on ''problematic time series'' plagued by high observation error and other non-process error. This noise masks the environmental variability that drives the statistical distributions of stochastic population trajectories. Indeed, it has been one of the criticisms of PVA analyses that they are sensitive to errors in the data (Ludwig 1999). Using a new method for correcting for such problems, we showed that unbiased estimates of parameters and probabilities of decline are possible. Our cross-validation results give empirical support for the Dennis-Holmes method for error-ridden data sets and, more generally, for count-based PVA analyses using diffusion approximations.

In any PVA analysis, one needs to address how to

present risk metrics, given parameter uncertainty. A strength of the Dennis-Holmes method (like other DA approaches) is that the parameter distributions can be calculated, have a strong theoretical foundation based on age-structured models, and, with these analyses, also have empirical support. These distributions allow one to calculate risk metric uncertainty in a rigorous manner. The most common approach has been to present point estimates of risk metrics with confidence intervals, but confidence intervals can be very broad and can give the mistaken impression that there is little support for any specific risk level. An alternate, arguably more informative, approach presents the overall support for risk ranges. For example, ''Given the data, what is the probability that the population is collapsing at >10% per year, or that it will go extinct in 20 yr or less?'' This approach is common in Bayesian methods for conservation and fisheries biology (Hilborn and Mangel 1997, Wade 2000, 2001). Frequentist approaches using likelihood inference are also available (for a review, see Wade 2001).

Practical, count-based PVA methods have proliferated in the last ten years in response to the needs within conservation biology, because sufficient data for full PVA models are not normally collected or even planned. Indeed, a recent survey found that of 136 recovery plans approved by the U.S. Fish and Wildlife Service, only 23% proposed collecting sufficient data for an age-structured PVA model, whereas 78% planned to collect data sufficient for a count-based PVA (Morris et al. 2002). Methods to deal with high data corruption greatly expand the utility of count-based PVAs, but new methods are still needed in many areas. Standardized methods for dealing with cycles and trends within population time series and for incorporating incomplete life history information are clearly needed. Algorithms used in computer graphics and engineering for analyzing corrupted signals may prove particularly useful in these regards. However, they will need to be adapted to fit the constraints particular to ecological and conservation data.

Literature Cited

Boyce, M. S. 1992. Population viability analysis. Annual Review of Ecology and Systematics **23**:481–506.

Brook, B. W., J. J. O'Grady, A. P. Chapman, M. A. Burgman, H. R. Akç akaya, and R. Frankham. 2000. Predictive accuracy of population viability analysis in conservation biology. Nature **404**:385–386.

Coulson, T., G. M. Mace, E. Hudson, and H. Possingham. 2001. The use and abuse of population viability analysis. Trends in Ecology and Evolution **16**:219–221.

Dennis, B., P. L. Munholland, and J. M. Scott. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs **61**:115–143.

Fagan, W. F., E. Meir, J. Prendergast, A. Folarin, and P. M. Kareiva. 2001. Characterizing vulnerability to extinction for 758 species. Ecology Letters **4**:132–138.

Fieberg, J., and S. P. Ellner. 2000. When is it meaningful to estimate an extinction probability? Ecology **81**:2040–2047.

Gerber, L., D. DeMaster, and P. Kareiva. 1999. Grey whales and the value of monitoring data in implementing the U.S. Endangered Species Act. Conservation Biology **13**:1215–1219.

Ghahramani, Z., and G. E. Hinton. 1996. Parameter estimation for linear dynamical systems. Technical Report **CRG-TR-96-2**. Department of Computer Science, University of Toronto, Toronto, Canada.

Hilborn, R., B. G. Bue, and S. Sharr. 1999. Estimating spawning escapements from periodic counts. Canadian Journal of Fisheries and Aquatic Sciences **56**:888–896.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Sciences (USA) **98**:5072–5077.

Ludwig, D. 1999. Is it meaningful to estimate a probability of extinction? Ecology **80**:298–310.

Morris, W. F., P. L. Bloch, B. R. Hudgens, L. C. Moyle, and J. R. Stinchcombe. 2002. Population viability analysis in endangered species recovery plans: past use and future improvements. Ecological Applications **12**:708–712.

Nicholls, A. O., P. C. Viljoen, M. H. Knight, and A. S. Van Jaarsveld. 1996. Evaluating population persistence of censused and unmanaged herbivore populations from the Kruger National Park, South Africa. Biological Conservation **76**:57–67.

NMFS (National Marine Fisheries Service). 2000. Reinitiation of consultation on operation of the federal Columbia River power system, including the juvenile fish transportation program, and 19 Bureau of Reclamation projects in the Columbia Basin. Endangered Species Act, Section 7 consultation. Biological Opinion.

Shumway, R. S., and D. S. Stoffer. 1982. An approach to time series smoothing and forecasting using the EM algorithm. Journal of Time Series Analysis **3**:253–264.

Tuljapurkar, S. D. 1989. An uncertain life: demography in random environments. Theoretical Population Biology **35**: 227–294.

Wade, P. R. 2000. Bayesian methods in conservation biology. Conservation Biology **14**:1308–1316.

Wade, P. R. 2001. The conservation of exploited species in an uncertain world: novel methods and the failure of traditional techniques. Pages 110–143 *in* J. D. Reynolds and G. Mace, editors. Conservation of exploited species. Cambridge University Press, Cambridge, UK.

APPENDIX A

A table presenting the salmon time series used in the analyses is available in ESA's Electronic Data Archive: *Ecological Archives* E083-047-A1.

APPENDIX B

Derivations of the distributions of the estimated parameters are available in ESA's Electronic Data Archive: *Ecological Archives* E083-047-A2.

APPENDIX C

A derivation of the distributions for cross-validating parameter estimates is available in ESA's Electronic Data Archive: *Ecological Archives* E083-047-A3.

SUPPLEMENT 1

The S-PLUS code for estimating Dennis-Holmes parameters and calculating risk metrics from a time series is available in ESA's Electronic Data Archive: *Ecological Archives* E083-047-S1.

SUPPLEMENT 2

Raw data for the salmon time series used in the cross-validations are available in ESA's Electronic Data Archive: *Ecological Archives* E083-047-S2.