Northwest Fisheries Science Center Processed Report

Final Report on the Technical Workshop on Population Trends and Extinction Metrics

Workshop held December 5, 2003

U.S. Department of Commerce
National Oceanic and Atmospheric Administration
NOAA Fisheries
Northwest Fisheries Science Center
2725 Montlake Blvd. East
Seattle, WA 98112-2097

October 2004

Contents

Introduction

This report, completed in March 2004, summarizes the December 5th, 2003 technical workshop on λ methods and the formal scientific review of methods used by NWFSC for estimating population trends and extinction metrics as pertains to the 2000 FCRPS Biological Opinion. This report includes the written material and comments submitted by federal, state, tribal and independent scientists before and after the workshop. Specific recommendations for the Regional Office pertaining to the use of these methods are attached at the end of this report.

Process

As part of a review of the current methods used for assessing status and risk in Pacific Northwest salmonids, a white paper reviewing the methods and research papers on these methods was prepared by Dr. Elizabeth E. Holmes (NWFSC), "Review of methods, progress and cross-validation studies pertaining to population trend and risk assessment for Columbia River salmonids". Subsequently a workshop was held on December 5, 2003 at the Northwest Fisheries Science Center. The workshop was led by two outside reviewers: Dr. Selina Heppell from Oregon State University and Dr. Douglas Deutschman from San Diego State University. The purpose of this workshop was to formally review the white paper and to review other scientific progress made by federal, state, tribal and independent scientists in estimating population trends and extinction metrics since the 2000 FCRPS Biological Opinion. An announcement was sent out to state, tribal and independent scientists inviting them to give presentations of their research at the workshop. Participants were also invited to submit written comments on the white paper by Dr. Holmes to the two reviewers.

After the workshop, Dr. Heppell and Dr. Deutschman prepared a report, which reviewed the methods as presented in the Holmes white paper. The reviewers were given the following initial guidance in regards to this report: "The review should be based on a) your professional impartial experience, b) the comments you will receive, c) the workshop presentations and discussions, and d) the background material. This review should consider both the methods, the research supporting those methods, and comment on the major areas needed for future research and methodologies." Additionally the following guidance was given on the day of the workshop and a copy of this guidance was handed out to all participants:
"Charge to the Reviewers

- Scientific review of the paper by Eli Holmes describing methods used in the FCRPS Biological Opinion.
- General Questions for the panel to consider when preparing their review of the workshop:
    - Given the type and quality of data we have across multiple stocks, are we using the best available methods to analyze this data?
    - Looking forward and given the progress made thus far, what are the most substantial challenges and opportunities for improvement of our capability of accessing and portraying status and trends in Columbia River stocks?"

The final report from Dr. Heppell and Dr. Deutschman was received on January 31, 2003.

Technical Workshop on Population Trends and Extinction Metrics

December 5, 2003

9:30am	Introductions, background and purpose of workshop
10:00	Eli Holmes Northwest Fisheries Science Center “Estimation and calculation risk metrics for stochastic population processes”
10:30	John Payne Northwest Fisheries Science Center “The flip side of extinction: when can a population be de-listed?”
11:00	Saang-Yoon Hyun Columbia River Inter-tribal Fish Commission “Risk metrics to spring/summer chinook salmon and steelhead in the Snaker River Basin.”
11:30-12:30	Lunch
12:30	Rich Hinrichsen Hinrichsen Environmental Services “State space approaches to estimating population growth rates”
1:00	Charlie Paulsen Paulsen Environmental Research, Ltd. “Lambda estimation and prediction: problems and (possible) solutions”
1:30	Earl Weber Columbia River Inter-tribal Fish Commission "A variable slope stock recruitment function for nonstationary data"
2:00-2:30	Break
2:30-4:30	Open discussion
4:30	Adjourn

December 5 Workshop Attendee List

Name	**Affliation**	**Email**
Anderson, Jim	University of Washington	jim@cbr@washington.edu
Athearn, Jim	US Army Corp of Engineers	jim.b.athearn@nwd01.usace.army.mil
Beaty, Roy	BPA	rebeaty@bpa.gov
Bellerud, Blane	NOAA	Blane.Bellerud@noaa.gov
Bilby, Bob	Weyerhaeuser	bob.bilby@weyerhaeuser.com
Blakley, Ann	WDFW	blaklab@dfw.wa.gov
Deutschman, Douglas	San Diego State U	doug@sciences.sdsu.edu
Espensen, Barry	CB Bulletin	bespensen@mev.com
Geisleman, Jim	BPA	jrgeiselman@bpa.gov
Grabowski, Steve	US Bureau of Reclamation	sgrabowski@pn.usbr.gov
Hassemer, Pete	Idaho Fish & Game	phassemer@idfg.state.id.us
Hatcher, Lynn	NOAA	lynn.hatcher@noaa.gov
Heppell, Selina	Oregon State U	selina.heppell@oregonstate.edu
Hinrichsen, Richard A.	Hinrichsen Environ. Services	hinrich@seanet.com
Holmes, Eli	NWFSC	eli.holmes@noaa.gov
Hyun, Saang-Yoon	CRITFC scientist	HYUS@critfc.org
Johnson, David H.	WDFW	johnsdhj@dfw.wa.gov
Masonis, Rob	American Rivers	reasonis@amriver.org
McCann, Jerry	Fish Passage Center	jmcann@fpc.org
McClure, Michelle	NWFSC	michelle.mcclure@noaa.gov
McDonald, Lyman	ISAB	lmcdonald@west-inc.com
Newsom, Michael	US Bureau of Reclamation	mnewsom@pn.usbr.gov
Paulsen, Charlie	Paulsen Environ. Research	cpaulsen@spiritone.com
Paulsen, Mike	Washington Farm Bureau	walf@3-cities.com
Payne, John	NWFSC scientist	John.c.Payne@noaa.gov
Petrosky, Charlie	Idaho Fish & Game	cpetrosk@idfg.state.id.us
Peven, Chuck		chuckp@chalenpud.org
Plummer, Mark	NWFSC	mark.plummer@noaa.gov

Ratner, Sue	NWFSC	sue.ratner@noaa.gov
Rudolph, Bill	NW Fishletter	billr@newsdata.com
Shutters, Marvin	US Army Corp of Engineers	marvin.k.shutters@nww01.usace.army.mil
Slatter, Dave	Nez Perce Tribe Fisheries	daves@nezperce.org
Smith, Stephen H.	BPA	huntersmith@canby.com
Stein, John	NWFSC	john.e.stein@noaa.gov
Toole, Chris	NMFS Portland Office	chris.toole@noaa.gov
Ungerecht, Todd	NOAA	todd.ungerecht@noaa.gov
Wang, Shizhen	Quinault Indian Nation	szwang@qwest.net
Weber, Earl	CRITFC scientist	WEBE@critfc.org
Wilson, Amilee	WDFW	wisonalw@dfw.wa.gov
Wilson, Paul	US Fish & Wildlife	paul_h_wilson@rl.fws.gov
Young, Frank	CBFWA	frank.young@cbfwa.org

Review of methods, progress and cross-validation studies pertaining to population trend and risk assessment for Columbia River salmonids

E. E. Holmes
Northwest Fisheries Science Center
National Marine Fisheries Service
2725 Montlake Blvd. East
Seattle, WA 98112-2097

November 2003

BACKGROUND FOR THIS WHITE PAPER

The 2000 Federal Columbia River Power System (FCRPS) Biological Opinion (FCRPS Biop) evaluated whether the operation of the FCRPS, when combined with survival rates expected to occur in all other life stages of ESA listed salmonids, would result in a "high likelihood of survival and a moderate-to-high likelihood of recovery." This qualitative determination was informed by quantitative estimates for several evolutionarily significant units (ESU). Specifically, NOAA Fisheries evaluated:

- whether or not there would be a 5% or lower probability of absolute extinction of natural spawners within 24- and 100-year periods as a "metric indicative of survival;"
- whether or not there would be at least a 50% probability of the 8-year geometric mean natural spawners being equal to, or greater than, interim recovery abundance levels in 48 and 100 years as a primary "metric indicative of recovery;"
- and whether or not there would be at least a 50% likelihood of the annual population growth rate ("lambda") being equal to, or greater than, 1.0 as an alternate "metric indicative of recovery" for populations lacking interim recovery abundance goals.

The basis for each of these indicator metrics was an analysis of the population growth rate associated with time series for relevant spawning aggregations. Population growth rate was calculated using the methods described in McClure et al. (2003). The Biological Opinion specified that several tests based on population growth rate would be conducted in 2005 and 2008 to ensure that implementation of the Biological Opinion was on track and that populations were not declining further. The Biological Opinion assumed that by 2005 there would be more information about methods of calculating population growth rate:

> "NMFS anticipates that methods of assessing annual population growth rates will have been refined, based on NMFS' research efforts, those of the Action Agencies, or those of independent scientists. In anticipation of this normal progress in scientific methods, NMFS does not now define a specific method by which population growth rate will be determined for its mid-point evaluations. By March 1, 2005, NMFS will choose the most appropriate method(s) to estimate population growth rate from the peer-reviewed literature, based on collaboration with the Action Agencies, USFWS, and the state and Tribal comanagers."

In June 2003, the Biological Opinion was remanded in National Wildlife Federation v. NMFS. NOAA Fisheries is currently in the process of revising the Biological Opinion and re-evaluating the effects of FCRPS operations and offsite mitigation activities. To facilitate this process, the NOAA Fisheries Northwest Regional Office (NWRO) requested that the Northwest Fisheries Science Center (NWFSC) conduct the above-referenced review of population growth rate estimation methods in 2003. In addition, the NWRO requested that that the NWFSC review related methods of characterizing population trends, especially those that had been suggested as alternatives to "lambda" estimation in comments on the draft of the original Biological Opinion and in comments or litigation since the Biological Opinion was issued.

INTRODUCTION

The purpose of this report is to review and discuss methods for estimating and presenting population trends and extinction risks for Columbia River salmonid populations to support management decisions, such as the ESA Section 7 determination in the 2000 FCRPS Biological Opinion and the anticipated 2005 and 2008 check-in tests. This report reviews research since 2000, which tests and validates diffusion approximation methods for estimating population trends and risks. This review summarizes information from the following publications:

Holmes, E. E. 2004. Beyond theory to application and evaluation: diffusion approximations for population viability analysis. In press in Ecological Applications.

Fagan, W. F., E. E. Holmes, J. J. Rango, A. Folarin, J. A. Sorensen, J. E. Lippe, and N. E. McIntyre. 2003. Cross-validation of quasi-extinction risks from real time series: an examination of diffusion approximation methods. Pre-print.

McClure, M., E. Holmes, B. Sanderson, and C. Jordan. 2003. A large-scale, multi-species risk assessment: anadromous salmonids in the Columbia River Basin. Ecological Applications 13: 964-989.

Holmes, E. E. and W. F. Fagan. 2002. Validating population viability analysis for corrupted data sets. Ecology 83: 2379-2386

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Science 98: 5072-5077.

Summary of work and changes as they pertain to the methods in FCRPS Biop

Changes in the methods for estimating trend and risk:

1) Running sum filter has been standardized to use a simple sum of four consecutive spawner counts. The work leading up to Holmes & Fagan (2002) clarified that this was better than the age-structure based running sum that was originally used.

Cross-validation work

The bulk of the work has focused on validating the methods using real time series (Holmes & Fagan 2002, Fagan et al. 2003) and more realistic simulations that include density-dependence (Holmes 2004). Also the underlying assumptions of the diffusion model were tested using simulations of salmon models with density-dependence (Holmes 2004).

Expressing uncertainty

Holmes & Fagan (2002) test the variability in parameter estimates from the Dennis-Holmes method and found that the variability is properly estimated. Holmes (2004) begins looking in-depth at how to express uncertainty in a way that it can best inform regulatory decision-making. Confidence intervals are commonly given, but are not very useful beyond showing that there is high or low uncertainty. Bayesian approaches are explored in Holmes (2004). A Bayesian metric is also used in McClure et al. (2003), specifically the probability that λ is less than 1.0 or less than 0.9.

THE NATURE OF POPULATION TRAJECTORIES AND RISK ESTIMATION

Real populations do not grow or decline at fixed rates, but rather show year-to-year variability in population growth rates, which leads to a population trend that varies about some long-term growth rate. Figure 1 shows an example of three population trends that each have the same long-term trend (5% per year decline) and the same year-to-year variability. Even though the population trends were generated with the same underlying dynamics, the trajectories are different. This is nature of populations: random chance means that there are a range of different possible population trajectories given some underlying population dynamics. Even though we cannot predict exactly what will happen in the future, if we could estimate the underlying dynamics governing the population trajectories, we could estimate the probability of different futures, i.e. we could estimate the probability of reaching critical thresholds. We can also estimate whether the population has long-term declining dynamics. To do this, we will need to estimate the following: the long-term rate of decline (or growth), the year-to-year variability in yearly population growth, and the amount of corruption in our data. Within the population dynamics literature, the year-to-year variability in yearly population growth is termed 'process error'; note that it is not technically 'error' in the layman's sense of the word, but rather variability. The rest of the variability is termed 'non-process error' and this includes actual observations errors. For the purpose of this review, one can think of process error as the variability that drives the long-term variability of future population size and the non-process error as the data corruption that is preventing us from estimating the process error.

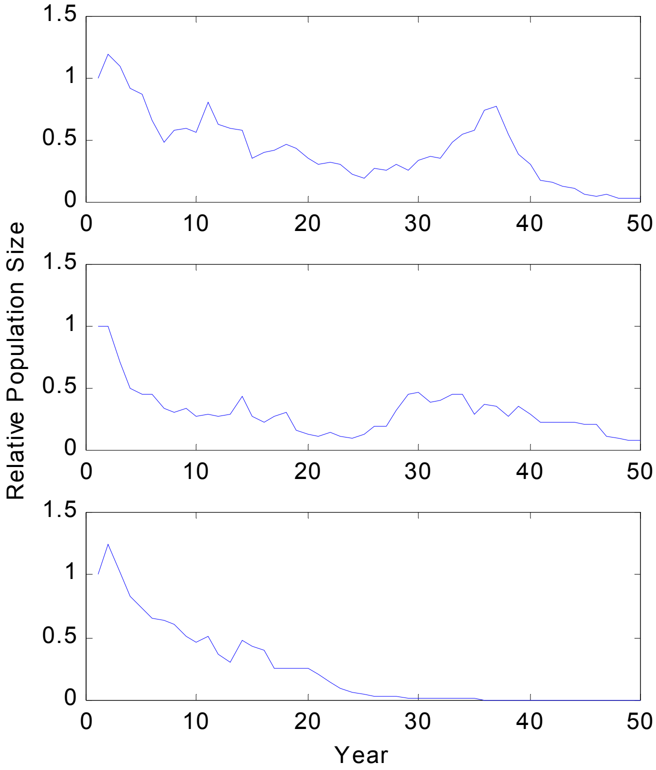

Figure 1. Sample simulated population trajectories from a population with average 5% yearly decline. The underlying population dynamics are identical. The differences are due to chance.

The term λ denotes the long-term rate of population decline (or growth). It is simply the long-term trend that you would observe if you had a very, very long time series of the population. The term λ is the standard notation in the conservation biology literature. $\lambda = 1$ means stable, $\lambda = 1.01$ means roughly increasing 1% per year, and $\lambda = 1.05$ means roughly increasing 5% per year. Similarly, $\lambda = 0.99$ and $\lambda = 0.95$ mean roughly declining 1% and 5% per year, respectively. Note that we can only estimate λ; we never know the true λ. Our estimates may be unbiased, but that still means that there is a 50/50 chance that the true λ is above or below our estimate.

One of the most common questions is "If λ is the population trend, why not just present the overall trend observed in the data, such as a regression of log numbers versus time?" as opposed to going through the analysis based on theory concerning the dynamics of population trajectories, which is presented in the next section. There are two main reasons why this is insufficient.

1) *We need to estimate uncertainty*. The trend tells you what happened but does not by itself tell you how likely it is that this trend happened by chance and that the long-term dynamics are actually quite different. For example, suppose we collect data on a population that has a true long-term average rate of decline of 12% yearly. Figure 2 shows an example of the population trend observed from 20-year consecutive time series from this population. Segment 1 is from year 1-20, segment 2 is from year 2-21, etc. The wavy lines show the estimates using different methods for estimating the trend; the true value is the straight dashed line. The solid line ("ML") is a simple regression of log natural abundance. The wavy dashed line shows the runsum method used in McClure et al. (2003) and the Biop. There is much variability in the observed trend in a 20-year segment. This variability is an unavoidable aspect of analyzing stochastic population processes. Population dynamics theory allows us to estimate this year-to-year variability and thus estimate how likely it is that a particular observed trend came from a population with a particular true λ (such as an increasing or declining population). But to do this, the estimate of the underlying process error in the population dynamics is needed. A natural response would be to argue that standard regression analyses will give you the uncertainty of the estimated trend, but unfortunately such analyses attribute all error to non-process error and will give you incorrect uncertainty estimates.

2) *We need to estimate probabilities of crossing critical thresholds*. The trend by itself does not give much information about the probability of dropping below critical population sizes. We cannot simply extend the trend into the future and see when our line crosses the threshold. Populations vary from year to year and even a population that has a positive growth rate still has some probability of dropping below the critical threshold by chance. To estimate this probability, we again need to estimate the process error driving the variability in long-term population sizes.

In the following section (section I), I review how the parameters driving a population process are estimated using diffusion approximation methods. This section also reviews the extensive cross-validation work that was done to verify the applicability of these methods for salmon populations. This section directly applies to the methods used in the FCRPS Biological Opinion. At the end of this section, there is a discussion of alternative risk estimation methodologies and why they were not used. The next section (section II) discusses work that goes beyond the methods used in the 2000 FCRPS Biological Opinion. One of the challenges when presenting scientific analyses is presenting the uncertainty in a useful and accurate manner. It is tempting to use the point estimates of risk metrics (i.e. 'this stock has a λ of

0.981') and ignore that this is a statistical estimate. λ = 0.981 may be the most likely value given the data, but λ = 0.99 is probably almost equally as likely and λ = 1.01 may be entirely plausible. Section II illustrates the use of probability curves as a way to formally express this uncertainty. This is a standard approach in decision theory for resource management.

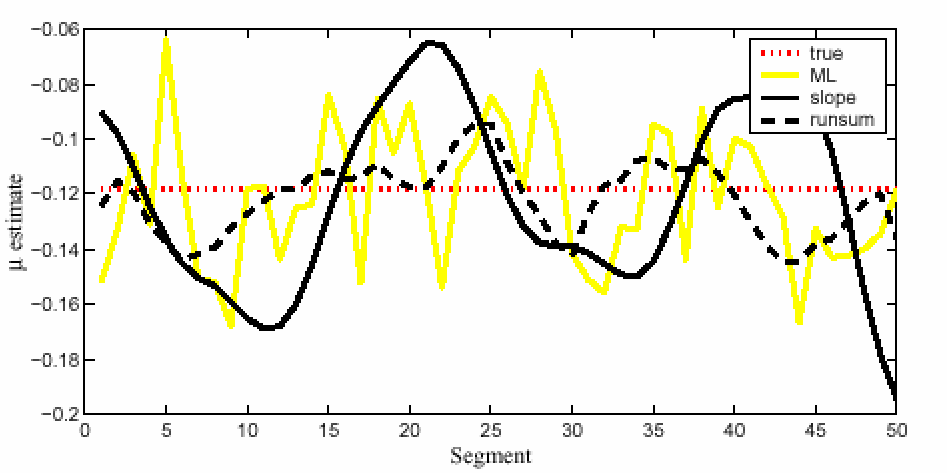

Figure 2. Estimated log(λ) from 20-year segments in a time series. Segment 1 is year 1-20, segment 2 is year 2-21, segment 3 is year 3-22, etc. This shows how the estimates vary depending on the segment observed. The "runsum" method is that used in McClure et al. (2003) and for the Biological Opinion calculations.

I. DIFFUSION APPROXIMATION METHODS FOR POPULATION VIABILITY ANALYSIS

In the last decade, diffusion approximation (DA) methods have been developed that use count data alone (for example, spawner counts) for the estimation of population viability analysis (PVA) risk metrics, such as the probability of crossing extinction thresholds, mean passage times, and average long-term rates of population growth or decline (Lande and Orzack 1988, Dennis et al. 1991). These methods have since been used to estimate extinction risks for numerous species of conservation concern (Dennis et al. 1991, Nicholls et al. 1996, Gerber et al. 1999, Morris et al. 1999, McClure et al. 2003). The appeal of DA methods from an applied standpoint is their simplicity and their reliance on simple census data alone (e.g. neither age-structure, cohort-level analyses, or total fish numbers are required). They have become one of the basic quantitative tools presented in recent books on PVA methods (Morris and Doak 2002, Lande et al. 2003).

Diffusion approximation methods stem from theory concerning the behavior of stochastic age-structured population models with no density-dependence,

$$\begin{bmatrix} n_{1,t+1} \\ n_{2,t+1} \\ n_{3,t+1} \\ \vdots \\ n_{k,t+1} \end{bmatrix} = \mathbf{A}_t \begin{bmatrix} n_{1,t} \\ n_{2,t} \\ n_{3,t} \\ \vdots \\ n_{k,t} \end{bmatrix} \qquad [1]$$

where $\mathbf{A}_t$ is the stochastic population transition matrix, e.g. a Leslie matrix, for time t. Note that most types of cohort or otherwise age-structured population simulations with no density-dependence are specific cases of the general model in Eq. 1. For such models, the asymptotic behavior of the total population size, $N_t = \sum_i n_{i,t}$, is a stochastic exponential process (Tuljapurkar and Orzack 1980, Tuljapurkar 1989):

$$N_t = N_0 \exp(\mu t + \varepsilon_p), \qquad \varepsilon_p \sim \text{normal}(0, \sigma^2 t) \ \text{ for } t \ \text{ big}. \qquad [2]$$

and log N_t/N_0 is distributed normal with mean=μt and variance=$\sigma^2 t$ for t big. The parameter μ in Eqn. 2 determines the rate at which the median log population size, log N_t, increases through time, while σ^2 determines the rate at which the distribution spreads, or in other words, the variability of potential population sizes at time $t+\tau$.

Diffusion approximation methods assume that Eqn. 2 holds for all $\tau > 0$ including small τ and that the ε are independently and identically distributed. This allows one to model the population as a diffusion process (Lande and Orzack 1988):

$$\begin{aligned} \partial p / \partial t &= -\mu \partial p / \partial x + (\sigma^2/2) \partial^2 p / \partial x^2 \\ p &= P(\log N_t = x \mid N_0 = x_0) \end{aligned} \qquad [3]$$

P(y) means the probability of y. The diffusion model has the property that log N_t/N_0 is distributed normal(mean=μt, variance=$\sigma^2 t$) like the stochastic exponential process it is used to approximate. See Dennis et al. (1991) for a much fuller discussion of the diffusion approximation.

This approximation opens a toolbox of parameterization methods for linear models with normal error. It also provides analytical estimates of quasi-extinction probabilities, i.e. the probability of crossing a particular threshold at some time within a given time frame. Strictly speaking, however, an age-structured population process is not a diffusion process. However despite the assumption violations, the diffusion model approximates many types of stochastic age-structured population processes, as seen both from simulated and real data (Lande and Orzack 1988, Dennis et al. 1991, Holmes and Fagan 2002, Fagan et al. unpublished manuscript, Holmes 2004). In particular as will be reviewed below, the diffusion approximation works well for salmon population models (Holmes 2004).

Parameter estimation methods

Diffusion approximations for a particular PVA must be carefully selected since a poor choice results in poor, highly biased estimates which lead to poor, highly biased risk estimates. Holmes (2004) discusses these issues and careful selection of parameterization methods using salmon data as an example. McClure et al. (2003) presents methods for estimating log(λ) and σ^2, which have been used by NWFSC scientists for salmon PVA. These methods have been extensively validated with real and simulated salmon data (see next section).

The basic estimation methods currently used for the Biop are presented here without discussion; see McClure et al. (2003) for a discussion and examples. The methods use a running sum transformed time series of spawner counts defined as

$$R_t = \sum_{i=0}^{3} O_{t+i} \qquad [4]$$

where O_t is the spawner count at year t. The estimate for log(λ), which is denoted μ, is

$$\begin{aligned} \hat{\mu}_{run} &= \text{mean of } \log(R_{t+1}/R_t) \\ &\text{for } t = 1,2,3,\ldots,k-3. \end{aligned} \qquad [5]$$

The estimate of σ^2 uses the rate that the variance increases within the time series:

$$\hat{\sigma}^2_{slp} = \text{slope of var}\left[\log\left(\frac{R_{t+\tau}}{R_t}\right)\right] \quad \text{versus } \tau, \text{ intercept free} \qquad [6]$$
$$\text{for } t = 1,2,3,\ldots,k-3 \text{ and maximum } \tau = 4.$$

These estimators will likely appear somewhat peculiar on first introduction. Note that the μ estimate is very similar to a linear regression of log population counts (typically log spawner counts). Why use the estimator with a running sum transformation of the data? Extensive testing described in Holmes (2001), Holmes & Fagan (2002) and especially Holmes (2004) indicates that the $\hat{\mu}_{run}$ gives the least variable estimates of μ (see also Figure 2). Estimation of the process error is an especially difficult problem. Holmes (2004) reviews the currently available methods in the literature. Again extensive cross-validation work (see especially Holmes 2004) found that $\hat{\sigma}^2_{slp}$ performs the best for salmon data.

One of the difficult problems with analyzing salmon spawner data is that hatchery fish are input into the stocks. Perhaps the easiest way to see how this presents a problem for estimating λ is to consider the analogy of a mutual fund. Suppose you put $1000 into a mutual fund 5 years ago and now you have $8000. You would like to know what the average rate of return (this is λ) has been so that you can decide whether to keep you money in this fund or move to another. Normally, you would just take (8000/1000)^(1/5) = 1.51, which means that your fund returned an incredible 51% per year. However, your benevolent aunt has been automatically adding $100 a month to your brokerage account, and you need to factor this in (these are the hatchery fish). Problem is you don't know whether her monthly gift was added to your mutual fund (the hatchery fish reproduce) or was simply deposited to your brokerage account but not invested (the hatchery fish don't reproduce). Without this information, you can only deduce the range of the possible average rates of return for your mutual fund. If not added to mutual fund, the rate of return was ((8000-100*12*5)/1000)^(1/5) = 1.15, which is still a nice 15% per year. If added to the mutual fund, rate of return is found by finding the λ that solves:

$$8000 = 1000^*\lambda\text{\^{}}5 + 100^*\sum_{i=0}^{12^*5} \lambda^{i/12},$$

which is λ = 1.05 and means a rather paltry 5% per year growth. Thus, knowing whether the monthly deposits were added to your mutual fund or not is a critical bit of information you need to evaluate how good a mutual fund you have. This is exactly the estimation problem we have with hatchery fish. We need to know whether or not they are reproducing in order to evaluate the underlying population growth rate. In McClure et al. (2003) and in Holmes (2004), the hatchery correction is presented. In the McClure et al. PVA, the range of λ for hatchery fish not reproducing versus are reproducing is shown. For the Biological Opinion, the range of λ is shown for hatchery fish reproducing 20% as effectively as wild-born fish versus 80% as well as wild-born fish.

Risk metrics

From the parameters μ and σ^2, a number of different risk metrics can be calculated. We have focused on two metrics. The first is the median yearly growth rate or the long-term yearly growth rate, which is denoted λ. Suppose you were able to observe 1000 20-year population trajectories with the same underlying dynamics (i.e. the same μ and σ^2 parameters) and each starting from the same initial size, much like Figure 1. The trajectories would all look different due to chance. The yearly growth rate you observed in the i^{th} (out of the 1000) trajectory is

λ_i = [(end population size)/(start population size)]^(1/(number of years)-1)

The median λ_i from all 1000 would be $\exp(\mu)$; on average 50% of trajectories would show a yearly growth rate greater than $\exp(\mu)$ in those 20 years and 50% would show a lower growth rate. An estimate of this median yearly growth rate is what we term λ. It also happens to be the yearly growth you would observe from a very long time series since λ_i goes to $\exp(\mu)$ as the number of years gets very large. This is why the λ estimate is referred as an estimate of the median yearly growth rate or the long-term yearly growth rate. For a particular time series with n years, the λ estimate is

$$\hat{\lambda} = \exp(\hat{\mu}_{run}) = \frac{1}{n-4}\log(R_{n-3}/R_1) \qquad [7]$$

The second metric is the probability of hitting a particular critical population threshold, N_e, within some period of time t_e, starting from the population size N_0. This is calculated from a diffusion approximation of the population process (Dennis et al. 1991):

$$\Pr(N_0 \to N_e \quad by \quad t_e) = \pi' * \Phi\left(\frac{-\ln(N_0/N_e) + |\hat{\mu}_{run}|\ t_e}{\hat{\sigma}_{slp}\sqrt{t_e}}\right)$$

$$+ \exp(2\ln(N_0/N_e)|\hat{\mu}_{run}|/\hat{\sigma}^2_{slp})\Phi\left(\frac{-\ln(N_0/N_e) - |\hat{\mu}_{run}|\ t_e}{\hat{\sigma}_{slp}\sqrt{t_e}}\right), \quad t_e > 0 \qquad [8]$$

$$\text{where} \quad \pi' = \left\{\begin{matrix} 1, \ if \quad \hat{\mu}_{run} \le 0 \\ \exp(-2\hat{\mu}_{run}\ln(N_0/N_e)/\hat{\sigma}^2_{slp}, \ if \quad \hat{\mu}_{run} > 0 \end{matrix}\right\}.$$

The function Φ is the standard normal cumulative distribution function. If you are interested in percentage-wise declines, e.g. 50%, 75% or 90%, then it is not necessary to know the actual population size, since $(N_0/xN_0) = (1/x)$. In this case, the probability of crossing critical thresholds can be estimated using on index data without information on the total number of spawners. If however, declines to specific absolute thresholds are of interest, total spawner counts are needed and it is also necessary to transform the spawner count into a count that reflects the total population rather than just spawners in a particular year. See McClure et al. (2003) for a discussion of this transformation.

Validation studies of diffusion approximations for salmon populations

Here I review cross-validation studies of the performance of the diffusion approximation model for salmon data and populations, including populations experiencing density dependence. Holmes (2004) discusses evaluation of the diffusion approximation and estimation methods using simulated data. This study used detailed population models for Upper Columbia River steelhead, Snake River fall chinook, and Snake River spring/summer chinook as examples. The models were parameterized from survivorship and fecundity data from these ESUs. The models include density-dependence in parr to smolt survivorship reflecting that found in low density Snake River chinook stocks (Achord et al. 2003). The models also include sampling error in the range of that observed for redd-count data (standard error 0.3 to 0.85).

The first question in this study was whether a diffusion approximation correctly described the behavior and probability of crossing thresholds for the age-structured models. The first test described in Holmes (2004) is an examination of the linearity assumptions inherent in the diffusion approximation. This key test is somewhat technical, and is described in Holmes (2004). The results of this test were that the linearity assumptions were satisfied for $t > 5$ years which means that (as is well-known) the diffusion approximation should be used to make medium and long-term projections not short-term projections ($t < 5$ years). The second test was whether a diffusion approximation would properly characterize the probability that the simulated time series would cross a threshold (in this case, 90% decline) in different time frames. This analysis is shown in Figure 3. The gray line shows the actual probability of crossing the 90% decline threshold within different time frames (determined by repeating the salmon simulations 1000s of times) versus the probabilities from a diffusion approximation. This illustrates that the probability of 90% decline in these salmon time series can be described by a diffusion approximation.

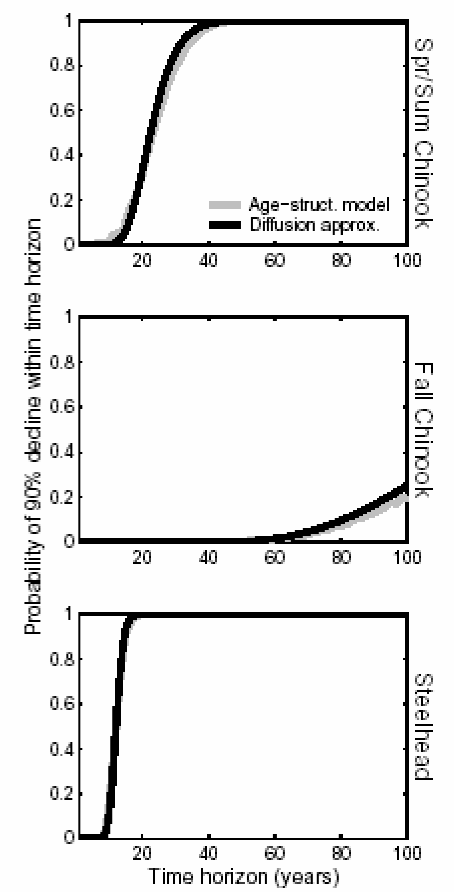

Figure 3. Actual versus predicted probability of 90% decline within different time horizons. From Holmes (2004).

Simply because a diffusion approximation exists which properly characterizes a particular population process does not mean that we can estimate the parameters for that process given realistic data constraints. Holmes (2004) also studies estimation performance given data constraints faced by the PVA of Columbia River salmon stocks (McClure et al. 2003): 1) counts of only the spawning segment of the populations, 2) time series limited to 20 years, 3) severe age-structure perturbations in the beginning of some time series due to reproductive collapses during dam construction (Williams et al. 2001), and 4) high observation error in the spawner counts. Figure 4 shows box plots of the estimates of log(λ) following the estimation methods described above (also in McClure et al. 2003) for 1000 random simulations from the three species' models. The output from the models (spawner counts) was 'corrupted' by different levels of sampling error: age (meaning an age perturbation due to no reproduction in one year), low, medium and high observer error. The dotted line in the graph shows the true value of log(λ). In the box plots, the middle line is the median estimate of log(λ) and the box encloses 75% of the estimates. As can be seen in the figure, the runsum method for estimating log(λ) works for these simulated salmon time series even within the data constraints; the median estimate is the true value even with added observer error in the spawner counts.

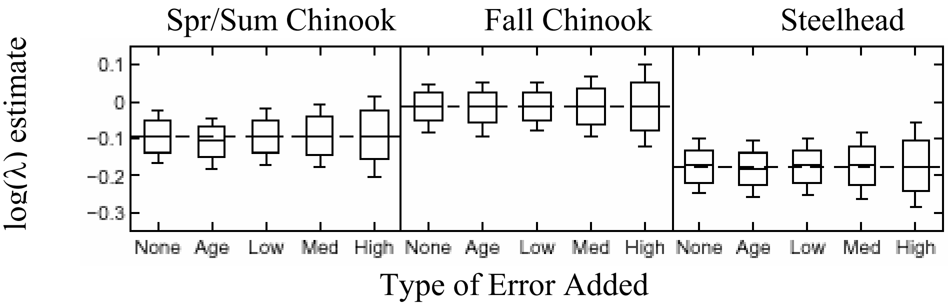

Figure 4. Distribution of log(λ) estimates using $\hat{\mu}_{run}$ from 1000 simulated time series from age-structured models of Snake River spring/summer chinook, Snake River fall chinook, and Upper Columbia steelhead. The models include density-dependent smolt survivorship. From Holmes (2004).

Figure 5 shows a similar analysis for the estimation of the process error, termed σ^2. Recall that the process error specifies the variability of potential future population trajectories and is a key parameter determining the probability of crossing thresholds. This analysis indicates that for low observation error $\hat{\sigma}^2_{slp}$ provides an unbiased estimate of the true value of σ^2, but as observation error increased, the estimate becomes increasingly biased. 'Medium' represents the average estimate of typical observation error in the Columbia River data based on studies of observer error in redd count data (see discussion in Holmes 2004).

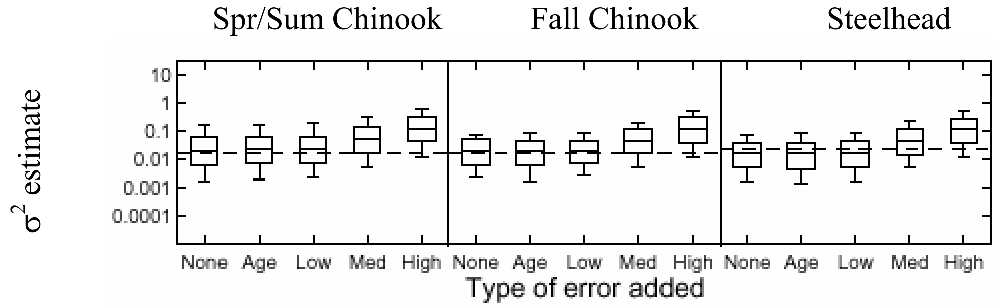

Figure 5. Distribution of $\hat{\sigma}^2_{slp}$ estimates from 1000 simulated time series from age-structured models of Snake River spring/summer chinook, Snake River fall chinook, and Upper Columbia steelhead. The models include density-dependent smolt survivorship. From Holmes (2004).

Simulated data is very useful, however it is ‘simulated’ and certainly lacks some aspects of real time series data. Another cross-validation (Holmes and Fagan 2002) involved testing the bias and precision of the diffusion approximation parameter estimates using hundreds of real time series. The strategy was to use the first 15 years of a time series to predict the second 15 years of the time series. The bias and variability of these predictions could then be tested against the predicted bias and variability. The two parameters tested were $\log(\lambda)$ and σ^2, which appears in the probability of crossing thresholds metric along with $\log(\lambda)$. Figure 6 shows the results of this analysis for the $\log(\lambda)$ estimates. This analysis involved 30-year time series within a 1920 to 1999 time frame.

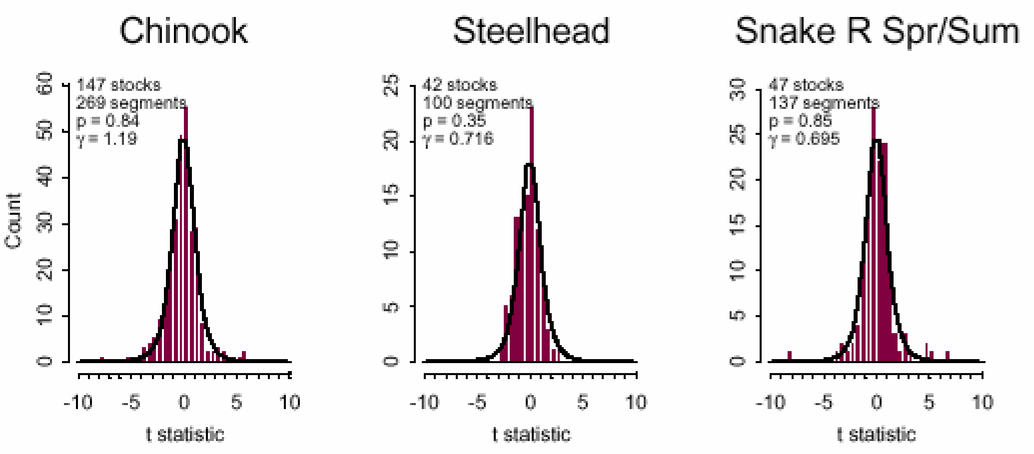

Figure 6. Distribution of actual $\log(\lambda)$ estimates (bars) versus predicted distribution (solid black line from 147, 42 and 47 chinook, steelhead and Snake R spring/summer time series respectively. From Holmes and Fagan (2002).

The close match between the observed and predicted distributions indicates that the $\log(\lambda)$ estimate was properly characterized in terms of its mean value (the peaks match). That is the mean trend in the first half of the time series was the same as the mean trend estimated in the second half of the time series. Figure 6 also demonstrates that the uncertainty in the $\log(\lambda)$ estimate (its variability) was also properly characterized since the width of the distributions match.

That the mean trend in the first 15 years was the same as the mean trend in the second 15 years appears at first glance to contradict the observations of strings of good years versus bad years. However keep in mind that this analysis used 30-year time series across the 1920 to 1999 period. It was asking about the average estimate across different time periods. What about estimates only during a specific time period? Figure 7 shows the difference between the trend in the first 15 years of a time series versus the following 15 years for specific time frames, i.e., not the average across all time periods, but the average if you only look at time series in a specific time period, say 1970-1999. The solid line is a measure of the difference between the trends in the first 15 years versus the following 15 years. Deviations above zero indicate that on average there was a more declining trend in the first 15 years versus the next 15; while deviations below zero means that on average the population was declining less in the first 15 years versus the next 20 years. These results show the average difference from all the West Coast time series put together. What you can see is that across the West Coast, stocks were on average declining more in 1959-1973 versus in 1974-1993 while the opposite was true for 1964-1978 versus 1979-1998.

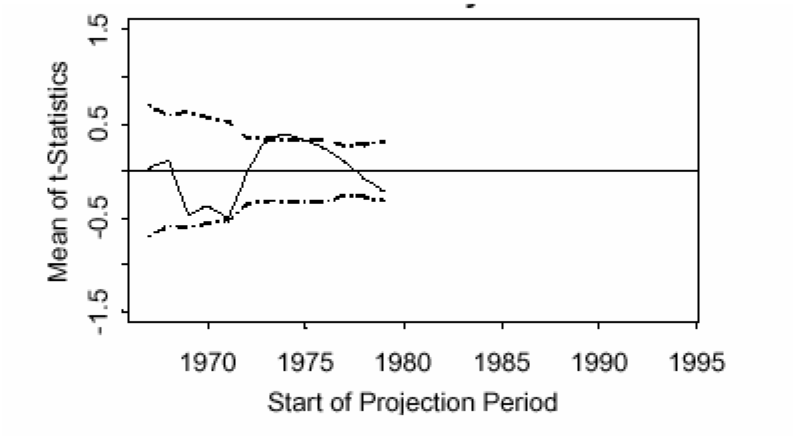

Figure 7. The solid line measures the difference between the average trend in the first 15 years versus in the second 15 years from a collection of 30 year time series of West Coast chinook and steelhead stocks (200+ stocks). "Mean of t-statistics" = 0 means that the average trend (across the whole West Coast) was similar in the first 15 years versus the following 15 years. "Mean of t-statistics" > 0 means that the on average stocks were declining more in the first 15 years relative to the following 15 years. The year on the x-axis denotes the start of the middle of the 30-year segment. The dashed line is the 95% confidence intervals for a random collection of time series, i.e., if there were no underlying environmental cycles causing "good" and "bad" series of years. Holmes unpublished analyses.

It is tempting to attribute these 'good' versus 'bad' strings of years to an environmental driver, such as ocean conditions that one could presumably model. While this may be the case, the data by themselves do not necessarily support this since this type of cycling good and bad strings of years can happen simply by chance in a collection of stochastic population time series. Indeed this is what Figure 2 illustrates. The dotted lines in Figure 7 show the 95% confidence intervals assuming that the time series were all completely independent. This is a conservative estimate since they are not all independent and the true 95% confidence intervals are farther apart. What we can see is that the solid line falls within the conservative 95% confidence intervals suggesting this West Coast pattern of good and bad strings of years is not inconsistent with the hypothesis that it occurred by chance.

The Holmes and Fagan (2002) analysis also looked how well the diffusion approximation predicted the probability of 90% decline. This analysis searched for a difference between the mean diffusion approximation estimates of the probability of 90% decline and the observed mean probabilities within the collection of West Coast salmon time series. Figure 8 shows the estimated versus actual mean probabilities. The gray solid line (Dennis-Holmes) is the method used in the salmon PVA (McClure et al. 2003). The close correspondence between the actual and observed indicates that first the diffusion approximation approach is correctly estimating the mean probabilities and second that the parameters of this approximation were not being systematically misestimated. Note that this analysis focuses on mean estimates of probability of decline. The issue of the variability in estimates of probability of decline is addressed later in this document.

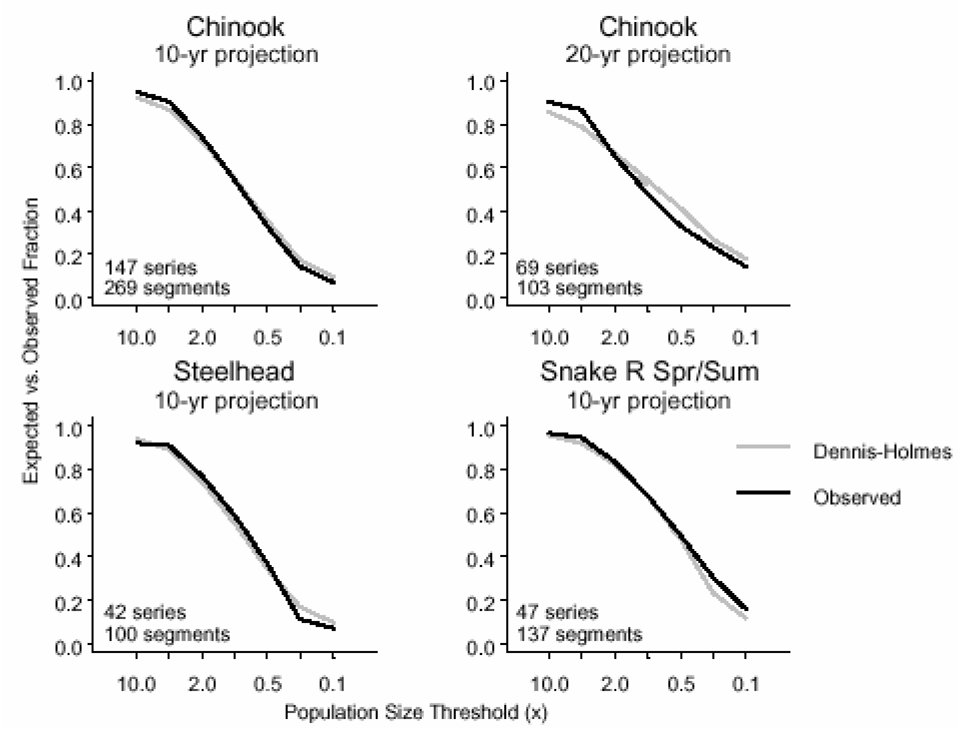

Figure 8. Probability of 90% decline versus observed probabilities with the West Coast salmon time series. From Holmes and Fagan (2002).

Why the diffusion approximation approach versus other approaches for describing trends and risks in salmon populations?

Diffusion approximation approaches for estimation of risk metrics are grounded in theoretical work on stochastic population processes (reviewed in Holmes and Fagan 2002 and Holmes 2004). These methods are one of the basic quantitative tools in population viability analysis and are featured in two current books on quantitative methods for analyzing population data (Lande et al. 2003, Morris and Doak 2003). The long-term rate of population growth is termed λ and is one of the most commonly used risk metrics within the field of conservation biology. Note that λ does not refer to a specific method of estimation, but rather simply the median or long-term trend in the population. There are a variety of methods for estimating λ. The most familiar within the conservation biology literature is to calculate λ from estimated Leslie matrix models. Diffusion approximation approaches present a way to estimate λ when only time series is available, and present a method for estimating the uncertainty in λ, which estimated Leslie matrix models do not provide.

However in the context of salmon management, traditionally other metrics of risk and population trend have been used. Some of the typical metrics that have been used or suggested are log recruits per spawner, SARs, 8-year geometric means of the natural cohort return rate, a simple regression of log natural abundance versus time, and residuals from a stock/recruit relationship. Some of these (log recruits per spawner and a regression of the log abundance versus time) have a close relationship to λ and indeed can be viewed as alternate methods for estimating λ. Many of the other methods, however, differ in a fundamental way in that they measure only a portion of the life cycle, i.e., survivorship or fecundity of only certain stages rather than from spawner to spawner. One of the key aspects of λ is that it integrates across the entire life-cycle. It is not a measure of one stage's survivorship or fecundity alone, but rather of the integration of survivorship and fecundity over the entire life cycle, much like a spawner-to-spawner ratio does. This is important when one is trying to assess a population rather than a particular stage since high survivorship in one stage can easily be offset by low survivorship in another stage.

Below the methods that have been more common in salmon management are discussed in terms of how they relate to λ and the estimation long-term population trends.

Log recruits to the spawning grounds per spawner

Log recruits (to the spawning ground) per spawner is another way to estimate log(λ) since the expected value of ln(R/S) = log(λ). This can be derived from theory on stochastic population processes (see review by Caswell 2001, 14.3.2) and is essentially what is shown by Eqn 14.47 in Caswell (2001) – although this probably will not be transparent on first glance. Obviously the estimates you get of log(λ) from Eq. 5 versus ln(R/S) are going to be different for a specific finite time series; you expect this using different methods even though the expected values (the average estimates) are identical.

If ln(R/S) can be used to estimate log(λ), why not use that since it is more familiar for fisheries biologists? First it is not a more accurate nor less variable estimator – a simple simulation demonstrates this. Second it requires much more data and effort to estimate – despite not providing an increase in precision in the estimation of log(λ). To the extent that the age-at-return data contains errors this adds additional errors to the ln(R/S) estimate. Third, if we want to compare stock status for example to prioritize recovery actions, using a consistent method across all stocks is critical. For the vast majority of stocks, the additional data to estimate R/S is not available so we can't estimate ln(R/S). Fourth, establishing the uncertainty in the estimate of ln(R/S) would be difficult. We would either have to model the error in age-at-return data, which would require some ad hoc assumptions since we have limited information on this error, or we would have to bootstrap from limited age-at-return data. Fifth, we would still have to estimate the process error and estimating this from ln(R/S) data alone is not possible if the population is affected by both process and non-process error.

8-year geometric means of the natural cohort return rate

This metric uses the 8-year geometric mean of the spawner-to-spawner ratio for the natural spawning component of the population. Like ln(R/S), this another way to estimate λ. The reasons for not using this metric are the same as those for not using ln(R/S); see above discussion.

Smolt-to-adult ratios (SARs)

SARs, along with other measures of survivorship, are clearly important for analyzing how survivorship changes within a portion of the salmon life cycle. However this metric leaves out the adult-to-smolt portion of the life cycle. For the purpose of tracking the long-term trends, the entire life cycle, spawner-to-spawner, must be included since increases in smolt-to-adult survivorship could be offset by decreases in adult-to-smolt ratios. Thus, SARs are not used for estimating long-term trends.

Note also that SARs detailed types of data, which are not available for many stocks and makes their analysis regionally limited.

Simple regression of log natural abundance versus time

λ is the regression of log spawner counts versus time for an infinite (i.e. very long) time series. One way to estimate λ is to use the regression of log spawners versus time for the available, finite, time series. This method could have been used, but simulations indicated that it gives estimates that are essentially the same as the runsum estimates (Figure 9). Even if one did use a regression, one needs to use the methods in McClure et al (2003) to get the confidence intervals on λ. The confidence intervals on the regression cannot be used since this attributes all error in the data to observation error. This is incorrect; part of the error is process error and part of it is observation error, and one needs to use a statistical framework that properly apportions the error into these two types. In addition, one still needs to obtain the estimates of environmental variability, which are critical for estimates of the probability of crossing thresholds. The regression will not provide this since again a simple analysis of the variance of the residuals attributes error to observation not process error. Holmes (2004) reviews the currently available methods for parameter estimation for population processes with process and observation error.

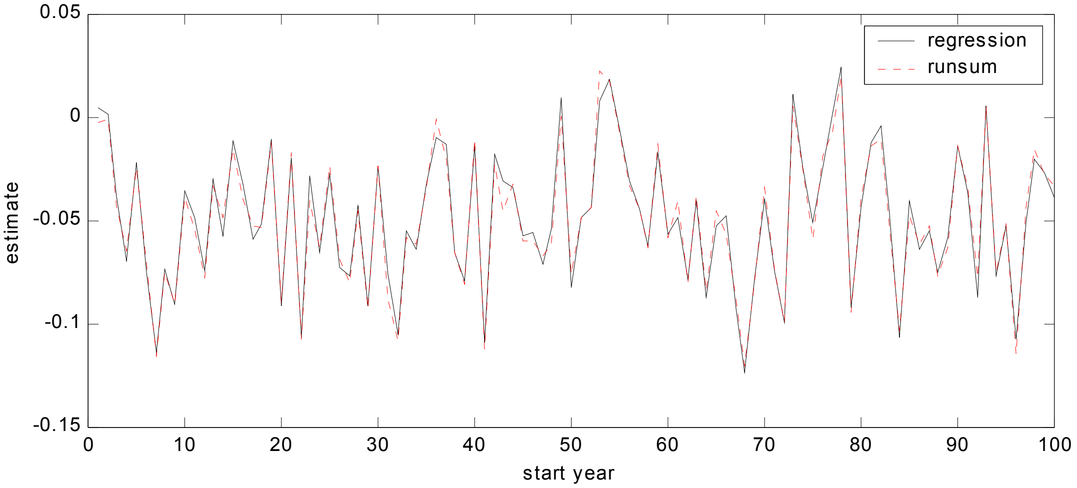

Figure 9. Example of stimated log(λ) from 20-year segments in a time series with $\mu = -0.05$ and $\sigma^2 = 0.02$. Segment 1 is year 1-20, segment 2 is year 2-21, segment 3 is year 3-22, etc. This shows how the estimates vary depending on the segment observed. The regression line (solid black) is from a regression of log counts versus time; the "runsum" method (red dashed) is that used in McClure et al. (2003) and for the Biological Opinion calculations.

Residuals from a stock/recruit relationship

Residuals for a stock/recruit relationship give information on how conditions in one year or cohort deviate from some longer trend. This can be useful for trying to determine if underlying changes for the long term trend has occurred, but is not useful for estimating the long term trend itself. Potentially these residuals could be used to estimate the environmental variability, although this is certainly not straight-forward. The variability in the residuals will be due not only to environmental variability but also variability due to density-dependence and the proclivity of salmon for "boom-bust" cycles. These latter types of variability are important for the short-term variability in population trajectories, but tend to dampen out with time and are less important for the long-term variability in population trajectories. Holmes (2004) gives an example of this using age-structured salmon models with density-dependence. Note also that residuals for a stock/recruit relationship also require age-structure data, which makes their analysis limited to stocks with that kind of detailed data.

Other methods

The methods used in McClure et al (2003) require very simple data, spawner time series, however, there are populations with much better and more detailed data, especially age-at-return and age-specific survivorship data. Incorporation of this data into the estimation of μ can increase the precision of the μ estimate, and consequently the λ estimate. Hinrichsen (2002) discusses estimation of λ using age-at-return information and shows how using this information increases precision although there is no change in bias relative to the λ estimate in McClure et al. (2003). The downside is that the methods in Hinrichsen (2002) are sensitive to high levels of observation error, for example, standard deviation of observation error greater than 0.7, which is certainly seen in redd count data (see discussion in Holmes 2004). More analytical work needs to be done to get around this sensitivity to observation error, but certainly this research suggests that more precise λ estimates can be obtained for those stocks with more extensive data. This is an area that is very promising, however, for regional analyses where we need to compare risks among stocks, some of which are data poor, we will have to continue to have and rely on methods that use only spawner time series for the sake of consistency.

Lindley (2003) presents state-space estimation for noisy time series and offers this as alternative to the estimation methods used in McClure et al. State-space estimation enables maximum-likelihood estimation of μ and σ^2 from noisy data (such as we have for salmon data). It has a strong statistical foundation. I have also been researching state-space estimation and tested Lindley's algorithm in Holmes (2004) and found that it gives much worse estimates of σ^2 than $\hat{\sigma}^2_{slp}$ given the particular characteristics and constraints we face with salmon data. The m estimates were similar to $\hat{\mu}_{run}$, however. I have also investigated a slightly different state-space algorithm for estimation and found similar results. State-space estimation is extremely promising, but a significant amount of research is still need to come up with algorithms that perform more robustly than the current methods in McClure et al.

Summary

While these commonly used metrics are useful for other questions, such as looking for survivorship changes in a particular habitat or life stage or understanding the contributions of particular age classes to recruitment to the population, they are limited in terms of estimating long-term trends, either because they look at just a segment of the population, lead to λ estimates that are more variable than the λ estimates used in the Biological Opinion, or require data that is not available across all populations. Furthermore these other methods do not lead us to an estimation of underlying variability in the population process (process error), which is essential for estimation of the probability of crossing critical population thresholds and for calculating the uncertainty in our risk estimates. The methods used for estimating λ and extinction metrics as described in McClure et al. (2003) have been extensively studied and validated with West Coast salmon time series (Holmes & Fagan 2002) and also salmon-specific simulations which include density dependence (Holmes 2004).

The time frame of ones data and λ estimates

Typically choices must be made about the data, specifically the years, to use to estimate λ. The point estimate of λ will depend on the time frame used, however keep in mind that in general the point estimate of λ should never be used alone since by itself the point estimate does not give an indication of the uncertainty in this estimate. One way present the uncertainty is to use confidence intervals, but confidence intervals are often misleading since they give the erroneous impression that the true value is equally likely within a large interval. Likelihood profiles or posterior probability distributions of λ are much more useful and give a rapid feel for the uncertainty in the estimate of λ. If one uses a posterior probability distribution, it becomes clear that the estimate of λ is not so sensitive to the time frame of the data or the addition of one extra year of data as would appear when only point estimates are presented.

This being said, selection of a reasonable time frame is very important. The following considerations should generally be kept in mind when selecting the time frame to use: a) more data is better, b) the time frame should be representative of historical trends, i.e. not be dominated by 'good' or 'bad' conditions and not dominated by an isolated perturbation and c) for the sake of uniformity and comparison, the time frame should be consistent across stocks. In McClure et al. (2003) the effect of using different time frames for estimation, specifically 1980-2000 versus 1960-2000, on risk metrics for the Columbia River ESUs is shown. The differences were not statistically significant nor in any consistent direction, i.e. for some stocks the 1980-2000 time period gave slightly more severe risk estimates and for others it gave less severe estimates.

From a management standpoint, λ estimates that vary widely depending on the exact starting year of the time series are problematic, and research showing that the estimates are statistically optimal while satisfying does not lessen this practical problem. There are a couple of strategies that I have proposed to deal with this:

1) Use robust estimators of the mean for the μ estimates. Currently in Eqn. 5, a straight mean is used, however a straight mean is highly sensitive to outliers. My preliminary tudies of the effect of different start years on λ estimates using Snake River spring/summer chinook time series indicated that a robust estimator of the mean eliminated much of the problem of λ estimates that vary widely depending on the start year. There are a variety of robust mean estimators; a trimmed mean is the simplest.
2) I examined the 1970s to present data throughout the Columbia River and found that the 1980-present data was affected by an especially unusual series of years between 1978-1982 or so. The estimates using the 1980-present time frame appeared to be more different that one would expect compared to estimates using any other time frame. My initial analysis suggested that 1976-present would generally be a better time frame to use, although this does suffer from dam effects in the early years for some stocks. The 1984-present data could also be used to avoid the 1978-82 period, however, a strong argument can be made that this overly emphasizes a period characterized by bad ocean conditions.

II. ACCOUNTING FOR UNCERTAINTY IN RISK ESTIMATES

A certain amount of variability in estimated parameters and risk metrics is an unavoidable aspect of the analysis of stochastic population processes, simply due to the nature of these processes. One of the strengths of diffusion approximation methods is that the statistical distributions of the estimated parameters are known. As a result, the uncertainty in the estimated risks can be calculated. This is often not the case for other PVA approaches. Even though the uncertainty in diffusion approximation risk metrics can be calculated, this uncertainty is definitely high. In this situation, examining either the likelihood functions or the posterior probability distributions for the risk metrics, rather than simply the point estimates and confidence intervals, will help to clarify the level of data support for different true risk levels. Statistical decision theory (e.g. Berger 1985 is one of many texts on decision theory) provides a framework for integrating estimates of the data support for different risk levels with the consequences of true risk levels. Wade (2000) and Dorazio and Johnson (2003) provide recent discussions of this Bayesian decision framework in conservation biology and resource management contexts.

The idea in a nutshell is to estimate the probability that the risk metric, for example λ, is within particular ranges that are important from a management perspective. For example:

Probability the true λ is in these ranges	< 0.9	0.15
	0.9 – 0.95	0.3
	0.95 – 1.0	0.5
	> 1.0	0.05

Table 1. Estimated probabilities that λ is within different ranges.

These probabilities are estimated using the posterior probability distribution that is estimated from the data. Figure 10 gives an example of the posterior probability distribution for λ estimated from a 38-year times series of spring chinook in the Upper Columbia River basin (data from T. Cooney, NMFS). The probability that λ is within the range *a* to *b* is calculated by integrating the posterior probability distribution between *a* and *b*.

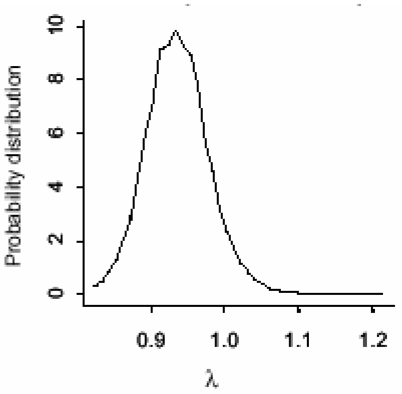

Figure 10. Estimated posterior probability distribution for λ for Upper Columbia spring chinook. From Holmes (2004).

There are a variety of ways these probabilities could be used. They might be used alone and qualitative thresholds set, such as if the probability that λ is less than 0.95 is greater than some threshold, then an such and such actions can (or cannot) occur. Note that the probability of a low λ can be high due to certainty that λ is low or due to uncertainty about λ. Thus, such a strategy leads to caution in the face of high uncertainty. A more quantitative, decision-theoretic, approach can be taken if the probability that actions will be ‘sufficient’ (however that might be defined) can be calculated given different true λs. For example,

		Probability of action being ‘sufficient’		
		Action A	Action B	Action C
λ range	< 0.9	0.3	0.1	0
	0.9 –0.95	0.5	0.3	0.2
	0.95-1.0	0.8	0.6	0.5
	> 1.0	1.0	1.0	0.8

Table 2. Estimated probabilities of action sufficiency given different true λ ranges.

These probabilities are multiplied by the probability of λ being within those ranges and then summed over all ranges to give the total probability that actions are ‘sufficient’. This probability incorporates the uncertainty in the estimated λ:

Probability of action being ‘sufficient’		
Action A	Action B	Action C
0.64	0.45	0.35

Table 3. Probabilities in Table 2 multiplied by those in Table 1 and summed over all λ ranges.

An example where the probabilities in Table 2 would be relatively easy to calculate is different harvest levels. Instead of giving a simple ‘yes/no’ answer as would be the case if using point estimates of λ, this approach quantifies the uncertainty in our estimate of λ and emphasizes that there is not a simple “100% or 0%” probability of an action being effective.

Probabilities of crossing thresholds are notoriously uncertain and variable, and analyzing the uncertainty connected with a proposed probability metric (e.g. ‘probability of extinction’) is especially critical when using these metrics. Figure 11 shows the estimated probability density distributions for the probability of 90% decline within 25, 50 or 100 years given a 20-year time series with an estimated λ of 1 or 0.93. The distributions when the estimated λ is 1 are fairly flat or U-shaped. This indicates that there is not much information about what the probability of 90% decline is. The estimation of the probability of 90% decline can be improved by using an informative prior on the process error. Twenty years of data is not sufficient for accurate process error estimates. If one argues that the variability driving long-term dynamics is similar across chinook throughout the basin, then one might use as an informative prior the distribution of process error estimates for a large number of stocks throughout the basin. Figure 12 shows how the estimation of the probability of 90% decline improves using an informative prior. Now it appears that estimation of the risk of 90% decline in 50 or 100 years is fairly informative for the stock with a low λ. For the stock with a λ equal to 1, 50 and 100-year probabilities are uncertain, but 25-year probability are much better.

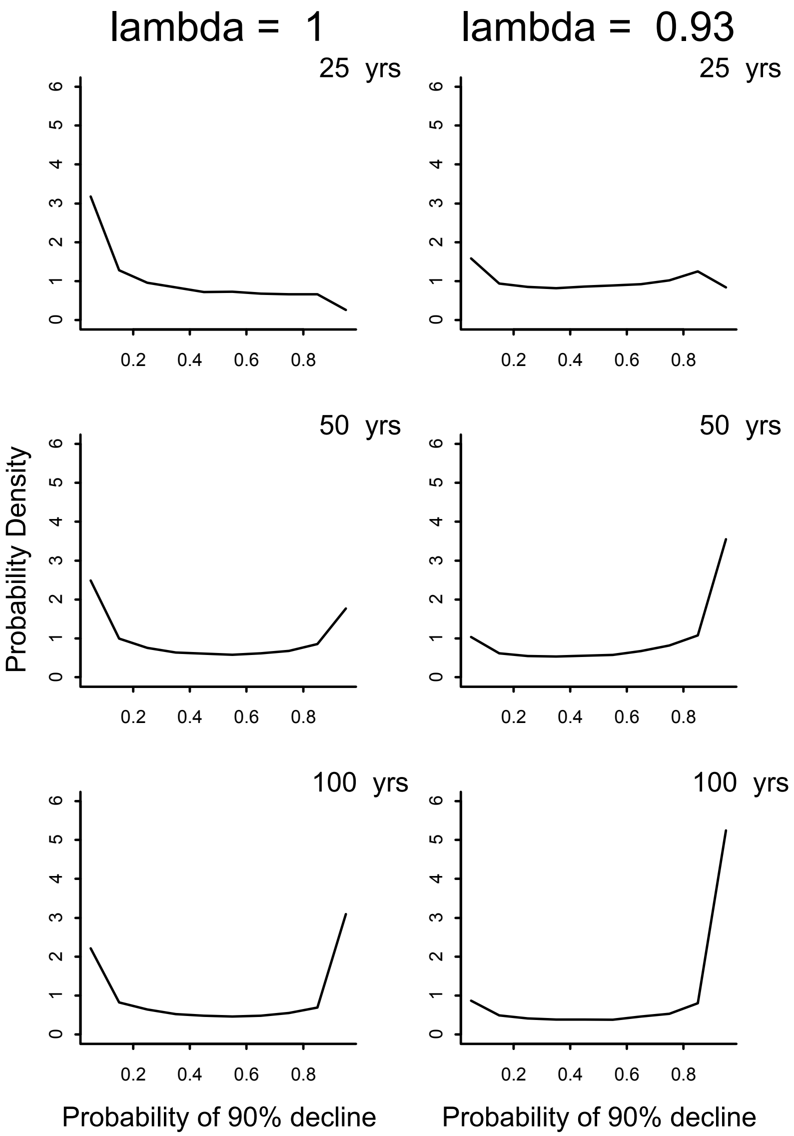

Figure 11. Estimated posterior probability distribution of the probability of 90% decline in 25, 50 and 100 years given a 20 year time series with estimated μ of 0 or –0.072 and an estimated process error of 0.08 and estimated non-process error of 0.71. A uniform prior on the process error was assumed.

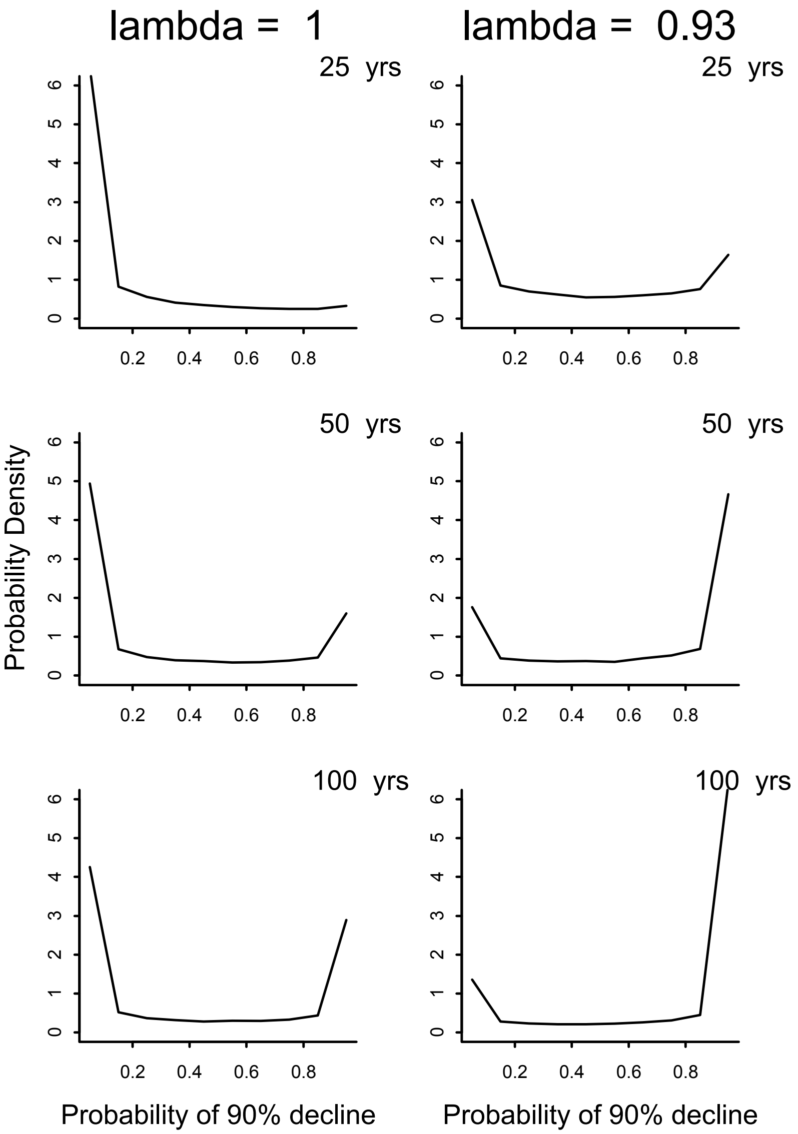

Figure 12. Estimated posterior probability distribution of the probability of 90% decline in 25, 50 and 100 years given a 20 year time series with estimated μ of 0 or –0.072 and an estimated process error of 0.08 and estimated non-process error of 0.71. A highly informative prior on the process error was assumed.

These last two figures focus on the probability of 90% decline in 25, 50 or 100 years. There are other ways to look at the probability of 90% decline that can be more informative. For example, here is an analysis of the probability of an eventual 90% decline for based on the Upper Columbia spring chinook time series (from Holmes 2004). The estimated probability of eventual 90% decline is almost 1.0, that is it is almost certain to occur (Figure 13, top), however there a great deal of uncertainty as to when this will occur (Figure 13, bottom) except that it is highly likely within 100 years. Figure 13, bottom panel, shows the expected probability of 90% decline within a given time frame. From the figure, on average there is a 70% probability of a 90% decline within 50 years for this population and an average 80% probability that the 90% decline occurs within 100 years.

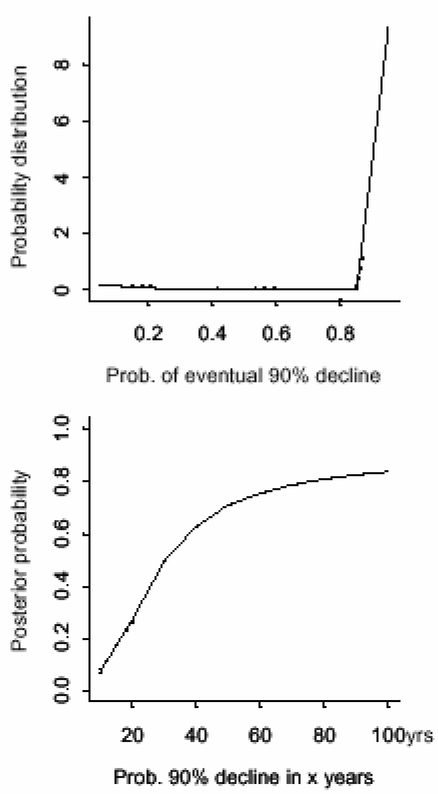

Figure 13. Estimated posterior probability distribution of eventual 90% decline in the Upper Columbia spring/summer chinook (top panel) and the probability that this decline has occurred within different time frames (bottom panel). From Holmes (2004).

The time to 90% decline can then be put into Table form similar to Table 1 for λ ranges (Table 4) which can then be combined with a Table similar to Table 2 for the probability that an action will be sufficient if 90% decline does not occur until after *x* years.

Probability 90% decline only after x years	> 25 yrs	0.7
	> 50 yrs	0.38
	> 75 yrs	0.2
	> 100 yrs	0.18

Table 4. Probability that 90% decline does not occur until after x years. From Figure 13, bottom panel.

Quasi-extinction threshold versus absolute extinction

Throughout this paper, the probability of 90% decline is discussed. The probability of 90% can be estimated for stocks for which we only have index data and not total spawner information. Thus it can be more widely applied. However, decline to specific critical population sizes are also of great importance in PVA analyses. Although estimating extinction to 1 individual is a popular risk metric, and unfortunately sometimes mandated, caution is required when using the diffusion approximation is to estimate extinction to very low numbers since factors that drive dynamics at very low population sizes (such as demographic stochasticity) and the catastrophes often associated with ultimate extinction will likely be poorly represented in a time series of a relatively larger population declining to low numbers. There have been a wide variety of papers published on this in the conservation biology literature. The general recommendation is to estimate the probability of decline to some critical population size (quite a bit greater than 1); this is termed a 'quasi-extinction' threshold. Fagan et al. (unpublished manuscript) studied a collection of actual time series of species that went extinct and compared diffusion approximations for quasi-extinction thresholds versus extinction to 1 individual. This analysis found that quasi-extinction estimates (to a size much greater than 1) using diffusion approximations fit the observed data, but that extinction estimates (to 1 individual) were very poor and underestimate the true risk.

With this in mind, one might ask why was the probability of decline to 1 individual estimated in the McClure et al. analysis. The reasons for this were four-fold. 1) The analysis was focused on estimating risks if current conditions continue as they appeared in the time series data. It was recognized that this would tend to underestimate risks if factors such as density-depensation occurred as the population got small, however baseline estimates of risks under current conditions are required in order to make meaningful statements about risks under hypothetical future scenarios, such as lower population growth rates as the population gets small. 2) The probability of 90% decline does not incorporate the actual population size. The implications of a 90% decline of a population of 10 individuals is quite different than a 90% decline for a population of 100,000 individuals. The probability of decline to 1 individual provided a risk metric that incorporated both the overall rate of decline of the population and the population size. Thus we could then compare ESUs in terms of a risk metric that integrates these two factors –regardless of whether this is an underestimate of the true probability of extinction. 3) Any other extinction threshold we could have specified would have been arbitrary – given the information we had on critical population size. Decline to 1 individual is meaningful for all populations. 4) The Fagan et al. analysis had one notable exception, i.e. one population where the probability of extinction would be properly estimated. That was the one population time series that followed an actual salmonid extinction (sockeye); the rest of the time series followed bird and reptile extinctions.

REFERENCES

Achord, S., P. S. Levin, and R. W. Zabel. 2003. Density-dependent mortality in Pacific salmon: the ghost of impacts past? Ecology Letters 6: 335-342.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer Verlag, NY, NY, USA.

Caswell, H. 2001. Matrix Population Models. Sinauer, Sunderland, Mass.

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

Dorazio, R. M., and F. A. Johnson. 2003. Bayesian inference and decision theory —a framework for decision making in natural resource management. Ecological Applications 13:556–563.

Fagan, W. F., J. Rango, A. Folarin, J. Sorensen, J. Lippe, and N. E. McIntyre. Cross-validation of quasi-extinction risks from real time series: an examination of diffusion approximation methods. Manuscript.

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.

Hinrichsen, R. A. 2002. The accuracy of alternative stochastic growth rate estimates for salmon populations. Canadian Journal Fisheries and Aquatic Sciences 59:1014–1023.

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Science USA 98:5072-77.

Holmes, E. E. and W. F. Fagan. 2002. Validating population viability analysis for corrupted data sets. Ecology 83: 2379-2386.

Lande, R. and S. H. Orzack. 1988. Extinction dynamics of age-structured populations in a fluctuating environment. Proceedings of the National Academy of Science USA 85:7418-7421.

Lande, R., S. Engen, and B. Saether. 2003. Stochastic population models in ecology and conservation: an introduction. Oxford University Press, Oxford, UK.

Lindley, S. T. 2003. Estimation of population growth and extinction parameters from noisy data. Ecological Applications 13: 806-813.

McClure, M. M., E. E. Holmes, B. L. Sanderson, and C. E. Jordan. 2003. A large-scale, multi-species risk assessment: anadromous salmonids in the Columbia River Basin. Ecological Applications 13:964-989.

Morris, W. F. and D. F. Doak. 2002. Quantitative conservation biology: theory and practice of population viability analysis. Sinauer Press, Sunderland, MA.

Morris, W., D. Doak, M. Groom, P. Kareiva, J. Fieberg, L. Gerber, P. Murphy, and D. Thomson. 1999. A practical handbook for population viability analysis. The Nature Conservancy.

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

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

Tuljapurkar, S. D. and S. H. Orzack. 1980. Population dynamics in variable environments. I. Long-run growth rates and extinction. Theoretical Population Biology 18:314-342.

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

Richard Hinrichsen

11 Dec. 2003

Review of white paper "Review of methods, progress and cross-validation studies pertaining to population trend and risk assessment for Columbia River salmonids (November 2003 draft)," and supporting documents by Elizabeth E. Holmes.

Summary

The white paper reviews and discusses methods for estimating population trends and extinction risks and summarizes recent work published by Holmes and others in the peer-reviewed scientific literature. Extensive work was done to validate the diffusion approximation model for salmonid population data subject to potentially high measurement error. My review focuses primarily on methods for strengthening the estimates of stochastic growth rate and dealing with nonstationarity. It also discusses evidence for ocean/climate regime shift effects on salmon population growth rates.

Extinction risk and trend

Extinction risk is estimated based on current estimate of population size, trend, and variability. The 2000 Biological Opinion jeopardy analysis focused on probability of extinction and probability of recovery as the critical metrics, not lambda *per se*. The reviewers should comment in detail on the strengths and weaknesses of using 100-year extinction probability in the jeopardy analysis as opposed to a more reliable measure such as lambda. How accurate is PVA in its ability to predict the future status of wild populations 100 years into the future? Holmes (2001) argues for using lambda as a risk measure because reliance on a risk metric with recalcitrant estimation problems (extinction risk) is hard to justify when an equally useful and more reliable measure is available.

Coulson et al. (2001) pointed out the many pitfalls in population viability analysis. They argue that “PVAs can only be accurate at predicting extinction probabilities if data are extensive and reliable, and if distribution of vial rates between individuals and years can be assumed to be stationary in the future, or if any changes can be accurately predicted.” Salmon populations are known to undergo large nonstationary changes in vital rates due to ocean/climate regime shifts and changes in harvest rates. These are not treated in the current CRI analysis because it assumes a stationary process.

Strengthening lambda estimation

The current approach to estimating and characterizing the uncertainty in lambda is to provide separate estimates of stochastic growth rate and sigma^2. Stochastic growth rate is estimated using a running sum approach that uses the same amount of smoothing on each of the data sets to which it is applied. The sigma^2 estimate is calculated using the the slope method of Holmes. This approach yields confidence intervals based on a t-distribution with degrees of freedom equal to

$$\text{d.f.} = .333 + 0.212*n - 0.387L \text{ (for } n > 15)$$

where *n* is the length of the time series and *L* is the number of counts summed to calculate the running sum (currently 4) (Holmes and Fagan 2002). In the normal i.i.d. case, the degrees of freedom are usually (d.f. = n-1), so it is clear that the current approach presents inefficient estimate of stochastic growth rate in order to reduce the effects of bias. Based on the formula above, in order for the CRI estimates to achieve 20 d.f., 100 spawner observations are needed. In order to achieve just 10 d.f., 53 spawner observations are needed. Using the 22 spawner observations over 1980-2001, which is the current time period for estimating stochastic growth rate, gives us just 4 d.f., which is quite poor. At 4 d.f., the variance of the t distribution is twice as large as it is for 21 d.f. Thus the slope based method comes at a price: a dramatic loss of precision.

To regain some accuracy in stochastic growth rate while also accounting for measurement error, I would recommend the following:
(1) Combine several populations from an ESU to make inferences about stochastic growth rate. Model selection criteria may support using a single stochastic growth rate for several populations.
(2) Do not treat the populations in (1) as independent. Model the covariability so precision estimate is not inflated.
(3) Choose a method that allows the level of smoothing of the spawner series to change with the estimate of measurement error. High measurement error should increase smoothing which low measurement error should reduce smoothing. As it stand the CRI method uses a high level of smoothing (4 year running sum) for all series.
(4) Check to see if common variances are supported by the data.
(5) Allow for the possibility of stochastic growth rate changing due to different harvest or ocean/climate regime shifts. The current models may be misinterpreting dramatic shifts in vitality rates as part of a noise process rather than nonstationarity.

Kalman filter

The Kalman filter approach I presented at the 5 December workshop can incorporate these suggestions naturally. Lindley (2003) recently applied the Kalman filter approach to model a single salmon population. More generally, a Kalman filter approach applied to multiple stocks is

$$\alpha_t = \alpha_{t-1} + \mu + \eta_t, \quad \mathrm{var}(\eta_t) = Q \text{ (state equation)}$$

$$y_t = \alpha_t + \varepsilon_t, \ \mathrm{var}(\varepsilon_t) + H \text{ (measurement equation)}$$

where α_t is a $m \times 1$ vector of states, μ is a $m \times 1$ vector of population-specific or common stochastic growth rates, η_t is a multivariate normal noise process with mean 0 and $m \times m$ variance matrix Q, y_t is a $m \times 1$ vector of log(spawner) observations, ε_t is a multivariate normal error term with mean zero and $m \times m$ variance matrix H.

Smoothing. It may be shown that the stock-specific stochastic growth rate estimates are

$$\hat{\mu} = \frac{a_{T|T} - a_{1|T}}{T-1}$$

where T is the number of yearly observations, and $a_{t|T}$ is the smoothed estimate of the state at time t. It is defined as

$$a_{t|T} = E(\alpha_t \mid y_1, y_2, \ldots, y_T)$$

When measurement error in the data increases, the state estimate is based on greater smoothing of the observations. When measurement error is low, the state estimates track the observations closely. Thus, unlike the running sum approach, the level of smoothing used to estimate the stochastic growth rate depends on the measurement error estimate (Figure 1).

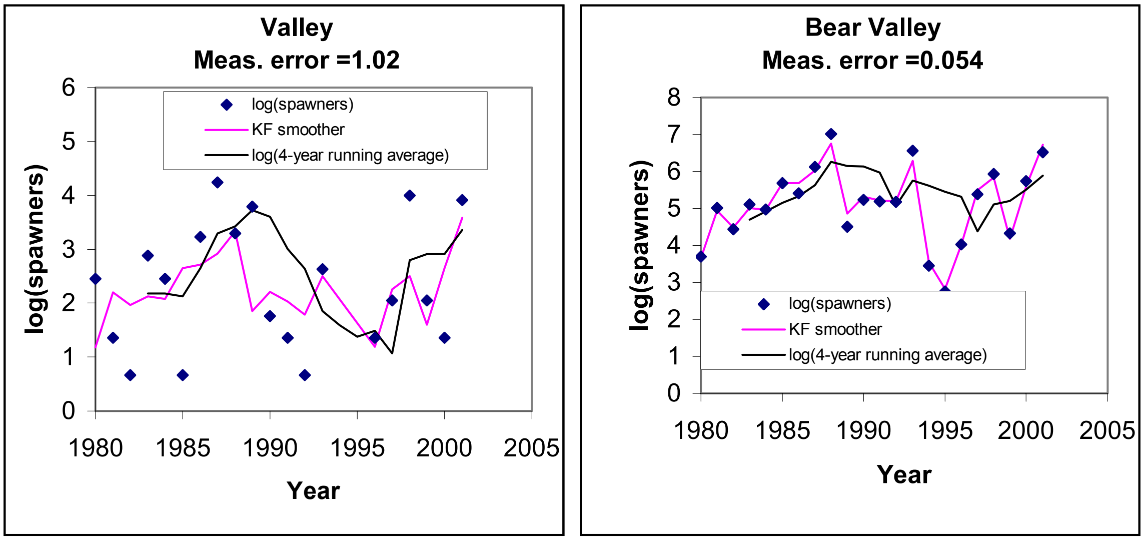

Figure 1. The level of smoothing used by the Kalman filter is greater when measurement error is larger as demonstrated in the above graphs.

<u>Common stochastic growth rate</u>. It is possible to estimate a common stochastic growth rate for the model described above so that all populations have the same stochastic growth rate parameter. This is accomplished by specifying a single parameter in the μ vector. We can then use the likelihood function to calculate AIC and see if the model with a common stochastic growth rate is better supported than the model with stock-specific stochastic growth rates. Preliminary results show that spring/summer chinook index stocks in the Snake River can be combined to increase the accuracy of stochastic growth rate (Table 1).

Table 1. Stochastic growth rate estimates. The model with common stochastic growth rate gives a more precise estimate as indicated by relatively low value of SE (see bolded value). Furthermore, the common stochastic growth rate model has lower AIC. The stock-specific stochastic growth rate model has AIC of 416 while the common stochastic growth rate model has AIC of 407.

Stock	stochastic growth rate	SE	Mean Bias*
Bear Valley and Elk Creek	0.15	0.25	-0.03
Marsh Creek	0.12	0.38	-0.03
Sulphur Creek	0.12	0.42	-0.03
Big Creek	0.12	0.27	-0.02
Lemhi River	0.10	0.19	-0.01
upper Valley Creek	0.11	0.18	-0.03
All (common stochastic growth rate model)	0.19	**0.08**	-0.04

*Mean bias as fraction of SE.

Process error variance bias. The Kalman process error variances tend to be biased downwards, but not to the degree described by Holmes (2003; Figure 3) in the paper "Beyond theory to application and evaluation: diffusion approximations for population viability analysis." At the workshop Holmes indicated that the log(process error variances) are not biased in mean, but are highly biased in median. My analysis shows that the median bias is actual smaller than the mean bias, but that bias is consistently downward. Some bias correction procedure appears to be needed. The largest mean bias was seen in the Lemhi River population which had log(process error variance) that was 50% of standard error (SE).

Table 2. Log(process error variance) estimates from Kalman filter assuming stock-specific stochastic growth rates.

Stock	Log(var estimate)	SE	Mean bias*	Median bias*
Bear Valley and Elk Creek	0.25	0.35	-0.28	-0.25
Marsh Creek	1.07	0.35	-0.28	-0.21
Sulphur Creek	1.28	0.43	-0.30	-0.22
Big Creek	0.37	0.41	-0.27	-0.19
Lemhi River	-0.29	0.48	-0.46	-0.29
upper Valley Creek	-0.44	0.47	-0.21	-0.10

*Bias as fraction of SE.

Correlations. The variance matrix for the process error indicates high correlations among the stocks (Table 3). This suggests that one should not use an estimation procedure that treats the process error of the stocks as independent. Careful consideration of the correlation structure is needed, and the Kalman filter can explicitly model this structure.

Table 3. Process error correlations among stock (Kalman filter estimates). The stock-specific stochastic growth rate model was used.

Bear Valley and Elk Creek	1.00					
Marsh Creek	0.99	1.00				
Sulphur Creek	0.90	0.94	1.00			
Big Creek	0.98	0.95	0.80	1.00		
Lemhi River	0.75	0.78	0.84	0.66	1.00	
upper Valley Creek	0.98	0.97	0.89	0.95	0.86	1.00

Nonstationarity

Because ocean/climate regime shifts and harvest rates can have large effects on stochastic growth rate, it is important to have a modeling framework that can accomodate nonstationarity. One way to deal with this is to explicitly model the nonstationarity using a state space model. One approach would be to rewrite that state equation as

$$\alpha_t = \alpha_{t-1} + \mu_t + \eta_t \text{ (state equation)}$$

Where μ_t is a step function described by two parameters (for each population) that can be estimated from the spawner abundance series. If the shift is thought to have a common effect on all of the stocks, that jump in the step function can be estimated as a parameter common to all stocks. As another possibility, if there are enough stocks in the analysis, it may be possible to model the nonstationarity by allowing μ_t to be determined by a series of parameters estimated for every year in the series. In this case,

$$\mu_t = \mu_0 + \theta_t e$$

where μ_0 is a vector representing the mean stochastic growth rate for the populations, θ_t is a scalar series such that $\sum_{t=1}^{T} \theta_t = 0$, and *e* is a vector of 1s. The various models can be compared based on their AIC scores or some other model selection criteria.

Validation work done by Holmes shows a rather large shift in stochastic growth corresponding to the 1977 regime shift, but indicates that the shift does not reject the null hypothesis of no shift at the 5% significance level. But it must be remembered that changes in harvest scenarios corrresponding to changes in population levels can mask the effects of regime shifts. For example, total fishing for ocean and in-river Snake River fall chinook fisheries was reduced by 30 percent or more from pre-listing rates (See http://www.nwr.noaa.gov/1salmon/salmesa/pubs/harvest2.html). This harvest shift took place during the period used for calculating lambdas in the 2000 BiOp (1980-1999). Furthermore, there was no attempt to quantify the power of the test aimed at detecting a regime shift. Regime shifts may be important, but not detectable using the Holmes estimators in the cross-validation test. The Kalman filter approach gives a natural way to test for a significant shift by directly estimating shifts in the state dynamics. Power to detect important shifts can also be easily quantified, and can be made greater by appropriately combining information from salmon populations in a single model.

Ocean/climate regime shifts

It is becoming increasing apparent that ocean/climate regime shifts play a role in salmon productivity and have a potential to create shifts in stochastic population growth rates that are distinct from a stationary mean noise process. Below, I summarize work on salmon and ocean/climate regime shifts over the last several years.

The National Research Council's Committee on Protection and Management of Pacific Northwest Anadromous Salmonids identified that what happens at sea is important to the conservation and management of Pacific salmon. Interdecadal changes in the ocean environment, especially in water temperature, currents and biological communities, influence the growth and survival rates, and in turn the returns of adults (National Research Council 1996).

The poor ocean conditions for Columbia River salmon over the last 20 years are in part due to the dramatic decrease in food available for young juvenile entering the ocean. Drops in zooplankton and primary productivity, which feed the salmon food chain, showed a dramatic decrease off the West Coast around 1977, creating a relatively barren ocean environment for West Coast juvenile salmon (Hare et al. 1999).

Ocean/climate regime shifts are signaled by large changes in an index called the Pacific Decadal Oscillation (PDO). In Figure 1, the PDO is illustrated. This is an index that characterizes the month-to-month current patterns, and temperature and sea level pressure distributions in the North Pacific. Researchers have found that positive values of the PDO correspond to a warmer North Pacific and negative values correspond to a cooler ocean (Hare et al. 1999, Mantua et al. 1997). Most significant is the cyclic nature of the PDO occurring approximately every 20-30 years during the last 100 years. The PDO and other indicators indicate that the North Pacific Ocean suddenly shifted between the warm and cool regimes four times in the past century. Historically these occurred in 1925, 1947 and 1977. The recent PDO information (http://tao.atmos.washington.edu/pdo/) suggests a shift may have occurred in 1998 with the ocean changing from the warm regime that began in 1977 to a cool regime. Recently, Peterson and Schwing (2003) used synchronized shifts in the PDO, zooplankton volumes, coho salmon survival, and biomass anomalies of cold-water copepod species as evidence of a 1998 shift.

The regime after the shift in 1947 favored Columbia River salmon populations, but the shifts after 1925 and 1977 were found to be unfavorable to Columbia River salmon.

Besides characterizing the ocean environment, the PDO is also related to streamflow, which affects the ecosystem of salmon in rivers. Streamflow records from British Columbia and Washington State indicate relatively dry conditions during the

1977-1997 period, Mantua et al. (1997) conclude that, "[t]o the extent that high streamflows favor high survival of juvenile salmon, PDO-related streamflow variations are likely working in concert with the changes to the near-shore marine environment in regard to impacts on salmon production."

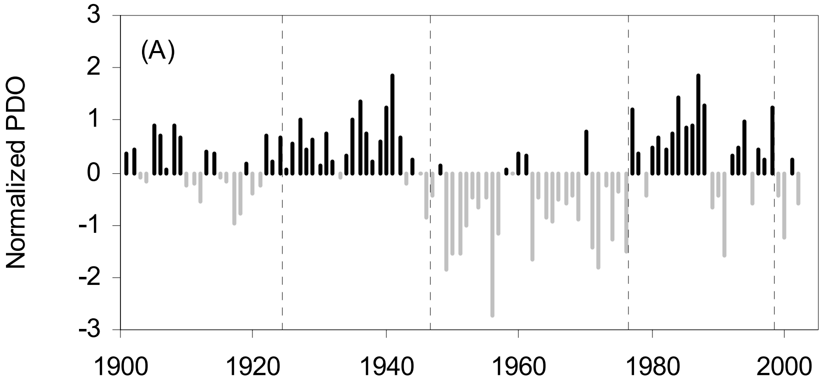

Figure 1. Winter values of the Pacific Decadal Oscillation, an index of sea surface temperature and temperature characterizes the state of the surface waters of North Pacific Ocean. The PDO index shows that the ocean cycles between a warm regime, identified by a positive PDO, and a cool regime, in which the PDO is negative.

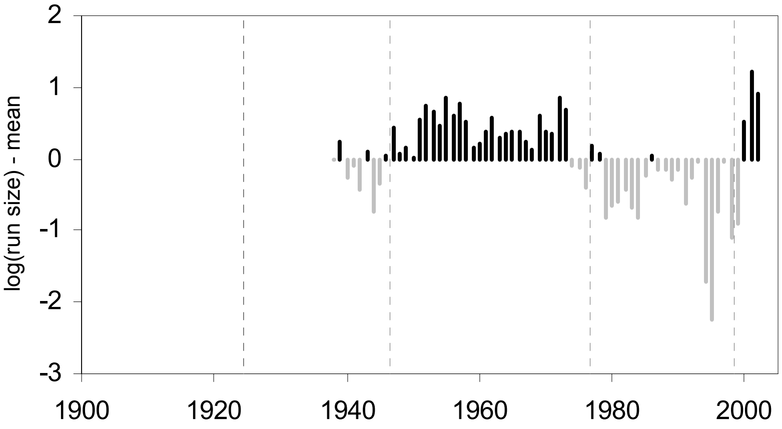

Figure 2. Depicted is log(run size) with the mean subtracted, where run size is the estimated number of upriver spring chinook (those that spawn above Bonneville Dam) arriving at the mouth of the Columbia. Data are available from Oregon Department of Fish and Wildlife, (http://www.dfw.state.or.us/odfwhtml/infocntrfish/interfish/crm.html). Positive value bars are black, negative are gray. Dotted vertical lines are drawn to mark the PDO polarity reversal times in 1925, 1947, 1977, and 1998.

Figure 2 illustrates how the abundance of upriver spring chinook has shifted with the PDO. Abundance is expressed as the estimated run size of spring chinook destined to areas upstream of Bonneville Dam. This run includes stocks from the Snake River spring/summer, the Upper Columbia River spring-run chinook ESUs. The data were updated versions of data in Oregon Department of Fish and Wildlife and Washington Department of Fish and Wildlife report (ODFW and WDFW 1999). Notice that run size was above average in the cool ocean regime, below average after the 1977 PDO shift, but returned to above average after the 1998 PDO shift.

The background level used in the 2000 FCRPS BiOP was used to project salmon populations over a century was derived from a period of poor ocean conditions for Columbia River salmon stocks (years 1980 through 1997) (Hare et al. 1999). For most of the past 20 years, ocean/climate conditions have been unfavorable to Columbia River salmon production. Assuming the base period's low background survival rate continues into the future overestimates the risk of stock extinction. Recent return rates are well above average, and some have hypothesized that ocean conditions may be shifting to a more favorable regime.

The scientific literature contains ample information from peer-reviewed scientific papers that the poor ocean conditions of the 1980s and 1990s are not permanent, and that the climate and ocean experience regime shifts which have large impacts on salmon production (Mantua et al. 1997; Hare et al. 1999). Finney et al. (2000) used sediment records to show that salmon abundance in Bristol Bay and Kodiak Island regions of Alaska have showed market shifts over the past 300 years, and that some pronounced changes appear to be related to climatic changes.

References

Coulson, T., Mace, G.M, Hodson, E., and Possingham, H. 2001. The use and abuse of population viability analysis. Trends in Ecology & Evolution. 16(5): 219-221.

Finney, B.P., Gregory-Eaves, I. Sweetman, J., M.S.V. Douglas, and J.P. Smol. 2000. Impacts of climate change and fishing on Pacific salmon abundance over the past 300 years. Science, 290:795-798.

Hare, S.R., N.J. Mantua, and R.C. Francis. 1999. Inverse Production Regimes: Alaska and West Coast Pacific Salmon. Fisheries, 24(1):6-14.

Holmes, E.E. 2001. Estimating risk in declining populations with poor data. PNAS 98 (9): 5072-5077.

Holmes and Fagan. 2000. Validating population viability analysis for corrupted data sets. Ecology 83(9):2379-2386.

Lindley, S.T. 2003. Estimation of population growth and extinction parameters from noisy data. Ecological Applications 13(3):806-813.

Mantua, N. J., S. R. Hare, Y. Zhang. 1997. A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. Bulletin of the American Meteorological Society, 78:1069-1079.

National Research Council. 1996. Upstream: Salmon and society in the Pacific Northwest. National Academy Press, Washington, D.C.

ODFW and WDFW. 1999. Status Report. Columbia River Fish Runs and Fisheries, 1938-98. Oregon Department of Fish and Wildlife, Clackamas, Oregon.

Peterson, W.T. and Schwing, F.B. 2003. A new climate regime shift in the northeastern pacific ecosystems. Geophysical Research Letters, 30(17), 1896, doi:10.1029/2003GL017528.

Charlie Paulsen

Dec. 12, 2003
cpaulsen@paulsenenvironmentalresearch.com
503-699-4115

Extinction predictions

Extinction predictions assume two things: that the future will be like the past (same mu and sigma) and that extinction will be caused by a continuation of past long-term trends, and not by novel or anomalous events not reflected in past data.

> Since the estimated distribution of $\ln(N_{t+1}/N_t)$ is used to forecast declines, if the extinction event is anomalous, extinction itself will not be predicted based on the information in the estimation period. (Fagan et al p. 16)
>
> This suggests some kind of anomalous final collapse. Though disconcerting from the perspective of population prediction, such sudden, catastrophic events do mirror our conceptual understanding of the extinction process (Fagan et al p. 20)
>
> In contrast to the successes we observed, the diffusion approximation model's most striking failure was its inability to predict actual extinction events. Examining the populations that actually went extinct, we found that both parameterization techniques (Dennis and Dennis- Holmes) did a distinctly poor job of signaling the real vulnerability of the populations… (Fagan et al p. 20)

Extinction, under the DA model, can only come about as the result of long-term, gradual declines (e.g., birth rate < replacement rate over many years, as in Italy or Germany). In fact, it appears that most extinctions are the result of catastrophes that are not well-predicted by DA methods. For salmon stocks in question, the only recent (past several decades years), well-documented extinctions have occurred due to stocks' exclusion from large spawning areas (upper Clearwater, above Grand Coulee and above Hells Canyon), and from deliberate poisoning with rotenone (of sockeye in Idaho lakes). None of these would have been predicted by DA methods alone.

Conclusion: either do not try to predict extinction, or place very large caveats around any such predictions.

Measurement Error

The Dennis-Holmes method differs from the Dennis method in that it assumes that some variation in abundance is due to measurement error rather than process error. Clearly, measurement error may be problematic. However, the examples cited (Dunham & Reiman 2000, Jones et al 1998, Hilborn et al 1999) are of very limited relevance for the stocks in question. The first deals with bull trout, a smaller, non-anadromous species which builds redds that are much smaller than those of chinook, steelhead, and coho. The other two address problems with live, transect-style counts (rather than redd or dam counts) of pink salmon spawning aggregates, a physically smaller and far more abundant species than listed salmon and steelhead. The Holmes November 2003 review (p. 9) notes that "[t]he models also include sampling error in the range of that observed for redd-count data (standard error 0.3 to 0.85)", but I can find no cite to sources for these estimates (they do not, so far as I can tell, appear in Achord et al 2003).

Conclusion: if examples of measurement error can indeed be found for listed stocks, or at least for the same species of salmonids, it would be very useful to see these in detail in the DA methods review. If not, appropriate disclaimers should be attached to the analyses, to the effect that while measurement error is assumed in the Dennis-Holmes method, we have no empirical estimates of it for the listed species.

Stationarity

A substantial and growing body of literature strongly suggests that cycles in salmonid survival – most likely due to climate effects - are the norm rather than the exception. Furthermore, these climate changes are typically considered to be large, step-function shifts, as opposed to simple autoregressive processes. If this is also the case for listed stocked, this would obviously violate the stationarity assumption of DA methods.

A closely related body of research suggests that one can expect that salmon stocks of the same species, rearing in close proximity, to exhibit strong correlations in abundance, recruits per spawner (R/S), and life-stage survival rates. These correlations are of technical interest because they enable one to use information from large collections of stocks to estimate model parameters. They are also of interest from a management perspective as well, since they strongly imply that the assumption of stock independence made in DA's to date do not hold in practice. Taken together, these would suggest that the failure to discover these relationships with DA methods (e.g., Holmes, "Review of methods …", Figure 7) are limitations of the DA method, and not indications that regime shifts and correlations among stocks do not exist.

Nearly all of these publications either compare multiple stocks with similar spawning locations (e.g., Snake River spring/summer chinook) or aggregates (e.g., all spring chinook spawning above Bonneville Dam), rather than the stock-by-stock approach taken for the majority of DA analyses. In addition, no other analysis that I am aware of that looked for climate effects or multi-stock correlations grouped different species and assumed similar responses to climate, as is the case in Figure 7 (steelhead and chinook). See next two sections for more details and conclusions.

Climate effects

A growing body of literature shows that sharp, step-wise changes in climate are common wherever long-term series of climate indices are available (e.g., Ware, 1995, Beamish et al 1999, Hare and Francis 1994, Downton and Miller 1998). The references merely scratch the surface of a rapidly growing field of study.

A number of recent publications strongly suggest that climate-induced changes in survival rates, R/S, and/or harvest follow step functions. These include Adkinson et al 1996 (BC sockeye R/S), Beamish et al 1999 (BC coho catch), Bradford and Irvine (BC coho R/S), Downton and Miller 1998 (sockeye, pink, and chum catch), Koslow et al 2002 (coho marine survival), Mantua et al 1997 (catch, various spp.), Mueter et al 2002 (*Oncorhynchus* spp. R/S), and Ware and Thompson 1991 (sardine and other spp. catch). All of these publications use either multi-stock R/S approaches (e.g., Adkinson et al 1996, Mueter et al 2002) or use catch aggregated across multiple stocks. To my knowledge, no one to date has detected climate step-function effects in single stocks – the data are simply too noisy to permit such analyses.

Conclusions: where data permit it (long time series and R/S having much higher power than short abundance time series) do systematic checks for step functions. Especially for longer time series, harvest estimates should be used to adjust spawning abundance, since rates were much higher pre-1980, the cut-off for most time series in the BiOp. Power and sensitivity analyses for the DA approach would also be useful, to se if climate effects are detectable (I'm guessing power will be very low) and how sensitive the DA approach is (guessing not very, due to very high variance).

Multi-stock approaches

Fortunately, there are many techniques available to model correlations or covariance in stock-specific spawner abundance, R/S, and life-stage survival rates. Botsford and Paulsen (2000) reviews methods for cross-stock correlations on abundance and R/S, and applies these to a number of Columbia basin spring/summer chinook stocks. Paulsen and Hinrichsen (2002) apply R/S regression methods to Snake River spring/summer chinook. Peterman et al (2000) apply Kalman filters to Pacific salmon, while Pyper et al (2002) apply regression methods to chum salmon, and Pyper et al (2001) apply similar methods to pink salmon. Paulsen and Fisher (2001) and Paulsen and Fisher (2003) use regression models applied to multi-stock estimates of juvenile survival. All R/S methods and Paulsen and Fisher (2003) find density-dependent effects in the vast majority of stocks, as well as strong covariation in indicators (abundance, R/S, or survival).

Note that only Botsford and Paulsen (2000) use abundance in their analyses. This suggests that information on age-at-return (for R/S) or life-stage survival rates will be needed for multi-stock modeling, although Hinrichsen's presentation at the lambda workshop used Kalman filters on abundance data to good effect.

Conclusion: multi-stock approaches have done very well at detecting effects that are too subtle for single-stock approaches. Where data permit, different analyses using abundance, R/S, and other metrics over time is recommended to compare stocks and investigate common effects due to climate, hydropower, etc.

References

Adkison, M. D., Peterman, R. M., Lapointe, M. F., Gillis, D. M., and J. Korman. 1996. Alternative models of climatic effects on sockeye salmon, *Oncorhynchus nerka*, productivity in Bristol Bay, Alaska, and the Fraser River, British Columbia. Fish Oceanogr. **5**: 137-152.

Adkison, M. D. and R. M. Peterman. 2000. Predictability of Bristol Bay, Alaska, sockeye salmon returns one to four years in the future. North American Journal of Fisheries Management **20**: 69-80.

Beamish, R. J., D. J. Noakes, G. A. McFarlane, L. Klyashtorin, V. V. Ivanov, and V. Kurahov. 1999. The regime concept and natural trends in the production of Pacific salmon. Canadian Journal of Fisheries and Aquatic Sciences **56**: 516-526.

Beamish, R. J., G. A. McFarlane, and R. E. Thomson. 1999. Recent declines in the recreational catch of coho salmon (*Oncorhynchus Kisutch*) in the Strait of Georgia are related to climate. Canadian Journal of Fisheries and Aquatic Sciences **56**: 506-515.

Botsford, L. W. and Paulsen, C. M. 2000. Assessing covariability among populations in the presence of intraseries correlation: Columbia River spring-summer chinook salmon (*Oncorhynchus tshawytscha)* stocks. Can. J. Fish. Aquat. Sci. **57**:616-627.

Bradford, M. J. and J. R. Irvine. 2000. Land use, fishing, climate change, and the decline of Thompson River, British Columbia, coho salmon. CJFAS **57**: 13-16.

Downton, M. W. and K. A. Miller 1998. Relationships between Alaskan salmon catch and North Pacific climate on interannual and interdecadal time scales. CJFAS **55**: 2255-2265.

Hare, S. R. and R. C. Francis. 1994. Climate change and salmon production in the northeast Pacific Ocean, p. 357-372. In R.J. Beamish [ed.] Climate Change and Northern Fish Populations. CJFAS **121**.

Koslow, J. A., Hobday, A. J., and G. W. Boehlert. 2002. Climate variability and marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon production area. Fisheries Oceanography **11**(2): 65-77.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis. 1997. A pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society **78**(6): 1069-1079.

Mueter, F. J., Peterman, R. M., and B. J. Pyper. 2002. Opposite effects of ocean temperature on survival rates of 120 stocks of Pacific salmon (*Oncorhynchus* spp.) in northern and southern areas. CJFAS **59**: 456-463.

Paulsen, C.M. and T.R. Fisher. 2001. Statistical relationship between parr-to-smolt survival of Snake River spring-summer chinook salmon and indices of land use. Transactions of the American Fisheries Society 130: 347-358.

Paulsen, C. M. and R. A. Hinrichsen, 2002. Experimental management for Snake River spring/summer chinook (*Oncorhynchus tshawytscha*): trade-offs between conservation and learning for a threatened species. CJFAS **59**: 717-725

Paulsen, C.M. and Fisher, T.R. 2003. Detecting juvenile survival effects of habitat actions: power analysis applied to endangered Snake River spring/summer chinook (*Oncorhynchus tshawytscha*). Can. J. Fish. Aquat. Sci. 60: 1122-1132

Peterman, R. M., B. J. Pyper, and J. A. Grout. 2000. Comparison of parameter estimation methods for detecting climate-induced changes in productivity of Pacific salmon (*Oncorhynchus spp.*). CJFAS **57**: 181-191.

Pyper, B. J., F. J. Mueter, and R. M. Peterman. 2002. Spatial covariation in survival rates of Northeast Pacific Chum Salmon. Trans. Am. Fish. Soc. **131**: 343-363

Pyper, B. J. et al 2001. Spatial covariation in survival rates of Northeast Pacific pink salmon (Oncorhynchus gorbuscha). CJFAS **58**: 1501-1515

Ware, D. M. 1995. A century and a half of change in the climate of the NE Pacific. Fisheries Oceanography **4**(4): 267-277.

Ware, D. M. and R. E. Thomson. 1991. Link between long-term variability in upwelling and fish production in the northeast Pacific Ocean. CJFAS **48**: 2296-2306.

Charlie Paulsen

The following comments were submitted by Dr. Kris Ryding of WDFW after the workshop on April 26, 2004. These comments are included here for the record but the reviewers Drs. Deutschman and Heppell did not have an opportunity to read these before they prepared their report. This memo was based on a review of the white paper, and many of the issues brought up in this memo are directly addressed in the original papers and their appendices: Holmes (2001), Holmes and Fagan (2002), and Holmes (2004).

TECHNICAL MEMO

TO: POPULATION TRENDS REMAND GROUP

FROM: KRIS RYDING, FISHERIES BIOMETRICIAN, WDFW

SUBJECT: POPULAITON TREND ASSESSMENT

DATE: 10/5/2004

Comments on the estimation of population trends presented at the March 23rd meeting, and presented in the paper ‘Review of methods, progress and cross-validation studies pertaining to population trend and risk assessment for Columbia River salmonids” (Holmes 2003). are divided into four areas. The first section concerns the separation of process and sampling error based on the conditional variance formula, and is a follow-up to the discussion on estimation of process error from regression from the March 23rd meeting. The second area concerns the distribution of the running sum. Third set of comments includes a general variance expression the finite rate of increase as estimated by $\ln\left(R_{t+\tau}/R_t\right)$, written in terms of spawner abundance and incorporating observational error. The fourth area involves a general discussion of parameter estimation methods. Suggestions for the analysis are presented last.

1.) ***Process error estimation****.*

The empirical variance of spawners can be written in terms of process and sampling error using the conditional variance formula as follows,

$$Var\left(\hat{S}_p\right) = Var\left(E\left(\hat{S}_p \middle| S_p\right)\right) + E\left(Var\left(\hat{S}_p \middle| S_p\right)\right),$$

where $Var\left(\hat{S}_p\right)$ = the empirical variance among all spawners,

$E\left(Var\left(\hat{S}_p \middle| S_p\right)\right)$ = the average measurement error and,

$Var\left(E\left(\hat{S}_p \middle| S_p\right)\right)$ = the process error.

If the sampling process for estimating spawners is known, then the average measurement error, $E\left(Var\left(\hat{S}_p \middle| S_p\right)\right)$, and subsequently process error, can be calculated.

The process error can also be estimated from the mean square error of a regression model. Recall that the empirical variance or total variance of the dependent variable is the sum of the regression variance and mean square error. Hence,

$$\begin{aligned}
Var\left(\hat{S}_p\right) &= MSR + MSE \\
Var\left(\hat{S}_p\right) &= \left(MSR + \sigma_\varepsilon^2\right) \\
Var\left(\hat{S}_p\right) &= MSR + \left(Var\left(E\left(\hat{S}_p \middle| S_p\right)\right) + E\left(Var\left(\hat{S}_p \middle| S_p\right)\right)\right) \\
Var\left(\hat{S}_p\right) &= MSR + \left(Var\left(\overline{\hat{S}_p \middle| S_p}\right) + \overline{Var\left(\hat{S}_p \middle| S_p\right)}\right)
\end{aligned}$$

where MSR = the total error from the regression,

MSE = the residual error,

$Var\left(\overline{\hat{S}_p \middle| S_p}\right)$ = natural variability (process error) which incorporates demographic stochastisity and,

$\overline{Var\left(\hat{S}_p \middle| S_p\right)}$ = average measurement error.

Consequently, variability in spawner abundances that are a result of time or environmental covariates can be accounted for using regression analysis through the mean squares regression term (MSR).

Kris Ryding

2.) Distribution of the running sum.

Estimation of the finite rate of increase, or population trend, is based on a running sum of spawner data, introduced in Holmes (2001), Holmes and Fagan (2002) and McClure et al (2003) and expressed as follows,

$$R_t = \sum_{j=1}^{4} S_{t+j-1} = S_t + S_{t+1} + S_{t+2} + S_{t+3},$$

where S_t = spawer abundance at time *t,*

Critical to the use of the diffusion approximation method is the lognormality of R_t. The distribution of an estimate depends on the sampling method. The usual assumption is that abundances are lognormally distributed, i.e., $\ln(S_t) \sim N(\mu, \sigma^2)$. The only way for $\ln(R_t)$ to be normally distributed from the above equation is for it to be the result of $\sum_{j=1}^{4} \ln(S_{t+j-1})$. The derivation of a lognormal distribution for R_t from the sum of spawner abundance is unclear. There is no well-defined distribution for $\ln\left(\sum_{j=1}^{4}(S_{t+j-1})\right)$, and subsequently, the theoretical properties are unknown.

McClure et al. (2003) provide the only stated arguments for the lognormality of $\ln(R_t)$. The bases of the arguments seem to arise from the presence or absence of significant outliers. Unbiased, random sampling of a target population can produce outliers regardless of the distribution. A lack of outliers does not constitute proof of normality. The most defensible arguments for normality arise from asymptotic (large sample) theory, which does not apply in this case. The other assumptions tested in McClure et al. (2003) are relevant to the analysis stochastic processes and can be violated even if the data are normally distributed.

The stochastic exponential process is presented in Eq. 2, pg 2.7 as,

$$N_t = N_0 e^{\left(\mu t + \varepsilon_p\right)}, \text{ where } \varepsilon_p \sim Normal\left(0, \sigma^2 t\right).$$

It is not clear in any manuscript or paper whether the spawner counts, S_t or running sum, R_t is substituted for N_t; however, all indications are that it is S_t.

3.) Independence between between $R_{t+\tau}$ and R_t, the variance of $\ln\left(\frac{R_{t+\tau}}{R_t}\right)$ and sources of error.

Two of the issues of concern are the independence between $R_{t+\tau}$ and R_t and the bias in estimating process error when sampling variability is high. Variances expressed analytically are invaluable in assessing how different assumptions and processes contribute to overall variability. These relationships can be hidden in simulation studies. Presented below is a countering the argument for independence between $R_{t+\tau}$ and R_t followed by two variance expressions for $\ln\left(R_{t+\tau}/R_t\right)$. The first variance estimator is an analytic expression for the regression method presented in McClure (2003). The second estimator of $Var\left(\ln\left(\frac{R_{t+1}}{R_t}\right)\right)$ is a more general expression for the variance of a function of two random variables. The second expression is then modified for the special case of τ = 1 and independence among spawner counts to illustrate how the addition of error to spawner counts can affect overall variability.

Independence between $R_{t+\tau}$ and R_t requires that each running sum is estimated from different spawner estimates *and* that the abundances, S_t, are also pairwise independent, i.e., $Cov\left(\hat{S}_{t+\tau+i}, \hat{S}_{t+j}\right) = 0$ for all *i, j = 0, 1, 2,3.* When τ = 1, 2, or 3, the first condition of independent data does not hold. For example, when τ = 1, then the values of $R_{t+\tau}$ and R_t calculated as,

$$R_{t+1} = \sum_{j=0}^{3} S_{t+1+j-1} = S_{t+1} + S_{t+2} + S_{t+3} + S_{t+4} \text{ and,}$$

$$R_t = \sum_{j=0}^{3} S_{t+j-1} = S_t + S_{t+1} + S_{t+2} + S_{t+3}$$

Common to each R is the sum $S_{t+1} + S_{t+2} + S_{t+3}$. Hence $R_{t+\tau}$ and R_t are not statistically independent when $\tau = 1$, 2, or 3. When $\tau = 4$, then spawner counts must be pairwise independent. Independence among spawner estimates is not addressed.

Estimation of the variance of $\widehat{\ln}(\lambda)$, i.e., $Var(\hat{\mu}_{run})$ is unclear in the manuscript, although it is stated as the estimated slope of a regression through the origin of $\widehat{Var}\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right)\right)$ versus τ for $t = 1,2,3,\ldots,k-3$, $\tau = 1,2,3,4$. Based on descriptions in previous work, the least squares estimate of the slope, i.e., the variance of $\widehat{\ln}(\lambda)$, can be expressed analytically as follows,

$$\widehat{Var}(\hat{\mu}_{run}) = \hat{\sigma}^2_{slp} = \frac{\sum_{\tau=1}^{4} \widehat{Var}\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right)\right)\cdot\tau}{\sum_{\tau=1}^{4}\tau^2},$$

where $\widehat{Var}\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right)\right) = \frac{\sum_{t=1}^{k-3}\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right) - \left(\overline{\ln\left(\frac{R_{t+\tau}}{R_t}\right)}\right)\right)^2}{k-3}$, the empirical variance of $\ln(R_{t+\tau}/R_t)$ across all t for each $\tau = 1,2,3,4$.

The manuscript mentions estimating process error from $\hat{\sigma}^2_{slp}$, yet is it unclear as to how the estimation process happens. Consider the analytical expression for the variance of $\ln(R_{t+\tau}/R_t)$, which can be written as follows

$$Var\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right)\right) = Var(\ln(R_{t+\tau})) + Var(\ln(R_t)) - 2Cov(\ln(R_{t+\tau}),\ln(R_t)),$$

In simulation studies, the observation error is added to the spawners. Expressed in terms of spawner abundance the variance is as follows,

$$Var\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right)\right) = Var\left(\ln\left(S_{t+\tau} + S_{t+\tau+1} + S_{t+\tau+2} + S_{t+\tau+3}\right)\right) + Var\left(\ln\left(S_t + S_{t+1} + S_{t+2} + S_{t+3}\right)\right)$$

$$-2Cov\left(\ln\left(S_{t+\tau} + S_{t+\tau+1} + S_{t+\tau+2} + S_{t+\tau+3}\right), \ln\left(S_t + S_{t+1} + S_{t+2} + S_{t+3}\right)\right),$$

Assuming that S_t can be substituted into the stochastic exponential process $N_t = N_0 e^{(\mu t + \varepsilon_p)}$, adding measurement error (σ_m^2) to the process error, and ignoring the covariance for simplification, the variance of $\ln(R_{t+\tau}/R_t)$ can be expressed as,

$$Var\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right)\right) = Var\left(\ln\left(\sum_{i=0}^{3} S_0 e^{(\mu(t+\tau+i)+\varepsilon_{t+\tau+1})}\right)\right)$$

$$+ Var\left(\ln\left(\sum_{i=0}^{3} S_0 e^{(\mu t+\varepsilon_{t+i})}\right)\right)$$

where $\varepsilon \sim Normal\left(0, \left(\sigma_p{}^2 + \sigma_m^2\right)t\right)$. As modeled in the above equation, measurement error is multiplicative, propagating through the time series. Alternatively, the measurement can be additive, expressed as,

$$Var\left(\ln\left(\frac{R_{t+\tau}}{R_t}\right)\right) = Var\left(\ln\left(\sum_{i=0}^{3} S_0 e^{\left(\mu(t+\tau+i)+\varepsilon_{p(t+\tau+i)}\right)} + \varepsilon_{m(t+\tau+i)}\right)\right)$$

$$+ Var\left(\ln\left(\sum_{i=0}^{3} S_0 e^{\left(\mu(t+i)+\varepsilon_{p(t+i)}\right)} + \varepsilon_{m(t+i)}\right)\right)$$

where $\varepsilon_m \sim Normal\left(0, \sigma_m^2 t\right)$.

Considering the above variance expressions, how are the measurement errors incorporated and process error estimated?

4.) Parameter estimation

Parameter estimation is not a trivial matter. Pages 2.18 - 2.19 of Final Report of the Technical Workshop on Population Trends and Extinction Analysis explicitly states, "One of the strengths of diffusion approximation methods is that the statistical distribution of the estimated parameters are known." Both frequentist and Bayesian methods rely on the likelihood, which require the data to be independent and identically distributed from a well-defined distribution. If the distribution of the random variable, in this case $\ln(R_t)$, is not well defined then the properties of the parameter estimators will be unclear. (N.B. - It is not the parameters that have distributions but the *estimators* of the parameters.)

Dennis et al. (1991) present maximum likelihood methods and resulting estimators for the parameters of the diffusion approximation method. The properties of these estimators, i.e., uniqueness, minimum variance, and unbiasedness, are sound. The reasons for not using these methods have been given in the manuscript and were based on simulations results where distributional assumptions that were most likely violated, leading the poor behavior of the estimators. Unfortunately, the properties of the distribution of R_t. are ill defined, leading to problems parameter estimation and quantifying uncertainty. Estimation of the mean, or $\log(\lambda)$, can be robust to distributional assumptions because of the Central Limit Theorem. When $\log(\lambda)$ is estimated using a Bayesian analysis the Central Limit Theorem no longer applies. Subsequently the resultant posterior distribution may more sensitive to distributional assumptions about R_t.

The estimation of the finite rate of increase is critical in assessing jeopardy. Although mentioned as a small part of the manuscript, four pages were devoted to the explanation and use of posterior distributions in decision theory, suggesting the use of Bayesian methods in the analysis. Estimation methods of the parameter and uncertainty must be scientifically sound, including methods and results that are falsifiable and repeatable. Arguments given for the use of Bayesian methods usually center on the ability to incorporate prior knowledge. Yet prior knowledge can take on many forms, such as values or limits of parameters taken from previous or similar studies, error structure, model processes that yield model specifications, or probabilistic distributions of the data. All of these are assumptions that can be incorporated into any analysis, and at times be tested for validity when using frequentist methods.

The use of probabilistic statements in decision-making and the quantification of uncertainty are often cited as driving forces in behind the use of Bayesian methods.

However, hypothesis tests can be conducted under several management strategies to examine differences and aid decision-making and uncertainty surrounding parameter estimates are reflected in variance estimators.

Another reason given for the use of Bayesian methods is a result of the interpretation of confidence intervals. A common, and wrong, perception is that the true parameter estimate is equally likely anywhere within the interval, implying a probability associated with the true parameter value. This is not true. The parameter value has no associated probability; it is either in the interval or it is not. Intervals, estimators, and estimated variances are random quantities not the parameter. These are frequentist concepts, but then so is the confidence interval so its attributes are best discussed from that aspect.

Suggestions:

1.) Resolve the distribution of the running sum, or choose a statistic whose distribution can be resolved. This will solve a variety of problems, including the estimation of process error.

2.) Clarify how errors are added to the simulated data set. Do the simulations match how the errors are modeled? Analytic expressions for expected values, variances, and exactly how simulations are conducted would help.

3.) Bayesian methods introduce parameters that are unnecessary in a frequentist analysis. At the minimum a sensitivity analysis should be conducted showing the effect of those parameters on results by,

    a. Estimating parameters and uncertainty with no prior distribution (maximum likelihood methods) and,

    b. Showing the changes in parameter estimates and uncertainty with different prior beliefs, and the meaning of the changes.

Note that limiting parameter values can be done by algebraic manipulation. For example, if the belief is that $a < \lambda < b,$ for $b > a$, the following expression can be substituted for λ, $\lambda = \frac{\exp(\theta)}{1+\exp(\theta)}\left(b-a\right)+a$. The likelihood would be maximized for values of θ, which can take on any value. Estimates and variances of λ can be calculated using a Taylor series expansion.

4.) Part of the analysis should examine if results are sensitive to prior belief (assumptions) and the implications in assessing jeopardy. Furthermore, the reasoning behind the selection of prior distributions should be clear.

Kris Ryding

Technical Review of Population Trends and Extinction Metrics used in the NMFS Hydropower Biological Opinion

Douglas Deutschman
Dept. of Biology
San Diego State University
San Diego, CA 92182-4614
ddeutschman@sciences.sdsu.edu

Selina Heppell
Dept. Fisheries and Wildlife
Oregon State University
104 Nash Hall
Corvallis, OR 97331
selina.heppell@oregonstate.edu

Submitted: January 30, 2004

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The **white paper** by Holmes is an excellent overview of the rationale supporting the use of the Dennis- Holmes parameterization of the Diffusion Approximation (DA) for estimating population trend. The white paper is both statistically sound and extremely well written. In the white paper, Holmes makes three strong claims: 1) that the DA method provides robust estimates that allow the separation of process and non-process error, 2) that the method allows the accurate calculation of metrics of (quasi) extinction risk and 3) that these methods have been extensively studied and validated with West Coast salmon time series. We agree with these claims. The white paper is a compelling review of the literature and rationale.

A total of **six presentations** were made by scientists from the Northwest Fisheries Science Center, the Columbia River Inter-tribal Fish Commission, Hinrichsen Environmental Services, and Paulsen Environmental Research. During the presentations and the discussion that followed, several criticisms were made of the methodology outlined in the white paper and alternatives were suggested. Participants questioned the relevance of extinction metrics offered by the DA approach, as well as model assumptions of stationarity of population growth rates and variance, with specific reference to large-scale shifts in the ocean regime, the hydrosystem, introduction of hatchery fish, and changes in harvest. Issues of data quality and appropriate methods for smoothing and pooling data were also addressed.

Several **alternative methods** and modifications to DA were proposed, including the use of a Kalman filter, development of new indices of population change, and modifications to traditional stock-recruitment models. The white paper and submitted manuscript by Holmes address most of these alternatives. Participants in the discussion also called for improved collection of monitoring data and continued work validating the DA methods against high-quality data.

Issues with the use of DA and the Dennis-Holmes parameterization for salmon management include data quality and estimation of confidence intervals, autocorrelation and stationarity effects on appropriate time windows for assessment, the use of aggregate predictions, and the choice of risk metrics. Our **recommendations** for future research include an analysis of alternative methods (particularly Dennis Holmes vs. Kalman filter) and risk metrics on high quality, real time series. Recommendations for management applications are also offered, including the use of a "moving window" to estimate population trends and variance. It is also important for managers to recognize the importance, and usefulness, of a robust risk metric and analysis to set thresholds for a wide range of populations and data sets.

Our **conclusions** are that the DA approach has been rigorously evaluated, has undergone better scientific peer review than any current methods used for threatened species assessment, and provides the best available approach for objectively estimating and comparing population status for salmonids. Although some further work may still be helpful, these methods are very strong and should be accepted as the current standard.

INTRODUCTION

This review consists of five parts. The first part (Part I) is an academic peer review of the white paper by Holmes. This review includes comments about the scientific ideas presented in the manuscript as well as suggestions for improvement. Part II of this report presents capsule summaries of the talks presented during the workshop and a brief review of the discussion that followed. Our summaries are not complete reviews of the talks. Instead, we highlight the key issues presented as they pertain to the white paper by Holmes and the larger question about estimating population status and trend. The discussion, while informal, was productive and focused on the technical concerns with estimating population dynamics. As a result, we felt that a brief summary of the discussion was informative, and provides the background for our synthesis of alternative methods (Part III) and issues (Part IV). This synthesis draws on published literature, information presented and discussed at the Workshop, and two commentaries received from Workshop attendees (Paulsen and Hinrichsen). The final section of our report (Part V) provides our recommendations for changes to methods used in the Biological Opinion, application to management, and directions future research.

PART I. PEER REVIEW OF HOLMES (2003) WHITE PAPER

The white paper by Holmes is an excellent overview of the rationale supporting the use of the Dennis- Holmes parameterization of the Diffusion Approximation (DA) for estimating population trend. The white paper is both statistically sound and extremely well written. As a result, it is compelling treatment of the material. The white paper covers an enormous amount of material and we discuss the issues presented in the white paper in five (5) major sections. We discuss each of these sections in turn.

Table 1*: Major sections of the white paper by Holmes. Each section is discussed in this review in the order presented in the white paper.*

Section	Description	MS Page(s)
Background	Regulatory and legal history that led to the development of the white paper and the subsequent workshop as part of the remanded BiOp.	1
Introduction	Introduction to population trend estimation in the face of large amounts of variability and uncertainty. Includes citations that support the methods.	2-5
Diffusion Approximation	A coherent overview of the application of the diffusion approximation method including parameter estimation, risk metrics, validation studies, and applications to salmon.	5-13
Comparison of DA to Other Estimators	Explores the performance of the DA methods compared to other methods including log recruits through time, geometric mean, smolt to adult ratio and others.	13-16
Extended Discussion	A discussion of important issues including choice of time frame, utility of confidence intervals, potential application of Bayesian methods, and quasi-extinction thresholds.	17-22

Background:

Holmes summarizes the history of the Federal Columbia River Power System (FCRPS) Biological Opinion (BiOp). The 2000 BiOp was based on estimating the probability of extinction at 24 and 100 years, as well as the probability that population growth rates were positive. The BiOp assumed that methods for estimating growth rates would advance and that the most appropriate method(s) would be chosen from the peer-reviewed literature. As part of the remand process, a critical review of methods for estimating population growth rate was required.

Introduction:

The white paper reviews research since 2000 and focuses on papers by Holmes, Fagan and McClure. These papers establish the utility of the DA method as well as present results of cross-validation work with salmon and other endangered species.

The introduction also provides a hypothetical example of a declining population to motivate the need for careful estimation. This simple example is important and can be improved with two simple modifications. Figure 1 presents three trajectories from a declining population. It is a nice motivating example but could be improved with a second simulation of a declining population with autocorrelated deviations in year-year population growth. This would provide a visualization more relevant to salmon where we expect boom-bust dynamics and/or the potential for long-term runs of good or bad years.

In Figure 2, Holmes presents growth rate estimates for several methods including the ML, slope and runsum methods (ML and "slope" are not clearly distinguished). The example would benefit from several minor changes. First, it would be helpful to visualize the raw data before showing the population estimates. Second, it would be helpful to show how the different estimators converge on the true population value as the data series increases in length. In year 50, we are not constrained to using only the most recent 20 years and this is not reflected in the example. Finally, this example was simulated using uncorrelated errors (high-frequency noise) and an illustration of the same comparison for data with runs of good and bad years (as suggested for Figure 1) might be useful.

Diffusion Approximation Methods:

Holmes does a nice job summarizing the history of, and theoretical basis for, the DA method. She refers to the classic literature particularly books and papers by Caswell, Tuljapurkar, Dennis, Morris, and Holmes. Direct comparisons with age-structured transition matrices are overstated, however; while both DA and matrix methods results in an exponential rate of population increase, the dominant eigenvalue of a transition matrix is the result of age- or stage-specific processes, rather than a simple statistical measure of observation data. Nevertheless, the use of exponential growth rates as indices for population health and potential for recovery have a well-established place in management applications. The introduction is followed by a very clear explanation of the methods used to estimate the parameters in datasets with significant amounts of process and non-process error. Holmes then asserts (with appropriate citations) that the running sum and slope methods are optimal for estimating μ and σ^2, respectively.

This is followed by a discussion of several key issues in the application of population growth estimates to conservation of salmon: uncertainty about the role of reproduction by hatchery fish, cycles in ocean productivity that affect survival rates, and the use of μ and σ^2 to create risk metrics. This last issue is fundamental and probably deserves a little more explanation. In particular the relationship between μ, λ_i and λ_{hat} (Equation 7 - error?) should be described more carefully and its relationship to the geometric mean be made explicit.

The validation studies of DA for salmon populations are very thorough. The work relies on simulated data from population models of steelhead and chinook as presented in Holmes (2004). The concordance between the DA estimated parameters and the simulated data is extremely high. Clearly the DA method is an accurate approximation under these conditions. However, there is still uncertainty about how this method performs with high levels of autocorrelation. Neither the simulated data presented in Holmes (2004) nor the bootstrapping method in Hinrichsen (2002) were designed to test the accuracy of estimation given extended runs of good and bad years. Although there is debate as to how credible the evidence is for regime shifts and long runs, the simulation method could be extended to test the sensitivity of the method under these conditions. Autocorrelation has been addressed more thoroughly in new simulations added to Holmes (2004), and should be highlighted; under most conditions, DA performs well. The assessment "window", or number of years used to estimate μ and σ^2, will affect the performance of the model.

The results of the validation study might be strengthened by using the simulated data to validate other metrics like the geometric mean and the regression method. This recommendation is a tentative one since the results are well understood from previous published work. The main benefit would be to illustrate the performance of these metrics under comparable conditions.

The match between observed and predicted values (Figure 6) provides a compelling demonstration that the DA technique is accurate. The aggregate mean and variance from the DA technique provide a nearly perfect match to the simulated data. It is important to reiterate that this does not mean that the prediction in any single application (stock or stream segment) will be precise. The degrees of freedom of the t distribution are difficult to understand. The paper should explain why 20 years of data leads to an expected t distribution with ~3 degrees of freedom. This is particularly important since t distributions with low degrees of freedom have very heavy tails and thus there is still substantial uncertainty associated with estimation.

In the final paragraphs about the DA method, Holmes addresses the evidence for strings of good and bad years and their impact on estimation. This idea is very important. The simplifying assumptions underlying Figure 7 seem restrictive. Averaging 200 stocks of chinook and steelhead stocks and looking at a trace of their average does not seem adequate. The rapid transition of the mean of 200 stocks moving from -0.5 to 0.5 in a few years may be evidence of non-stationarity.

Comparison of Methods:

Holmes details several proposed methods for estimating population change including log recruits per spawner, SARs, 8-year geometric means of return rate, simple regression of log natural abundance through time, and residuals from a stock/recruit relationship. The relative merits of each technique are presented in a clear and straightforward manner. For example, regression of log abundance through time, log recruits per spawner and 8-year geometric mean can be used to estimate μ or λ and thus are analogous to the DA method. Holmes asserts these methods are not better than the DA and can be much worse because they may require more data, are often more variable, and do not discriminate between process and non-process error.

In the latter part of this section, Holmes briefly discusses the use of stochastic Leslie matrices (Hinrichsen 2002) and Kalman filtering (Lindley 2003) as alternative methods. Both of these methods have some merit and should be explored in more detail. Our expectation is that neither will prove superior to the DA approximation, but that they may be comparable (at least under some conditions). There is no reason to expect one approach to be uniformly most powerful under all conditions and a pluralistic approach is justified until more comparisons have been made.

Holmes concludes this comparison with three very strong and broad claims.

- The DA method is optimal (our wording, not hers) because it provides robust estimates that allow the separation of process and non-process error.
- DA provides accurate estimates of process error that allow the accurate calculation of metrics of (quasi) extinction risk.
- The methods outlined for estimating λ and extinction methods have been extensively studies and validated with West Coast salmon time series.

We agree with these claims. The white paper is a compelling review of the literature and rationale. Although some further work may still be helpful, these methods are very strong and should be accepted as the current standard.

Extended Discussion

Following the review of the DA approximation and other estimators, Holmes moves into a more speculative discussion of uncertainty. She asserts that decision-making will be improved by examining the likelihood functions (ML) or the posterior probability distributions (Bayesian) rather than point estimates and confidence intervals. This work builds on an increasing literature base about decision making, in general, and Bayesian statistics in particular. In Figures 11 and 12, Holmes points out that informative priors can lead to substantial improvements (increased precision) in risk metrics.

The white paper does not discuss how controversies about the prior distribution would affect this approach. Holmes (2004) provides a more complete treatment of the importance of priors in Bayesian analyses. In particular, Figures 8 and 9 show the relationship between the prior and posterior distribution. In some cases (μ, λ) the results are robust to the choice of prior distribution. In other cases (σ^2) the choice of priors influences the results quite dramatically.

The white paper ends with a discussion of quasi-extinction thresholds. Holmes refers to a large body of literature that supports the notion that risk should be evaluated relative to a quasi-extinction threshold set much greater than N=1. Although there is uncertainty about how to set such a threshold, there are compelling reasons to do so. Both the empirical data and the estimation procedures become unreliable as populations become very, very small due to a variety of processes including Allee effects, demographic stochasticity, and increased (relative) measurement error. We strongly agree with Holmes rationale for establishing metrics that are related to relative declines (say 90%) or quasi-extinction thresholds.

PART II. SUMMARY OF WORKSHOP

Eli Holmes, Northwest Fisheries Science Center

The workshop began with a presentation titled **"Estimation and calculation of risk metrics for stochastic population processes."** In the presentation, Holmes described the work summarized in the white paper. She began the talk discussing the challenges in estimating growth rates from stochastic matrix models. She argued that diffusion approximations require fewer parameters and relied on data that was generally available.

She made the case that cross-validation studies were the most appropriate way to establish the accuracy of the DH methods. Her presentation drew heavily from her excellent recent publications (Holmes and Fagan 2002, Holmes 2004). Her presentation reiterated the claims in her published works that demonstrate that the DA method and the parameterizations used to estimate parameters were accurate. She concluded with a bold statement that the science in the 2000 BiOp is justified.

In the final minutes of her talk, she introduced the ideas on expressing uncertainty contained in the final pages of the white paper. She advocates moving away from point estimates and toward expressing risk using likelihood functions or Bayesian posterior probability distributions.

John Payne, Northwest Fisheries Science Center

Payne followed with a presentation titled **"The flip side of extinction: when can a population be de-listed."** In the presentation, Payne presents several alternatives to the DH approximation outlined by Holmes. Initially, he discussed the relationship between population size, intrinsic productivity, and extinction risk.

He discussed the utility of stock-recruitment patterns as potential predictors of population dynamic. In doing so, he fit alternative models (Ricker, Hockey Stick, Constant Recruitment, etc) to several salmon datasets. Model selection using AIC indicated that the data were too variable to reject the notion of constant recruitment in favor of more biologically meaningful models. This exercise underscores the tremendous variability in salmon time series.

He then proposed that potential for population growth (Population Change Criteria) could be useful in the absence of a good estimate of λ. This is an interesting idea, but one with several limitations that make its application to salmon data very dangerous because it sets fixed targets for recovery. He concluded that the data available should drive the approach used.

At the end of his talk, he described the software that he has developed (with colleagues). The software, currently nicknamed SPAZ, is a JAVA simulation environment that allows the simulation and/or estimation of population growth rates. The software will implement the algorithms currently proposed in the white paper and Holmes (2004).

Saang-Yoon Hyun, Columbia River Inter-Tribal Fish Commission

Hyun's talk titled **"Risk metrics to Spring/Summer Chinook Salmon and Steelhead in the Snake River basin"** focused on the estimation of population dynamics for different populations within an ESU. The models presented were primarily modifications of DA. The talk included a discussion of criteria for "excessive extinction" although this concept was not clearly defined. The talk also reviewed the process of risk estimation using the approaches outlined in McClure et al. and Holmes. The data presented was derived from data on escapement, harvest, and spawning. Hyun derived an index of abundance (I) as a function of annual escapement and harvest data. This index was used as a surrogate for population size in the development of risk metrics for 31 Chinook and 23 Steelhead populations.

Hyun made the point that extinction estimates should include three quantities, the long run mean, μ, the variance, σ^2, and the current population size, N. He suggested using ratio estimators of the functional form:

$$PSR = \frac{N_{current}}{N_{t<1975}} = \frac{\text{constant} \bullet I_{current}}{\text{constant} \bullet I_{t|t<1975}}$$

where PSR is the population size ratio, also referred to as "population size resilience", the *I* values are the abundance values derived from available indices of abundance (escapement, spawning, and harvest data) and N values are the true (unobserved) population sizes. This assumes a constant relationship between the index and N, a common issue for many assessment models. The general idea behind this simple model is that a population can be healthy if a current population index is equal to or greater than the same index measured over a series of years prior to dam construction. The method relies on a point estimate ($I_{current}$) which may not be indicative of current population health, due to stochasticity.

Ultimately, he proposed that extinction metrics be based on the joint distribution of μ (log(λ)) and the PSR. Specifically, he presented the idea that:

$$\Pr(\mu < 0, PSR < 1) = \Pr(\mu < 0) \bullet \Pr(PSR, 1)$$

It must be noted that this formula can only be true if the two events { Pr(μ<0) and Pr(PSR<1) }are independent. Logically, these two quantities cannot be independent. In fact the data presented by Hyun show that the two quantities are strongly correlated. This makes sense since observed population declines (PSR<1) must be associated with negative growth rates (μ<0).

Despite our concern over some of the methods proposed by Hyun, he argued eloquently that Bayesian methods are difficult to interpret when the posterior distributions are too flat. Finally, he reiterated the idea that alternative methods may still be useful and that rigorous comparisons between approaches can be illuminating.

Rich Hinrichsen, Hinrichsen Environmental Services

Hinrichsen's talk titled **"State space approaches to estimating population growth rates"** contained three distinct sections. The first section discussed the use of stochastic Leslie matrices to estimate population growth rates. This summarized work published by Hinrichsen in 2002 in the CJFAS. The second section of the talk focused on the utility of the Kalman filter. In the third section, he discussed several ideas including stock-specific growth rates and evidence for non-stationarity.

In the first portion of his talk, Hinrichsen reviewed the material published in his 2002 paper. He argues that the Dennis-Holmes method is not superior to an alternative approach based on resampling stochastic Leslie matrices. He argues the stochastic Leslie approach is superior to DH when measurement error is small. The argument appears to boil down to a trade-off between bias and variance. The stochastic Leslie matrix approach is not always unbiased. As measurement error increases, bias increases. The DH method is unbiased regardless of measurement error. On the other hand, the stochastic Leslie matrix approach has slightly smaller variance when measurement error is small.

Hinrichsen then transitioned into a discussion of the Kalman filter. His work borrowed heavily from the recent paper by Lindley in Ecological Applications. He concluded that the Kalman filter may be a useful tool that should be explored further.

In the final segment of his talk, Hinrichsen explored the need for more complex models that include stock-specific growth rates, correlated process error, and non-stationarity in salmon time series. He argued that analysis of multiple stocks provides evidence for stock-specific growth rates. He also presented evidence that process errors among stocks were strongly positively correlated. Finally, he argued that there was compelling evidence of an ocean regime shift in 1977 and perhaps again in 1998. He concludes that these shifts are important perturbations to the long-term time series and thus violate the implicit assumption of stationarity.

Charlie Paulsen, Paulsen Environmental Research Ltd.

Paulsen presented the next talk titled **"λ estimation and prediction: problems and (possible) solutions."** This talk, while amusing and thought provoking, led to some of the most contentious debate. His talk was based on several clearly stated premises. First, that estimation is inherently difficult (unreliable) because of measurement and process error. Second, that salmon stocks are extremely variable. Third, that regime shifts are real (with the implication that estimates that ignore this are unreliable).

Early in his talk, Paulsen introduced the stock market as an analogy to the problem of predicting salmon growth rates. He argues that salmon populations, like the stocks, are inherently unpredictable and thus any long-term predictions are simply fiction. In specific, he asked the audience to predict the stock value of IBM and GM 100 years out. The analogy, while seductive, is flawed. A more appropriate analogy might be to compare a start-up company to IBM. Clearly there are strong, empirically-based metrics that show that a start-up has a much higher extinction risk. In addition, it is dangerous to compare the (relatively) simple life-history of salmon to a multinational company. Salmon shareholders cannot vote to divest from the Columbia river and decide to manufacture electric cars instead.

Paulsen followed the analogy with a discussion about short-term and long-term prediction success. He argued that many recent papers have concluded that PVA models are unable to predict extinction. In another engaging quip, he stated that statistics (like those in Holmes) are neat, but beside the point. Instead, he suggests that managers focus exclusively on short-term predictions, be pluralistic, and always communicate uncertainty.

Earl Weber, Columbia River Inter-Tribal Fish Commission

The final presentation was titled **"A variable slope stock-recruitment function for nonstationary data."** Weber began his presentation with a stock recruitment graph from Marsh Creek Idaho. He suggested that the relationship was dominated by three clusters of points from the different time periods (50s and 60s, 70s, 80s and 90s). He proposed adding a 3rd parameter, c, to the standard Ricker function where:

$$R = caSe^{-cbS}$$

This modification deforms the Ricker curve and "stretches" it further across the X axis. The approach is predicated on the idea that the stock-recruitment function is not stationary and that large-scale changes due to ocean regime shifts, damns etc mediate the stock-recruitment relationship.

Weber's presentation was another reminder that salmon populations are impacted by multiple factors and that overly simplistic models may not capture important biological variation. On the other hand, the modified Ricker function seems somewhat arbitrary. It is also unclear how the results from this function could be used to calculate μ, σ^2, and extinction risks.

Summary of the Discussion

The discussion that followed the presentations was lively and interesting. Several topics were brought up repeatedly.

- **Non-stationarity:** several participants in the workshop suggested that salmon populations were strongly impacted by large-scale shifts in the ocean regime, the hydrosystem, introduction of hatchery fish, and changes in harvest. There was concern expressed that these processes are largely ignored by "pure statistical" methods that ignore biologically relevant data.
- **Alternative methodologies:** several alternative methodologies were discussed including the use of a Kalman filter as an alternative to the DA method. There was also interest in seeing population change estimates lagged over a range of time periods. This might provide some insight into multiyear excursions of μ_i from an ensemble mean.
- **Improved data collection:** improved modeling of population change likely requires additional general monitoring data. There were calls for better data on the role of hatchery fish, evidence of density dependent survivorship, and stock specific harvest rates. It is likely that more intense data collection is needed in order to identify minimum data needs for long-term monitoring.
- **More complete validation studies:** there was some dissatisfaction expressed with the validation of DA methods proposed in the BiOp. In particular, the stability of the methods to violation of important assumptions (e.g. stationarity, hatchery fish etc) was questioned. It is possible that some limited additional modeling is appropriate to confirm the applicability of the current methodology.

These discussion points provide directions for future research but in our opinion they did not refute the method outlined in the Holmes white paper. Although methods will undoubtedly continue to evolve, the DA approximation should be adopted as the *de facto* gold standard.

PART III. REVIEW OF ALTERNATIVE METHODS

Alternative methods and modifications to the DH diffusion approximation fall into four categories: trend estimator techniques (age-structured model, regression, stock recruit, geomean), data smoothing techniques to separate observation and process error (Kalman filter, running sums and slope), the importance of the time window for assessment, and ways to characterize express uncertainty. We discuss each of these alternatives in the following sections.

Trend Estimator Techniques

The Holmes white paper provides a good discussion of alternative methods to calculate trends and trend variance. However, all trends for salmon populations require analysis of multiple data sets (see Data Quality, below), and more attention should be given to the relative effects of pooling data in different ways. Data pooling will have an impact on 95% confidence interval estimation for μ and σ^2.

Uncertainty in the estimates of μ and σ^2 is important to appreciate. CIs are largely determined by the number of years in the assessment, providing a measure of reliability that promotes a) improvements in data collection and b) long term monitoring. In his review comments, Hinrichsen points out that the current CI estimates for the stochastic growth rate of a population are based on separate calculations for μ and σ^2. This leads to very wide estimates, effectively counting the uncertainty twice. Hinrichsen suggests that this is inefficient, and provides recommendations for pooling and smoothing data.

Regardless of methods used, the confidence intervals for DA metrics and extinction risk should account for uncertainty *within* annual estimates, not just year-to-year variability. See Data Uncertainty in **Issues**, below.

Adult to adult survivorship or stock recruit relationships, where an entire cohort is measured, requires an assessment lag (all spawners from a cohort must be accounted for) but could be more informative to help interpret trends and correlations (Weber presentation). However, data requirements for good age structure estimates are high. Like population trends and variance, age structure is not stationary. The DH approximation utilizes a stationary age distribution, but does not depend heavily on it; any assessment method that relied on recruits per spawner would require precise assessment of every cohort, every year. Such information is valuable for detecting shifts in survival rates, as might be expected with changing ocean conditions, which could lead to changes in population growth. Despite these important benefits, a recruits per spawner approach to determine population status across an ESU fails as a management model because of the intensive monitoring required to make accurate estimates. Also, projecting population status in the future with recruits per spawner methods has the same problems of stationarity assumptions as the DH methods, but with the added cost of uncertainty in parameter estimates.

Data Smoothing

The DH method provides a stable estimate of μ in the face of both process and non-process error. Hinrichsen proposed using a Kalman filter as an alternative (Hinrichsen comments, Lindley 2003). One potential advantage ascribed to the Kalman filter is that the level of smoothing will be dependent on the data. When estimates track the observations closely, the level of smoothing is low, and when estimates fail to track the observations, the level of smoothing is high. Hinrichsen also argues that nonstationarity can be easily included using the Kalman filter. He proposes a step function be used in the initial characterization of the state equation (Hinrichsen comments). Because this approach yields a variety of models, Hinrichsen recommends that AIC scores be used to compare models.

These ideas are attractive since they address several of the concerns expressed by participants in the workshop. The very flexibility proposed by Hinrichsen also poses significant problems. First, smoothers (in general) are often very sensitive to choices made about the window size and tension used. It is not clear how this approach would be standardized across stocks and years. Second, creating more flexible ad hoc models (i.e. allowing μ to vary as a step function) introduces new sources of potential error and or bias. Adoption of such an approach would require the kind of large-scale data and simulation modeling that has been performed for the DH method outlined in the white paper. The arguments presented by Hinrichsen deserve further consideration, but it is premature to advocate the use of the Kalman filter.

Time Window

Whether using the current DH approach or the Kalman filter advocated by Hinrichsen, the assumption of stationarity and the length of the window will be very important. These two concerns are intertwined because of the potential tradeoff between variance and bias (Figure 1A). As the window length increases, the variance in the estimator decreases (Solid Line). If the system is not stationary, the estimator may (or may not) become increasingly biased (Dashed lines a, b and c). The optimal window size depends critically on the rate at which potential bias grows. In other words, the optimal window size depends on the rate at which older data becomes misleading. If the system is stationary, then the full data series should be used, regardless of length (Figure 1B, panel I). However, if the system has 2 stable states (good versus bad conditions based on the PDO) as advocated by some, the window size should be constrained (Figure 1B, panel II)

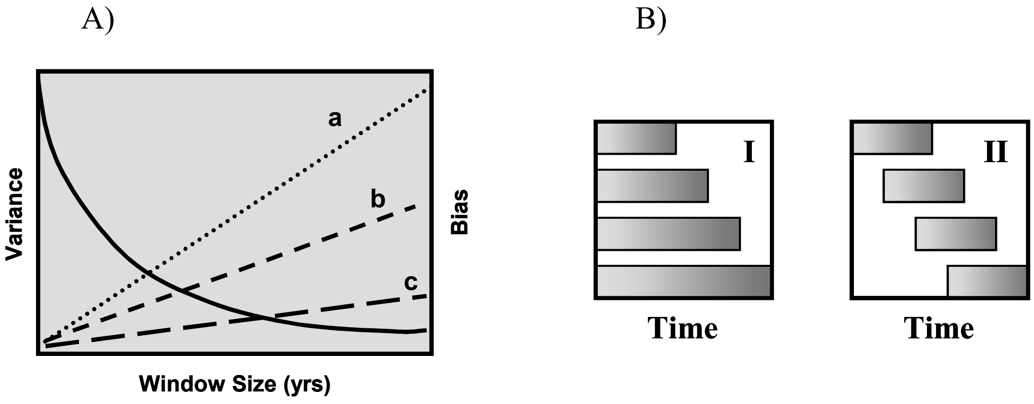

***Figure 1**: **A)** Conceptual relationship between window size and error. Variance should decrease monotonically with increasing length. Potential bias is introduced if the conditions change. The rate at which conditions change is represented by the 3 dashed lines (a - rapid change, b - modest change, c - slow change). **B)** Representation of two strategies (I - accumulating data so window size increases, II – using a moving window based on some optimal window size.*

PART IV. ISSUES

This part of our report reviews a number of issues regarding the appropriateness of the DH method and resulting metrics for assessment of population status. We have identified these particular issues through our reviews of provided materials, peer-reviewed literature, and comments (verbal and written) by workshop attendees. At the end of each issue review, we provide our own assessment of whether the issue has been adequately addressed in peer reviewed publications (methods) or in the BiOp itself (application), based on McClure et al. (2003).

Data Quality and Expressing Uncertainty

Many workshop participants expressed concerns about the choice of metrics and the need to express uncertainty. The written comments by Paulsen, in particular, reflect these concerns.

The white paper by Holmes includes an extended discussion about extinction metrics and expressing uncertainty that arises from the number of years in an assessment. There is some disagreement between Holmes and Hinrichsen about the appropriate degrees of freedom, depending on how data are smoothed – the methods espoused by Holmes need clarification.

Data from multiple sources are used in all assessments of spawner abundance, and are often indices that rely on a range of methods to extrapolate to population size. Sources of uncertainty may be easy to identify, such as redd count errors, or more subtle, such as biases that accumulate from the pooling of several sources of information. Holmes and co-authors have done a good job of addressing *between* year variability due to observation error, and there are new methods proposed to refine this more satisfactorily (Kalman filter approaches, above). However, all of the methods presented assume normally or log-normally distributed measurement error, which may be violated in abundance data. Measurement error is not random. Some stream segments are easy to monitor, others more difficult. Estimates may improve through time with better methods, or decrease due to restricted budgets. A project is underway to document methods used to estimate abundance across all West coast stocks (WDFW), and future efforts may attempt standardize count data or extrapolation methods.

In the white paper, Holmes discusses the potential for developing Bayesian distributions of μ and describes z-transform vs. Bayesian approaches. We suggest that there may be a narrow window for the usefulness of Bayesian analysis to estimate μ distributions or confidence limits. When data are very good, i.e., long time series with relatively small within year error, there may be no need for a Bayesian approach. When data are sparse, the posterior distributions from such an analysis will be strongly affected by priors. Priors are likely to be contentious and not very quantifiable, with different. priors on process vs. observation error. This issue should be explored more thoroughly, with real time series and non-normally distributed measurement error.

Lindley (2003) found that very high levels of measurement errors in simulated data may lead to biases in DH (underestimate process errors), but substantial differences between methods only occurred when measurement error greatly exceeded process error. The real time series examples given by Lindley for sea otter and grizzly are not really comparable to salmon, due to differences in life history and apparently low process error in the data for these species (more overlapping generations?). Holmes (2004) addresses measurement error with simulations and shows very little effect on μ but more on σ^2.

The methods used to pool or massage data to create time series for the DH model are not specified in McClure et al. (2003). Data assessment should be more transparent. Pooling could lead to accumulation of bias due to correlation of errors among data sources. Also, by making data pooling methods an explicit part of the assessment process, the management agency can show how better estimates of population size, or longer time series, can narrow the confidence limits of μ. This could provide incentive for improved data collection if method accounted for uncertainty in the data.

Has this issue been addressed sufficiently? No. This is an ongoing area of research and NMFS should continue to explore and evaluate methods.

Autocorrelation, Stationarity and Biological Realism

The diffusion approximation of population trend and extinction risk includes very little biology. It is unappealing to rely on a model that treats variance as a random process, rather than a mechanistic one that is the result of biological processes. Surely population dynamics and extinction risk are rooted in biology – underlying relationships between spawners and recruits, changes in abundance and population growth due to external forcing on vital rates, and so on. These relationships and stressors should result in cohort dynamics that produce autocorrelation in time series of abundance. Nevertheless, the diffusion approximation works, whether because stochastic processes truly randomize abundance from year to year or cohort to cohort (4 year running sum estimates), or because our ability to count salmon is so compromised by observation error that patterns due to biological processes are undetectable.

Holmes (2001) suggested the running sum transformation in part to account for autocorrelation due to cohort dynamics. The effect of autocorrelation on μ, σ^2 and extinction risk metrics has not been examined explicitly in validation studies, and could be done with simulated data sets and a range of lags. Autocorrelation in the time series data has been addressed to some extent by using age-structured models to generate “true” population trajectories, but could be investigated further to explore long term cycles. The sensitivity of model results to autocorrelation will be strongly dependent on the true variance of the time series. If there is a potential for bias due to autocorrelation, time series could be formally tested and transformed appropriately.

The assumption of stationarity in μ and σ over long time intervals is an important criticism of DH. There are two primary sources of concern: the relevance of a 100 year time frame for prediction of extinction risk, and the potential for long time scale bias or cycling of survival rates due to management, climate, and ocean productivity. Relevancy of predictions that assume constant rates is an interesting philosophical debate (Paulsen), but all model projections require assumptions about the continuation of current conditions. Caswell (2001) provides a clear distinction between *prediction* and *projection*; in order to make relevant comparisons among time periods and stocks, we use models to project potential outcomes given that conditions (trend and variance) remain generally constant.

Ocean cycling effects on west coast salmon stocks are uncertain and undetectable when stocks are considered in aggregate. However, it is premature to dismiss the potential for long-term changes in vital rates due to ocean conditions and their effect on abundance and trends. Francis and Hare (1994) have shown step functions in abundance of spawners in Alaska (pink and sockeye stocks) that correspond with shifts in climate. There is less evidence for such abrupt changes in west coast stocks, possibly due to hatchery influence and differences in life history traits. Relative effects of ocean productivity on population numbers will likely be ESU or stock-specific and depend on life history (number of years spent in the estuary, nearshore and offshore), harvest levels, and hatchery input. Some of these analyses have been done by Payne and McElhany with simulated data (*in prep*), but no consistent biases in error of the DH estimator have been identified. Holmes (2004) provides several simulation results that show robustness of the DH approximation to cycles in abundance and correlation among populations.

Population trends change in response to management, harvest, ocean conditions, and new stressors, such as disease or exotic species. A trend measured over a short time period may be biased, particularly if long-period cycles in abundance occur. The upshot of this is that μ should not be set in stone, but updated frequently. The minimum time window for assessment should be based on biology – for example, 3 generations (15 years) – or the periodicity of ocean cycles, if it can be reliably determined. Time series or assessment windows of shorter length are not to be excluded, but are less reliable indicators of population status. For example, calculations of μ or extinction risk from recent data may, indeed, include only years of decline, but the fact remains that many of those populations were declining to extinction.

Has this issue been addressed sufficiently? Yes. We believe the method has been adequately tested.

Aggregate Predictions

Most of the analyses presented in papers by Holmes and Fagan represent validations of the DH model based on aggregations of time series. Predictions are unbiased, but the spread is often wide; this means that, on average, the DH method is an excellent predictor of average behavior but may not be satisfactory for estimation of a single stock. This is illustrated very well by the confidence intervals given for individual stocks in McClure et al. (2003). Unfortunately, the sample sizes for each time series were not provided, which would have made uncertainty in model predictions much easier to judge.

High variability is the rule, rather than the exception, in most time series of animal populations (Fieberg and Ellner 2000). In his commentary, Paulsen points out that The stakes may be very high on a population by population basis. With the DH approximation, we're right most of the time - is that good enough? Probably not for managers working at small spatial scales. However, this is not the scale at which NMFS, nor the Endangered Species Act, is operating.

Has this issue been addressed sufficiently? Not to manager's satisfaction.

Choice of Metrics for Assessing Population Status

The Dennis-Holmes diffusion approximation provides estimates of trend, variance and extinction probability due to stochastic variability. In comments provided to the review panel, Hinrichsen questions whether the probability of extinction in 100 years is the most appropriate metric for comparing the status of threatened salmonid populations. He argues that other measures provided by DH, specifically the average trend in the population through time, may be a more reliable metric. The arbitrary nature of "5% chance of extinction in 100 years", and the implausible assumption that a trend of any sort would be sustained for such as long time period, are concerns expressed by others attending the review session. Hinrichsen also expresses concern that extinction risk results can be strongly dependent on the choice of the quasi-extinction threshold, as found by Holmes (2001), Payne and McElhany (ref) and others.

The mean and 95% confidence intervals for μ (or its discrete counterpart, λ) could, indeed, be used as assessment metrics, and would eliminate the need to agree on 3 arbitrary variables (QET, time horizon, acceptable risk level). Hinrichsen suggests improved methods for calculating estimates and CIs, utilizing results from the DH analysis and the length of the time series. However, it may be equally difficult to agree on an acceptably high λ. Under most stochastic conditions, λ must be considerably greater than 1 to maintain a population. Confidence intervals around λ or μ are likely to be quite large for salmonids, particularly when sampling error plays a large role in year-to-year estimates of population size (although the Holmes correction reduces the impact of sample variance). Should the 95% CIs for λ not include 1.0? This could be excessively precautionary, particularly for large populations at or near carrying capacity.

Any metric used for assessment will, at minimum, require a choice of thresholds that are not based on science. McClure et al. (2003) do state that the extinction risk metric includes arbitrary values. The advantage of extinction risk is that it includes population size as well as trend; large populations are, generally, less likely to go extinct than small ones. A problem arises when extinction risk is taken literally, rather than regarded as an index. There is a certain preposterousness about the 100 year time frame. However, short-term extinction risk (e.g., probability of extinction in, say, 10 or 15 years), while possibly more palatable to decision-makers, is not an appropriate metric because populations should not be delisted unless their status assures long term viability. Even if we don't believe in a 100 year prediction, extinction risk provides a comparable assessment metric: high extinction risk is not desirable, low extinction risk is.

Has this issue been addressed sufficiently? Not entirely. More work needs to be done.

PART V. RECOMMENDATIONS

1. Perform a Quantitative Analysis of Alternative Methods

We suggest that an assessment of alternative methodologies on real and computer-generated time series (from age-structured models) could clear up some of the debate about the utility and accuracy of different approaches. Specifically, the relative merits of stochastic Leslie matrices and Kalman filtering methods for estimating μ (λ) and, more importantly, the confidence intervals for λ should be explored more thoroughly. Alternative methods and modifications to the DH diffusion approximation should address the importance of the time window for assessment, potential nonstationarity, the costs and benefits of data pooling, and non-independence of populations.

2. Acknowledge and Incorporate Data Quality

Provide distributions of μ whenever possible, using moving windows, Kalman filters, or resampling methods. Acknowledge uncertainty by being up front about it – provide sample sizes and variance for within year population size estimates and the number of years used in the analysis (missing from data table in McClure et al. 2003). Simulations with more complex models have been used to test the robustness of DH under a range of conditions (Holmes 2004). We suggest that a similar assessment be done on monitoring methods and metrics, to determine the level of monitoring required to provide good estimates of population size. A project underway at the EPA Western Ecology Division (Corvallis, OR) may provide some guidance for monitoring effort and data reliability.

We recommend establishment of minimum criteria for including data in assessments. In some cases, poor data of insufficient length can bias estimates dramatically. Criteria could be based on a simple minimum number of years or a minimum confidence interval on μ. If data for a stock do not meet the criteria, that stock will have to remain "unclassified" until more data can be collected. This will be problematic for management, but some level of quality control needs to be established to assure that appropriate results are obtained.

3. Time Frame for Assessment

We advocate the use of a "moving window" of fixed duration (3 generations) to estimate μ and its confidence interval for each stock. The data used in the DH approximation should be smoothed by a consistent method that gives the most robust results (to be determined – see recommendations above). By evaluating μ over several moving windows, we can narrow our confidence estimates of μ and detect changes due to low frequency noise (cycles). The robust estimate of μ suggested by Holmes would remove "outliers" and could be based on standard deviations of population size, but may be more contentious than simply including all the data (smoothed by 4 year running sum or Kalman filter).

4. Density Dependence and Increasing Populations

It is possible that listed populations could rebound relatively quickly and reach a carrying capacity for their environment. Under this circumstance, μ would increase and then level off, possibly masking population health and providing an overly cautious assessment. A population with $\mu = 0$ and high year-to-year variance will exhibit moderately high extinction probability, dependent on population size. Reaching carrying capacity may not be sufficient for delisting if carrying capacity has been compromised by reductions in habitat quality. With a moving window calculation of μ and σ^2, the estimates of the true μ and σ^2 should improve through time. We suggest further scenario testing with real and artificial time series to explore the potential for Type I (delisting prematurely) and Type II (failure to delist a healthy population) assessment errors (Payne and McElhany).

5. Subjective Thresholds: Extinction risk and Quasi-Extinction Population Size

Acceptable risk thresholds are subjects of debate among stakeholders and cannot be addressed by science. The commonly used "5% chance of extinction in 100 years" is an arbitrary risk level over an arbitrary time scale. However, all assessment metrics will include some "arbitrary" thresholds. These may be based on mutually agreed upon risk levels (e.g., through "expert opinion" panels) or, in a few cases, on statistical analyses of simulation results (for an example, see the recovery factor determination for marine mammal bycatch limits in Wade 1998). The DA models can be run with any accepted risk criteria. Standardizing the thresholds enables cross-population and cross-ESU assessment comparisons. The "quasi-extinction" threshold should represent a population size below which recovery is unlikely to occur, at least without intervention. Quasi-extinction thresholds may also be arbitrary, because this is a difficult parameter to estimate from field data. However, quasi-extinction thresholds for salmonids could depend on factors that contribute to density-depensation. Depensation is a reduction in survival or reproductive output that occurs at low densities, and has been noted in statistical analyses of some fish stocks (Liermann and Hilborn 2001). For example, the density of spawners returning to a watershed may need to be above a threshold to assure adequate fertilization rates. Further research and synthesis of existing data may narrow the range of subjective thresholds, but the acceptable risk level for assessment and recovery criteria is largely a societal choice.

6. Use of Qualitative vs. Quantitative Results

Although extinction risks may be unreliable for particular stocks, there is a qualitative “weight of evidence” assessment that should be acknowledged by managers. Although several studies have shown that extinction itself is due to stochastic processes affecting individuals (Fagan et al. 2004), at the ESU level we may need to be satisfied with a relative ranking of risk, given that we probably can’t do much better with data available. Too much attention to the apparent lack of precision of extinction risk for individual stocks may overwhelm the fact that a technique works well most of the time. Nevertheless, while aggregate assessments of DH are usually “right”, the particular attributes of simulations that do not match prediction should be investigated thoroughly.

We suggest that NMFS explore alternative metrics for assessment that compromise between practicality and the best estimate of population status. As long as it is clearly defined and used as a relative index, rather than an absolute, there is no obvious reason why extinction risk is an unacceptable choice.

7. Provide Guidance for Applications to Management

Many of the management application issues discussed at the workshop are under analysis by the TRTs and do not pertain directly to an assessment of the DH approach.

The Lower Columbia-Willamette TRT has an unpublished report on assessment and delisting recommendations that is available through the NWFSC.

Management models applied across a wide range of data quality should be consistent to assure that results are comparable, even if biological realism must be sacrificed. Holmes (2004) provides a number of examples from simulations to show that the diffusion approximation is robust to a wide range of potential effects and assumption violations. However, the robustness of the DH method should not provide an excuse for avoiding better data collection. Good data sets provide critical information for testing the simpler models and developing new and better approaches. Age structure information, in particular, may reveal changes in population behavior that are not evident from a time series of spawner counts, particularly population responses to density and harvest (Hinrichsen).

8. Management Alternatives

Measures of population status, including μ and extinction risk, do not provide managers with information about how to promote population recovery. The relative impacts of natural and anthropogenic stressors on populations requires a more detailed analysis of life history and survival rates (e.g., a sensitivity analysis of an age- or stage-structured model) specific to the population of watershed under consideration. The important thing to acknowledge is that different models are good at different things; DH is designed to provide a robust assessment of population trend, and, potentially, changes in extinction risk if management can increase μ or decrease σ^2. It is not a good model for ranking alternative management strategies unless those strategies can be translated into predicted changes in μ.

CONCLUSIONS

In closing, we suggest that the Dennis-Holmes parameterization for DA, and resulting metrics of extinction risk, are scientifically sound, rigorous to many uncertainties and serve as a robust method to compare the status of salmonid stocks. How model results are interpreted and used is a separate issue for debate by managers and biologists. The simulation tests for model robustness clearly show that, on average, the DA model accurately predicts the probability of extinction. Probability of extinction is an acceptable metric, but should be used as an index. Improvements to the DA estimation and uncertainty estimation may marginally improve these estimates, and should be explored through simulation analysis. However, we concur that the methods outlined in the Holmes white paper provide the best available assessment tool for salmonid stocks across ESUs.

REFERENCES

Caswell, H. 2001. Matrix population models : construction, analysis, and interpretation. 2nd edition. Sinauer Associates. Sunderland, MA.

Fagan, W.F., E.E. Holmes, J.J. Rango, A. Folarin, J.A. Sorensen, J.E. Lippe, and N.E. McIntyre. in press. Cross-validation of quasi-extinction risks from real time series:an examination of diffusion approximation methods.

Francis, R.C. and S.R. Hare. 1994. Decadal scale regime shifts in the large marine ecosystems of the North-east Pacific: a case for historical science. Fish. Oceanogr.3: 279-291.

Hinrichsen, R.A. 2002. The accuracy of alternative stochastic growth rate estimates for salmon populations. Canadian Journal of Fisheries and Aquatic Sciences. 59:1014-1023.

Holmes, E.E. in press?. Beyond theory to application and evaluation: Diffusion approximations for population viability analysis.

Holmes, E.E. and W.F. Fagan. 2002. Validating population viability analysis for corrupted data sets. Ecology 83:2379-2386.

Liermann, M. and Hilborn, R. 2001. Depensation: evidence, models and implications Fish and Fisheries 2:33-58.

Lindley, S.T. 2003. Estimation of population growth and extinction parameters from noisy data. Ecological Applications 13:806-813.

McClure, M.M., E.E. Holmes, B.L. Sanderson, and C.E. Jordan. 2003. A large-scale multispecies status assessment: Anadromous salmonids in the Columbia River basin. Ecological Applications 13:964-989.

Wade, P.R. 1998. Calculating limits to the allowable human-caused mortality of cetaceans and pinnipeds Marine Mammal Science 14: 1-37.

Weber, E. 2003. A more dynamic, variable slope, stock recruitment function for nonstationary data. Draft presented at workshop.

After reviewing the report by Dr. Deutschman and Dr. Heppell, the workshop presenters were given the opportunity to submit clarifications of their presentations. Dr. Hyun sent in the attached clarifications. Dr. Hyun's talk is attached along with the other presentations in section 4. The manuscript upon which this talk and the following clarifications are based is

Hyun, S. and A. Talbot. 2004. Status of Snake River Spring/Summer Chinook Salmon and Steelhead by an Integrated Risk Metric. Columbia River Inter-tribal Fish Commission, report.

Drs. Deutschman and Heppell were given Dr. Hyun's comments (below) and the Hyun and Talbot (2004) paper (above) to review and given the opportunity to change their report if they saw fit. They indicated that their report should stand as originally written.

Dr. Hyun's clarification:

Dear Drs. Deutschman and Heppell

In response to your comments about my presentation on 5 December 2003, I am writing this.

Concern about independence for integrated risk metric

One of the key components in my methods is to make valid the independence between the two events {mu<0} and {PSR<1}. Because of the independence concern, I did not use the same data set for the respective calculation. For example, post-1980 data series (1980 - the most recent year) was used for the calculation of Pr(mu<0), whereas pre-1975 data series (the earliest year - 1975) and abundance index from the most recent year were used for the calculation of Pr(PSR<1). The data used repeatedly for these calculations is only single datum from the most recent year. Though there is an autocorrelation concern in time series of annual salmon abundance, the cohort strength before 1975 is less likely to be correlated to the current population size. Salmon longevity is at most about five or six years.

After my talk in the workshop, Heppell asked about an alternative idea of using data from the recent few years rather than using datum from only the most recent year in calculating Pr(PSR<1). My first thought was it would be possible. But later I realized that the alternative idea would make the independence weak, because quantity of Pr(mu<0) already has the information of data from the recent few years (in other words, estimates of mu an sigma^2 already contain the information). For the valid independence, I favor my original idea. Heppell's question helped me to re-think and justify the independence.

You say '*In fact the data presented by Hyun show that the two quantities are strongly correlated* (line 3-4 on p.8).' These comments surprise me very much. Where did you find them? My presentation file did not show full data set for the analysis. Because you may want to check my talk file, I attach that file here.

Indeed, the actual correlation coefficient value between Pr(mu<0), and Pr(PSR<1) was small (rho = 0.166) enough to support the independence. Because you may want to calculate the correlation coefficient, I show Table 1 below, which has two columns Pr(mu<0), and Pr(PSR<1).

Table 1. Summary of risk metrics estimated with post-1980 data series of spring/summer chinook salmon populations. Code is the population code, Unit is the abundance index's unit, n is the length of annual time series data, and y1 and y2 are the range of annual time series data. Integ. risk is 'Integrated risk metric'.

Code	Unit*	n	yr1	yr2	$\hat{\sigma}^2$	$\hat{\mu}$	Pr(μ<0)	Pr(PSR<1)	Integ. risk
4	RC	20	1980	2001	0.020	-0.103	0.998	1.000	0.998
23	RC	17	1980	2001	0.020	-0.068	0.960	1.000	0.960
20	TLC	19	1980	2001	0.224	-0.107	0.809	1.000	0.809
28	RPM	22	1980	2001	0.318	-0.006	0.519	1.000	0.519
6	RC	14	1981	2001	0.381	0.006	0.489	1.000	0.489
24	RC	20	1980	2001	0.208	0.023	0.420	1.000	0.420
25	RC	22	1980	2001	0.200	0.024	0.411	0.974	0.400
19	TLC	20	1980	2001	0.393	0.042	0.395	0.972	0.384
7	Exp.RC	22	1980	2001	0.050	0.005	0.466	0.789	0.368
2	RC	22	1980	2001	0.159	0.027	0.388	0.622	0.241
14	TLC	22	1980	2001	0.178	0.068	0.248	0.943	0.234
11	Exp.RC	22	1980	2001	0.039	0.031	0.251	0.721	0.181
15	TLC	22	1980	2001	0.431	0.066	0.335	0.526	0.176
1	RC	22	1980	2001	0.062	-0.002	0.511	0.314	0.161
16	RC	19	1980	2001	0.054	-0.049	0.795	0.186	0.147
10	RC	22	1980	2001	0.048	0.067	0.097	0.705	0.068
3	TLC	22	1980	2001	0.029	0.061	0.062	0.787	0.049
9	RC	22	1980	2001	0.044	0.040	0.211	0.233	0.049
21	RC	22	1980	2001	0.467	0.195	0.113	0.415	0.047
5	RC	22	1980	2001	0.023	0.070	0.026	0.416	0.011
31	RC	14	1981	2001	0.233	0.839	0.000	1.000	0.000
26	RC	21	1980	2000	0.020	0.202	0.000	0.996	0.000
27	TLC	18	1980	2001	0.452	-0.116	0.741	NA	NA
8	TLC	22	1980	2001	NA	-0.056	NA	NA	NA

Concern about one point ($I_{current}$)

I agree to your opinion that one point ($I_{current}$) may not be indicative of current population health due to stochasticity. However, I don't reach conclusions only with values of Pr(PSR<1). Conclusions are based on both estimates of mu and sigma^2 in the DA model, and the distribution of PSR. Estimates of mu and sigma^2 in the DA model contain information about stochasticity in population size over time.

Bayesian methods

Your comments about Bayesian methods (the second paragraph on p. 8) also have me feel that you may not have understood my presentation well, or my presentation may not have been clear to you. The section of Bayesian methods has nothing to do with the joint probability of two events, {mu<0} and {PSR<1}. Please refer to my presentation file again.

White paper of E. Holmes

You say, *'... in my opinion, they did not refute the method outlined in the Holmes white paper* (line 25-26 on p. 10).' Though I did not refute Holmes' paper, I raised some concerns in my presentation (again please refer to my presentation file). In the workshop, I reported that Holmes' estimator of sigma^2 (namely slope sigma^2) can lead to its negative value. Because the variance should be non-negative, the estimator needs to be improved. I believe someone else is likely to raise this concern in time.

Another minor issue that I raised was the algorithm of Bayesian calculation. This minor concern can be quickly fixed by a statistician fluent in Bayesian programming, so you may ignore the second concern.

Though I raise two concerns about Holmes' paper, I highly appreciate her contribution in risk/extinction analyses. She is a good scientist.

General comments

Now I feel the workshop audience including you may not have had enough time to digest my presentation at that time. Please feel free to refute or discuss my writing in this letter again. I would sincerely appreciate your re-evaluating my presentation.

Saang-Yoon Hyun, Ph.D.
Columbia Inter-Tribal Fish Commission
729 NE Oregon St. Suite 200
Portland, OR 97232
Email: hyus@critfc.org

Review of methods, progress and cross-validation studies pertaining to population trend and risk assessment for Columbia River salmonids

Updated June 2004 in response to outside comments at the December 5th, 2003 "lambda workshop", written comments, and the review by Drs. Deutschman and Heppell

E. E. Holmes
Northwest Fisheries Science Center
National Marine Fisheries Service
2725 Montlake Blvd. East
Seattle, WA 98112-2097

June 2004

BACKGROUND FOR THIS WHITE PAPER

The 2000 Federal Columbia River Power System (FCRPS) Biological Opinion (FCRPS Biop) evaluated whether the operation of the FCRPS, when combined with survival rates expected to occur in all other life stages of ESA listed salmonids, would result in a "high likelihood of survival and a moderate-to-high likelihood of recovery." This qualitative determination was informed by quantitative estimates for several evolutionarily significant units (ESU). Specifically, NOAA Fisheries evaluated:

- whether or not there would be a 5% or lower probability of absolute extinction of natural spawners within 24- and 100-year periods as a "metric indicative of survival;"
- whether or not there would be at least a 50% probability of the 8-year geometric mean natural spawners being equal to, or greater than, interim recovery abundance levels in 48 and 100 years as a primary "metric indicative of recovery;"
- and whether or not there would be at least a 50% likelihood of the annual population growth rate ("lambda") being equal to, or greater than, 1.0 as an alternate "metric indicative of recovery" for populations lacking interim recovery abundance goals.

The basis for each of these indicator metrics was an analysis of the population growth rate associated with time series for relevant spawning aggregations. Population growth rate was calculated using the methods described in McClure et al. (2003). The Biological Opinion specified that several tests based on population growth rate would be conducted in 2005 and 2008 to ensure that implementation of the Biological Opinion was on track and that populations were not declining further. The Biological Opinion assumed that by 2005 there would be more information about methods of calculating population growth rate, so it specified the following:

> "NMFS anticipates that methods of assessing annual population growth rates will have been refined, based on NMFS' research efforts, those of the Action Agencies, or those of independent scientists. In anticipation of this normal progress in scientific methods, NMFS does not now define a specific method by which population growth rate will be determined for its mid-point evaluations. By March 1, 2005, NMFS will choose the most appropriate method(s) to estimate population growth rate from the peer-reviewed literature, based on collaboration with the Action Agencies, USFWS, and the state and Tribal comanagers."

In June 2003, the Biological Opinion was remanded in National Wildlife Federation v. NMFS. NOAA Fisheries is currently in the process of revising the Biological Opinion and re-evaluating the effects of FCRPS operations and offsite mitigation activities. To facilitate this process, the NOAA Fisheries Northwest Regional Office (NWRO) requested that the Northwest Fisheries Science Center (NWFSC) conduct the above-referenced review of population growth rate estimation methods in 2003. In addition, the NWRO requested that that the NWFSC review related methods of characterizing population trends, especially those that had been suggested as alternatives to “lambda” estimation in comments on the draft of the original Biological Opinion and in comments or litigation since the Biological Opinion was issued.

INTRODUCTION

The purpose of this report is to review and discuss methods for estimating and presenting population trends and extinction risks for Columbia River salmonid populations to support management decisions, such as the ESA Section 7 determination in the 2000 FCRPS Biological Opinion and the anticipated 2005 and 2008 check-in tests. This report reviews research since 2000, which tests and validates diffusion approximation methods for estimating population trends and risks. This review summarizes information from the following publications:

Holmes, E. E. 2004. Beyond theory to application and evaluation: diffusion approximations for population viability analysis. In press in Ecological Applications.

Fagan, W. F., E. E. Holmes, J. J. Rango, A. Folarin, J. A. Sorensen, J. E. Lippe, and N. E. McIntyre. 2003. Cross-validation of quasi-extinction risks from real time series: an examination of diffusion approximation methods. Pre-print.

McClure, M., E. Holmes, B. Sanderson, and C. Jordan. 2003. A large-scale, multi-species risk assessment: anadromous salmonids in the Columbia River Basin. Ecological Applications 13: 964-989.

Holmes, E. E. and W. F. Fagan. 2002. Validating population viability analysis for corrupted data sets. Ecology 83: 2379-2386

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Science 98: 5072-5077.

Summary of work and changes as they pertain to the methods in FCRPS Biop

Changes in the methods for estimating trend and risk:

1) Running sum filter has been standardized to use a simple sum of four consecutive spawner counts. The work leading up to Holmes & Fagan (2002) clarified that this was better than the age-structure based running sum that was originally used.

Cross-validation work

The bulk of the work has focused on validating the methods using real time series (Holmes & Fagan 2002, Fagan et al. 2003) and more realistic simulations that include density-dependence and correlated ocean conditions (Holmes 2004). Also the underlying assumptions of the diffusion model were tested using simulations of salmon models with density-dependence (Holmes 2004).

Expressing uncertainty

Holmes & Fagan (2002) test the variability in parameter estimates from the Dennis-Holmes method and found that the variability is properly estimated. Holmes (2004) begins looking in-depth at how to express uncertainty in a way that it can best inform regulatory decision-making. Confidence intervals are commonly given, but are not very useful beyond showing that there is high or low uncertainty. Bayesian approaches are explored in Holmes (2004). A Bayesian metric is also used in McClure et al. (2003), specifically the probability that λ is less than 1.0 or less than 0.9.

THE NATURE OF POPULATION TRAJECTORIES AND RISK ESTIMATION

Real populations do not grow or decline at fixed rates, but rather show year-to-year variability in population growth rates, which leads to a population trend that varies about some long-term growth rate. Figure 14 shows an example of three population trends that each have the same long-term trend (5% per year decline) and the same year-to-year variability. Even though the population trends were generated with the same underlying dynamics, the trajectories are different. This is nature of populations: random chance means that there are a range of different possible population trajectories given some underlying population dynamics. Even though we cannot predict exactly what will happen in the future, if we could estimate the underlying dynamics governing the population trajectories, we could estimate the probability of different futures, i.e. we could estimate the probability of reaching critical thresholds. We can also estimate whether the population has long-term declining dynamics. To do this, we will need to estimate the following: the long-term rate of decline (or growth), the year-to-year variability in yearly population growth, and the amount of corruption in our data. Within the population dynamics literature, the year-to-year variability in yearly population growth is termed 'process error'; note that it is not technically 'error' in the layman's sense of the word, but rather variability. The rest of the variability is termed 'non-process error' and this includes actual observations errors. For the purpose of this review, one can think of process error as the variability that drives the long-term variability of future population size and the non-process error as the data corruption that is preventing us from estimating the process error.

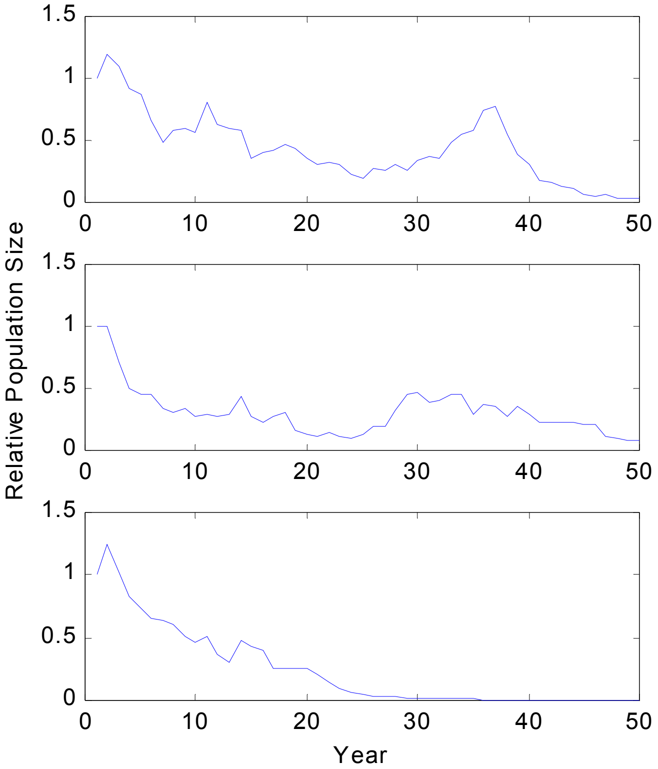

Figure 14. Sample simulated population trajectories from a population with average 5% yearly decline. The underlying population dynamics are identical. The differences are due to chance.

The term λ denotes the long-term rate of population decline (or growth). It is simply the long-term trend that you would observe if you had a very, very long time series of the population. The term λ is the standard notation in the conservation biology literature. λ = 1 means stable, λ = 1.01 means roughly increasing 1% per year, and λ = 1.05 means roughly increasing 5% per year. Similarly, λ = 0.99 and λ = 0.95 mean roughly declining 1% and 5% per year, respectively. Note that we can only estimate λ; we never know the true λ. Our estimates may be unbiased, but that still means that there is a 50/50 chance that the true λ is above or below our estimate.

One of the most common questions is "If λ is the population trend, why not just present the overall trend observed in the data, such as a regression of log numbers versus time?" as opposed to going through the analysis based on theory concerning the dynamics of population trajectories, which is presented in the next section. There are two main reasons why this is insufficient.

1) *We need to estimate uncertainty*. The trend tells you what happened but does not by itself tell you how likely it is that this trend happened by chance and that the long-term dynamics are actually quite different. For example, suppose we collect data on a population that has a true long-term average rate of decline of 12% yearly. Figure 15 shows an example of the population trend observed from 20-year consecutive time series from this population. Segment 1 is from year 1-20, segment 2 is from year 2-21, etc. The wavy lines show the estimates using different methods for estimating the trend; the true value is the straight dashed line. The solid line ("ML") is a simple regression of log natural abundance. The wavy dashed line shows the runsum method used in McClure et al. (2003) and the Biop. There is much variability in the observed trend in a 20-year segment. This variability is an unavoidable aspect of analyzing stochastic population processes. Population dynamics theory allows us to estimate this year-to-year variability and thus estimate how likely it is that a particular observed trend came from a population with a particular true λ (such as an increasing or declining population). But to do this, the estimate of the underlying process error in the population dynamics is needed. A natural response would be to argue that standard regression analyses will give you the uncertainty of the estimated trend, but unfortunately such analyses attribute all error to non-process error and will give you incorrect uncertainty estimates.

2) *We need to estimate probabilities of crossing critical thresholds*. The trend by itself does not give much information about the probability of dropping below critical population sizes. We cannot simply extend the trend into the future and see when our line crosses the threshold. Populations vary from year to year and even a population that has a positive growth rate still has some probability of dropping below the critical threshold by chance. To estimate this probability, we again need to estimate the process error driving the variability in long-term population sizes.

In the following section (section I), I review how the parameters driving a population process are estimated using diffusion approximation methods. This section also reviews the extensive cross-validation work that was done to verify the applicability of these methods for salmon populations. This section directly applies to the methods used in the FCRPS Biological Opinion. At the end of this section, there is a discussion of alternative risk estimation methodologies and why they were not used. The next section (section II) discusses work that goes beyond the methods used in the 2000 FCRPS Biological Opinion. One of the challenges when presenting scientific analyses is presenting the uncertainty in a useful and accurate manner. It is tempting to use the point estimates of risk metrics (i.e. 'this stock has a λ of 0.981') and ignore that this is a statistical estimate. $\lambda = 0.981$ may be the most likely value given the data, but $\lambda = 0.99$ is probably almost equally as likely and $\lambda = 1.01$ may be entirely plausible. Section II illustrates the use of probability curves as a way to formally express this uncertainty. This is a standard approach in decision theory for resource management.

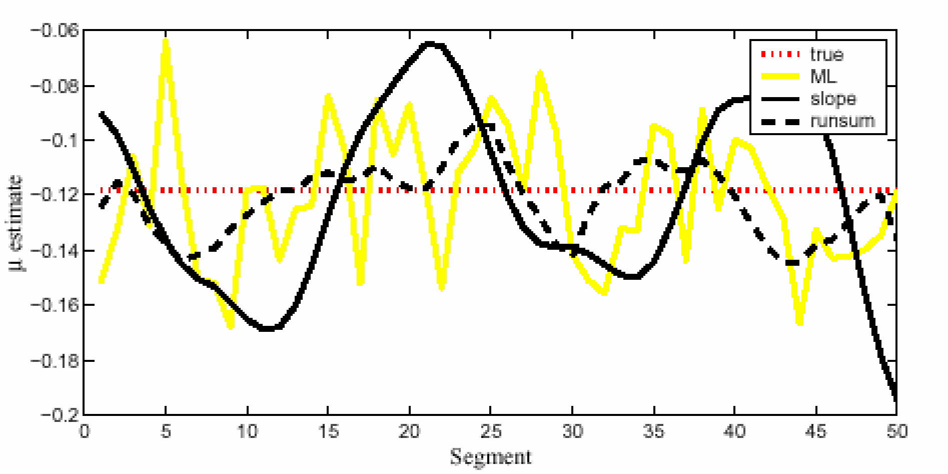

Figure 15. Estimated log(λ) from 20-year segments in a time series. Segment 1 is year 1-20, segment 2 is year 2-21, segment 3 is year 3-22, etc. This shows how the estimates vary depending on the segment observed. The "runsum" method is that used in McClure et al. (2003) and for the Biological Opinion calculations.

I. DIFFUSION APPROXIMATION METHODS FOR POPULATION VIABILITY ANALYSIS

In the last decade, diffusion approximation (DA) methods have been developed that use count data alone (for example, spawner counts) for the estimation of population viability analysis (PVA) risk metrics, such as the probability of crossing extinction thresholds, mean passage times, and average long-term rates of population growth or decline (Lande and Orzack 1988, Dennis et al. 1991). These methods have since been used to estimate extinction risks for numerous species of conservation concern (Dennis et al. 1991, Nicholls et al. 1996, Gerber et al. 1999, Morris et al. 1999, McClure et al. 2003). The appeal of DA methods from an applied standpoint is their simplicity and their reliance on simple census data alone (e.g. neither age-structure, cohort-level analyses, or total fish numbers are required). They have become one of the basic quantitative tools presented in recent books on PVA methods (Morris and Doak 2002, Lande et al. 2003).

Diffusion approximation methods stem from theory concerning the behavior of stochastic age-structured population models with no density-dependence,

$$\begin{bmatrix} n_{1,t+1} \\ n_{2,t+1} \\ n_{3,t+1} \\ \vdots \\ n_{k,t+1} \end{bmatrix} = \mathbf{A}_t \begin{bmatrix} n_{1,t} \\ n_{2,t} \\ n_{3,t} \\ \vdots \\ n_{k,t} \end{bmatrix} \qquad [9]$$

where $\mathbf{A}_t$ is the stochastic population transition matrix, e.g. a Leslie matrix, for time t. Note that most types of cohort or otherwise age-structured population simulations with no density-dependence are specific cases of the general model in Eq. 1. For such models, the asymptotic behavior of the total population size, $N_t = \sum_i n_{i,t}$, is a stochastic exponential process (Tuljapurkar and Orzack 1980, Tuljapurkar 1989):

$$N_t = N_0 \exp(\mu t + \varepsilon_p), \qquad \varepsilon_p \sim \text{normal}(0, \sigma^2 t) \text{ for } t \text{ big}. \qquad [10]$$

and log N_t/N_0 is distributed normal with mean=μt and variance=$\sigma^2 t$ for t big. The parameter μ in Eqn. 2 determines the rate at which the median log population size, log N_t, increases through time, while σ^2 determines the rate at which the distribution spreads, or in other words, the variability of potential population sizes at time $t+\tau$.

Diffusion approximation methods assume that Eqn. 2 holds for all $\tau > 0$ including small τ and that the ε are independently and identically distributed. This allows one to model the population as a diffusion process (Lande and Orzack 1988):

$$\begin{aligned} &\partial p / \partial t = -\mu \partial p / \partial x + (\sigma^2 / 2) \partial^2 p / \partial x^2 \\ &p = P(\log N_t = x \mid N_0 = x_0) \end{aligned} \qquad [11]$$

P(y) means the probability of y. The diffusion model has the property that log N_t/N_0 is distributed normal(mean=μt, variance=$\sigma^2 t$) like the stochastic exponential process it is used to approximate. See Dennis et al. (1991) for a much fuller discussion of the diffusion approximation.

This approximation opens a toolbox of parameterization methods for linear models with normal error. It also provides analytical estimates of quasi-extinction probabilities, i.e. the probability of crossing a particular threshold at some time within a given time frame. Strictly speaking, however, an age-structured population process is not a diffusion process. However despite the assumption violations, the diffusion model approximates many types of stochastic age-structured population processes, as seen both from simulated and real data (Lande and Orzack 1988, Dennis et al. 1991, Holmes and Fagan 2002, Fagan et al. unpublished manuscript, Holmes 2004). In particular as will be reviewed below, the diffusion approximation works well for salmon population models (Holmes 2004).

Parameter estimation methods

Diffusion approximations for a particular PVA must be carefully selected since a poor choice results in poor, highly biased estimates which lead to poor, highly biased risk estimates. Holmes (2004) discusses these issues and careful selection of parameterization methods using salmon data as an example. McClure et al. (2003) presents methods for estimating log(λ) and σ^2, which have been used by NWFSC scientists for salmon PVA.

These methods have been extensively validated with real and simulated salmon data (see next section).

The basic estimation methods currently used for the Biop are presented here without discussion; see McClure et al. (2003) for a discussion and examples. The methods use a running sum transformed time series of spawner counts defined as

$$R_t = \sum_{i=0}^{3} O_{t+i} \qquad [12]$$

where O_t is the spawner count at year t. The estimate for log(λ), which is denoted μ, is

$$\begin{aligned} \hat{\mu}_{run} &= \text{mean of } \log(R_{t+1} / R_t) \\ &\text{for } t = 1,2,3,\ldots,k-3. \end{aligned} \qquad [13]$$

The estimate of σ^2 uses the rate that the variance increases within the time series:

$$\begin{aligned} \hat{\sigma}^2_{slp} &= \text{slope of var}\left[\log\left(\frac{R_{t+\tau}}{R_t}\right)\right] \text{ versus } \tau, \text{ intercept free} \\ &\text{for } t = 1,2,3,\ldots,k-3 \text{ and maximum } \tau = 4. \end{aligned} \qquad [14]$$

These estimators will likely appear somewhat peculiar on first introduction. Note that the μ estimate is very similar to a linear regression of log population counts (typically log spawner counts). Why use the estimator with a running sum transformation of the data? Extensive testing described in Holmes (2001), Holmes & Fagan (2002) and especially Holmes (2004) indicates that the $\hat{\mu}_{run}$ gives the least variable estimates of μ (see also Figure 15). Estimation of the process error is an especially difficult problem. Holmes (2004) reviews the currently available methods in the literature. Again extensive cross-validation work (see especially Holmes 2004) found that $\hat{\sigma}^2_{slp}$ performs the best for salmon data.

One of the difficult problems with analyzing salmon spawner data is that hatchery fish are input into the stocks. Perhaps the easiest way to see how this presents a problem for estimating λ is to consider the analogy of a mutual fund. Suppose you put $1000 into a mutual fund 5 years ago and now you have $8000. You would like to know what the average rate of return (this is λ) has been so that you can decide whether to keep you money in this fund or move to another. Normally, you would just take (8000/1000)^(1/5) = 1.51, which means that your fund returned an incredible 51% per year. However, your benevolent aunt has been automatically adding $100 a month to your brokerage account, and you need to factor this in (these are the hatchery fish). Problem is you don’t know whether her monthly gift was added to your mutual fund (the hatchery fish reproduce) or was simply deposited to your brokerage account but not invested (the hatchery fish don’t reproduce). Without this information, you can only deduce the range of the possible average rates of return for your mutual fund. If not added to mutual fund, the rate of return was ((8000-100*12*5)/1000)^(1/5) = 1.15, which is still a nice 15% per year. If added to the mutual fund, rate of return is found by finding the λ that solves:

$$8000 = 1000*\lambda\hat{}5 + 100*\sum_{i=0}^{12*5} \lambda^{i/12},$$

which is λ = 1.05 and means a rather paltry 5% per year growth. Thus, knowing whether the monthly deposits were added to your mutual fund or not is a critical bit of information you need to evaluate how good a mutual fund you have. This is exactly the estimation problem we have with hatchery fish. We need to know whether or not they are reproducing in order to evaluate the underlying population growth rate. In McClure et al. (2003) and in Holmes (2004), the hatchery correction is presented. In the McClure et al. PVA, the range of λ for hatchery fish not reproducing versus are reproducing is shown. For the Biological Opinion, the range of λ is shown for hatchery fish reproducing 20% as effectively as wild-born fish versus 80% as well as wild-born fish.

Risk metrics

From the parameters μ and σ^2, a number of different risk metrics can be calculated. We have focused on two metrics. The first is the median yearly growth rate or the long-term yearly growth rate, which is denoted λ. Suppose you were able to observe 1000 20-year population trajectories with the same underlying dynamics (i.e. the same μ and σ^2 parameters) and each starting from the same initial size, much like Figure 14. The trajectories would all look different due to chance. The yearly growth rate you observed in the i^{th} (out of the 1000) trajectory is

λ_i = [(end population size)/(start population size)]^(1/(number of years)-1)

The median λ_i from all 1000 would be $\exp(\mu)$; on average 50% of trajectories would show a yearly growth rate greater than $\exp(\mu)$ in those 20 years and 50% would show a lower growth rate. An estimate of this median yearly growth rate is what we term λ. It also happens to be the yearly growth you would observe from a very long time series since λ_i goes to $\exp(\mu)$ as the number of years gets very large. This is why the λ estimate is referred as an estimate of the median yearly growth rate or the long-term yearly growth rate. For a particular time series with *n* years, the λ estimate is

$$\hat{\lambda} = \exp(\hat{\mu}_{run}) = \frac{1}{n-4}\log(R_{n-3}/R_1) \qquad [15]$$

The second metric is the probability of hitting a particular critical population threshold, N_e, within some period of time t_e, starting from the population size N_0. This is calculated from a diffusion approximation of the population process (Dennis et al. 1991):

$$\Pr(N_0 \to N_e \quad by \quad t_e) = \pi' * \Phi\left(\frac{-\ln(N_0/N_e) + |\hat{\mu}_{run}| \; t_e}{\hat{\sigma}_{slp}\sqrt{t_e}}\right)$$

$$+ \exp(2\ln(N_0/N_e)|\hat{\mu}_{run}|/\hat{\sigma}^2_{slp})\Phi\left(\frac{-\ln(N_0/N_e) - |\hat{\mu}_{run}| \; t_e}{\hat{\sigma}_{slp}\sqrt{t_e}}\right), \quad t_e > 0 \qquad [16]$$

$$\text{where} \quad \pi' = \begin{Bmatrix} 1, \; if \; \; \hat{\mu}_{run} \le 0 \\ \exp(-2\hat{\mu}_{run}\ln(N_0/N_e)/\hat{\sigma}^2_{slp}, \; if \; \; \hat{\mu}_{run} > 0 \end{Bmatrix}.$$

The function Φ is the standard normal cumulative distribution function. If you are interested in percentage-wise declines, e.g. 50%, 75% or 90%, then it is not necessary to know the actual population size, since $(N_0/xN_0) = (1/x)$. In this case, the probability of crossing critical thresholds can be estimated using on index data without information on the total number of spawners. If however, declines to specific absolute thresholds are of interest, total spawner counts are needed and it is also necessary to transform the spawner count into a count that reflects the total population rather than just spawners in a particular year. See McClure et al. (2003) for a discussion of this transformation.

Validation studies of diffusion approximations for salmon populations

Here I review cross-validation studies of the performance of the diffusion approximation model for salmon data and populations, including populations experiencing density dependence. Holmes (2004) discusses evaluation of the diffusion approximation and estimation methods using simulated data. This study used detailed population models for Upper Columbia River steelhead, Snake River fall chinook, and Snake River spring/summer chinook as examples. The models were parameterized from survivorship and fecundity data from these ESUs. The models include density-dependence in parr to smolt survivorship reflecting that found in low density Snake River chinook stocks (Achord et al. 2003). The models also include sampling error in the range of that observed for redd-count data (standard error 0.3 to 0.85).

The first question in this study was whether a diffusion approximation correctly described the behavior and probability of crossing thresholds for the age-structured models. The first test described in Holmes (2004) is an examination of the linearity assumptions inherent in the diffusion approximation. This key test is somewhat technical, and is described in Holmes (2004). The results of this test were that the linearity assumptions were satisfied for $t > 5$ years which means that (as is well-known) the diffusion approximation should be used to make medium and long-term projections not short-term projections ($t < 5$ years). The second test was whether a diffusion approximation would properly characterize the probability that the simulated time series would cross a threshold (in this case, 90% decline) in different time frames. This analysis is shown in Figure 16. The gray line shows the actual probability of crossing the 90% decline threshold within different time frames (determined by repeating the salmon simulations 1000s of times) versus the probabilities from a diffusion approximation. This illustrates that the probability of 90% decline in these salmon time series can be described by a diffusion approximation.

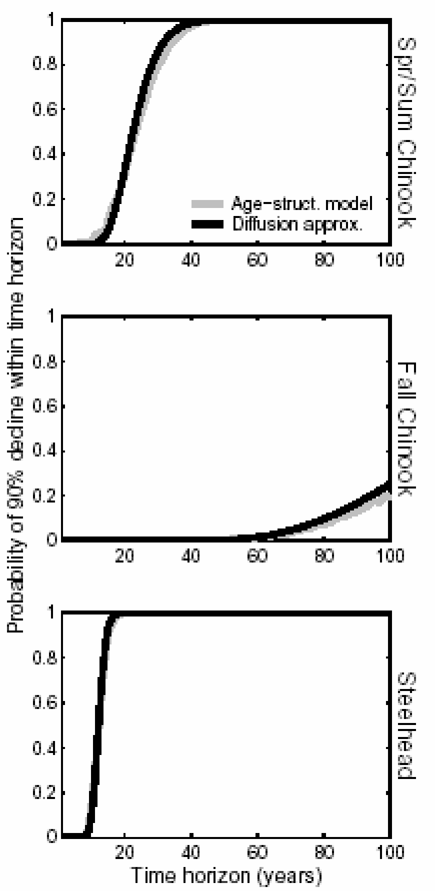

Figure 16. Actual versus predicted probability of 90% decline within different time horizons. From Holmes (2004).

Simply because a diffusion approximation exists which properly characterizes a particular population process does not mean that we can estimate the parameters for that process given realistic data constraints. Holmes (2004) also studies estimation performance given data constraints faced by the PVA of Columbia River salmon stocks (McClure et al. 2003): 1) counts of only the spawning segment of the populations, 2) time series limited to 20 years, 3) severe age-structure perturbations in the beginning of some time series due to reproductive collapses during dam construction (Williams et al. 2001), 4) high observation error in the spawner counts, and 5) high year-to-year correlation in ocean survival (correlation coefficient of 0.8 between adjacent years). Figure 17 shows box plots of the estimates of log(λ) following the estimation methods described above (also in McClure et al. 2003) for 1000 random simulations from the three species' models. The output from the models (spawner counts) was 'corrupted' by different levels of sampling error: age (meaning an age perturbation due to no reproduction in one year), low, medium and high observer error. The dotted line in the graph shows the true value of log(λ). In the box plots, the middle line is the median estimate of log(λ) and the box encloses 75% of the estimates. As can be seen in the figure, the runsum method for estimating log(λ) works for these simulated salmon time series even within the data constraints; the median estimate is the true value even with added observer error in the spawner counts.

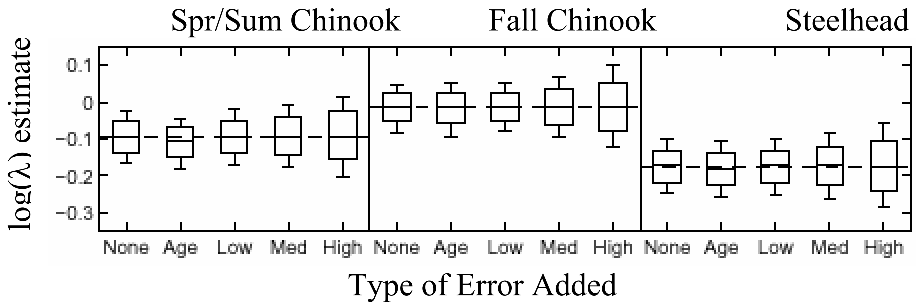

Figure 17. Distribution of log(λ) estimates using $\hat{\mu}_{run}$ from 1000 simulated time series from age-structured models of Snake River spring/summer chinook, Snake River fall chinook, and Upper Columbia steelhead. The models include density-dependent smolt survivorship. From Holmes (2004).

Figure 18 shows a similar analysis for the estimation of the process error, termed σ^2. Recall that the process error specifies the variability of potential future population trajectories and is a key parameter determining the probability of crossing thresholds. This analysis indicates that for low observation error $\hat{\sigma}^2_{slp}$ provides an unbiased estimate of the true value of σ^2, but as observation error increased, the estimate becomes increasingly biased. 'Medium' represents the average estimate of typical observation error in the Columbia River data based on studies of observer error in redd count data (see discussion in Holmes 2004).

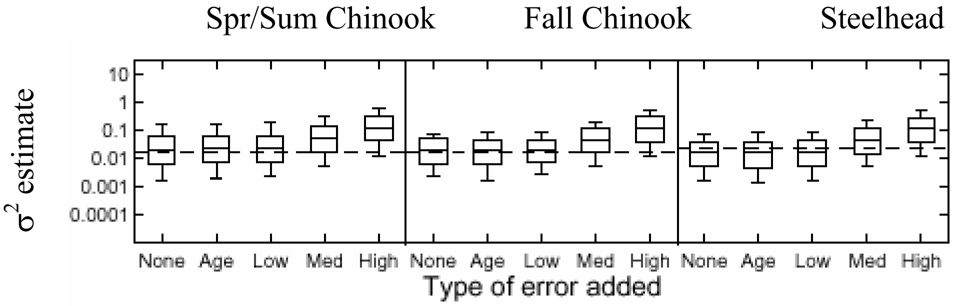

Figure 18. Distribution of $\hat{\sigma}^2_{slp}$ estimates from 1000 simulated time series from age-structured models of Snake River spring/summer chinook, Snake River fall chinook, and Upper Columbia steelhead. The models include density-dependent smolt survivorship. From Holmes (2004).

Simulated data is very useful, however it is 'simulated' and certainly lacks some aspects of real time series data. Another cross-validation (Holmes and Fagan 2002) involved testing the bias and precision of the diffusion approximation parameter estimates using hundreds of real time series. The strategy was to use the first 15 years of a time series to predict the second 15 years of the time series. The bias and variability of these predictions could then be tested against the predicted bias and variability. The two parameters tested were log(λ) and σ^2, which appears in the probability of crossing thresholds metric along with log(λ). Figure 19 shows the results of this analysis for the log(λ) estimates. This analysis involved 30-year time series within a 1920 to 1999 time frame.

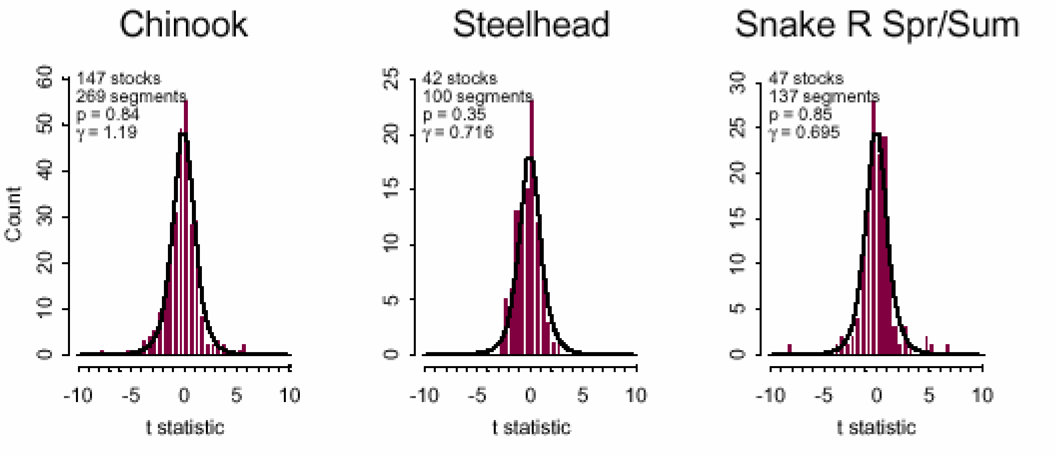

Figure 19. Distribution of actual log(λ) estimates (bars) versus predicted distribution (solid black line from 147, 42 and 47 chinook, steelhead and Snake R spring/summer time series respectively. From Holmes and Fagan (2002).

The close match between the observed and predicted distributions indicates that the log(λ) estimate was properly characterized in terms of its mean value (the peaks match). That is the mean trend in the first half of the time series was the same as the mean trend estimated in the second half of the time series. Figure 19 also demonstrates that the uncertainty in the log(λ) estimate (its variability) was also properly characterized since the width of the distributions match.

That the mean trend in the first 15 years was the same as the mean trend in the second 15 years appears at first glance to contradict the observations of strings of good years versus bad years. However keep in mind that this analysis used 30-year time series across the 1920 to 1999 period. It was asking about the average estimate across different time periods. What about estimates only during a specific time period? Figure 20 shows the difference between the trend in the first 15 years of a time series versus the following 15 years for specific time frames, i.e., not the average across all time periods, but the average if you only look at time series in a specific time period, say 1970-1999. The solid line is a measure of the difference between the trends in the first 15 years versus the following 15 years. Deviations above zero indicate that on average there was a more declining trend in the first 15 years versus the next 15; while deviations below zero means that on average the population was declining less in the first 15 years versus the next 20 years. These results show the average difference from all the West Coast time series put together. What you can see is that across the West Coast, stocks were on average declining more in 1959-1973 versus in 1974-1993 while the opposite was true for 1964-1978 versus 1979-1998.

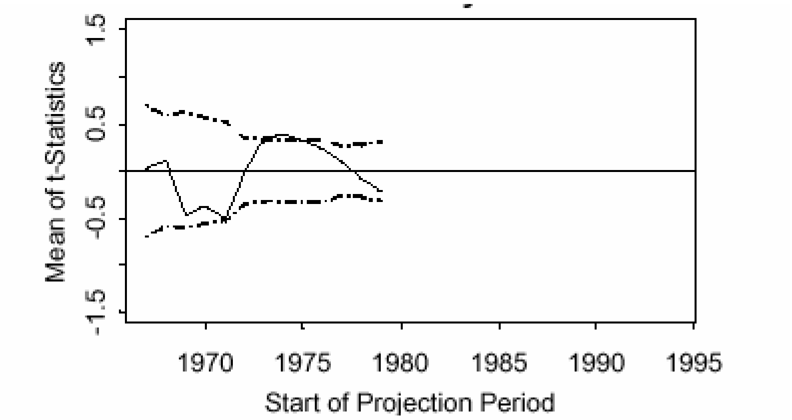

Figure 20. The solid line measures the difference between the average trend in the first 15 years versus in the second 15 years from a collection of 30 year time series of West Coast chinook and steelhead stocks (200+ stocks). "Mean of t-statistics" = 0 means that the average trend (across the whole West Coast) was similar in the first 15 years versus the following 15 years. "Mean of t-statistics" > 0 means that the on average stocks were declining more in the first 15 years relative to the following 15 years. The year on the x-axis denotes the start of the middle of the 30-year segment. The dashed line is the 95% confidence intervals for a random collection of time series, i.e., if there were no underlying environmental cycles causing "good" and "bad" series of years. Holmes unpublished analyses.

It is tempting to attribute these 'good' versus 'bad' strings of years to an environmental driver, such as ocean conditions that one could presumably model. While this may be the case, the data by themselves do not necessarily support this since this type of cycling good and bad strings of years can happen simply by chance in a collection of stochastic population time series. Indeed this is what Figure 15 illustrates. The dotted lines in Figure 20 show the 95% confidence intervals assuming that the time series were all completely independent. This is a conservative estimate since they are not all independent and the true 95% confidence intervals are farther apart. What we can see is that the solid line falls within the conservative 95% confidence intervals suggesting this West Coast pattern of good and bad strings of years is not inconsistent with the hypothesis that it occurred by chance.

The Holmes and Fagan (2002) analysis also looked at how well the diffusion approximation predicted the probability of 90% decline. This analysis searched for a difference between the mean diffusion approximation estimates of the probability of 90% decline and the observed mean probabilities within the collection of West Coast salmon time series. Figure 21 shows the estimated versus actual mean probabilities. The gray solid line (Dennis-Holmes) is the method used in the salmon PVA (McClure et al. 2003). The close correspondence between the actual and observed indicates that first the diffusion approximation approach is correctly estimating the mean probabilities and second that the parameters of this approximation were not being systematically misestimated. Note that this analysis focuses on mean estimates of probability of decline. The issue of the variability in estimates of probability of decline is addressed later in this document.

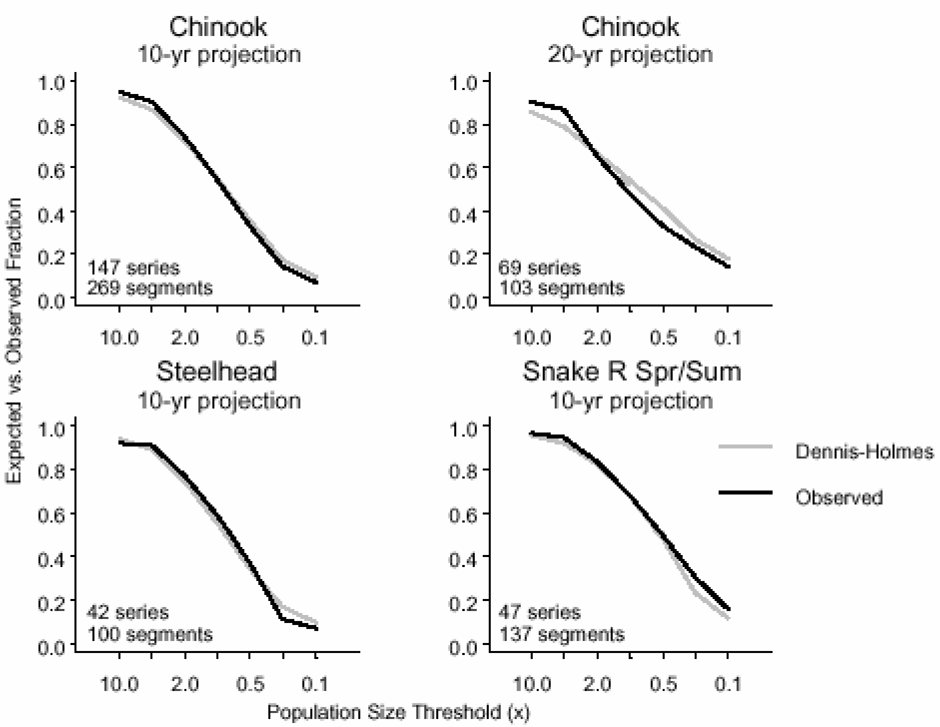

Figure 21. Probability of 90% decline versus observed probabilities with the West Coast salmon time series. From Holmes and Fagan (2002).

Why the diffusion approximation approach versus other approaches for describing trends and risks in salmon populations?

Diffusion approximation approaches for estimation of risk metrics are grounded in theoretical work on stochastic population processes (reviewed in Holmes and Fagan 2002 and Holmes 2004). These methods are one of the basic quantitative tools in population viability analysis and are featured in two current books on quantitative methods for analyzing population data (Lande et al. 2003, Morris and Doak 2003). The long-term rate of population growth is termed λ and is one of the most commonly used risk metrics within the field of conservation biology. Note that λ does not refer to a specific method of estimation, but rather simply the median or long-term trend in the population. There are a variety of methods for estimating λ. The most familiar within the conservation biology literature is to calculate λ from estimated Leslie matrix models. Diffusion approximation approaches present a way to estimate λ when only time series is available, and present a method for estimating the uncertainty in λ, which estimated Leslie matrix models do not provide.

However in the context of salmon management, traditionally other metrics of risk and population trend have been used. Some of the typical metrics that have been used or suggested are log recruits per spawner, SARs, 8-year geometric means of the natural cohort return rate, a simple regression of log natural abundance versus time, and residuals from a stock/recruit relationship. Some of these (log recruits per spawner and a regression of the log abundance versus time) have a close relationship to λ and indeed can be viewed as alternate methods for estimating λ. Many of the other methods, however, differ in a fundamental way in that they measure only a portion of the life cycle, i.e., survivorship or fecundity of only certain stages rather than from spawner to spawner. One of the key aspects of λ is that it integrates across the entire life-cycle. It is not a measure of one stage's survivorship or fecundity alone, but rather of the integration of survivorship and fecundity over the entire life cycle, much like a spawner-to-spawner ratio does. This is important when one is trying to assess a population rather than a particular stage since high survivorship in one stage can easily be offset by low survivorship in another stage.

Below the methods that have been more common in salmon management are discussed in terms of how they relate to λ and the estimation long-term population trends.

Log recruits to the spawning grounds per spawner

Log recruits (to the spawning ground) per spawner is another way to estimate log(λ) since the expected value of (1/T)*ln(R/S) = log(λ), where T is the mean age at which fish spawn. This can be derived from theory on stochastic population processes (see review by Caswell 2001, 14.3.2) and is essentially what is shown by Eqn 14.47 in Caswell (2001) – although this probably will not be transparent on first glance. Obviously the estimates you get of log(λ) from Eq. 5 versus (1/T)*ln(R/S) are going to be different for a specific finite time series; you expect this using different methods even though the expected values (the average estimates) are identical.

If (1/T)*ln(R/S) can be used to estimate log(λ), why not use that since it is more familiar for fisheries biologists? First it is not a more accurate nor less variable estimator – a simple simulation demonstrates this. Figure 22 shows the performance of (1/T)*ln(R/S) as an estimator of long-term growth, λ. The salmon simulations are the same as those used for Figure 17 to test the performance of the runsum estimator. As reviewed previously, this is age-structured simulation (read cohort-based model) with correlated ocean conditions and density-dependent juvenile survivorship. To estimate "recruits per spawner" perfect age of return information was used, however the estimates did not seem to degrade significantly with less precise age information and even 20 fish random samples gave sufficiently accurate age of return information in these simulations. Row 2 shows the estimates from the runsum estimator and Row 3 shows the estimates using recruits per spawner. First note that the recruit per spawner estimates is not less variable than the runsum estimator. Second, note that the R/S estimator is prone to bias, i.e. the median estimate is not equal to the true value, unlike the runsum estimator or the other estimators: the dennis estimator or the slope estimator. This may be due to using a mean age of return and may be rectified by using a cohort-specific mean age of return for each cohort. However, note that the estimates are sensitive to sampling error in the spawner count and that the bias increases significantly when there is high spawner count errors. This bias is unlikely to be improved by using a cohort-specific age of return. The runsum estimator, and the other two estimators, do have these bias problems. Thus the estimate of λ using recruits per spawner estimates requires much more data and effort to estimate – with no apparent decrease in the variability of the log(λ) estimates and with an increase in bias and sensitivity to sampling error. To the extent that the age-at-return data contains errors this adds additional errors to the ln(R/S) estimate.

There are additionally reasons the less data intensive methods are preferable. If we want to compare stock status for example to prioritize recovery actions, using a consistent method across all stocks is critical. For the vast majority of stocks, the additional data to estimate R/S is not available so we can't estimate ln(R/S). Fourth, establishing the uncertainty in the estimate of ln(R/S) would be difficult. We would either have to model the error in age-at-return data, which would require some ad hoc assumptions since we have limited information on this error, or we would have to bootstrap from limited age-at-return data. Fifth, we would still have to estimate the process error and estimating this from ln(R/S) data alone is not possible if the population is affected by both process and non-process error.

These comments are specific to estimation of λ and extinction risks estimation that relies on λ. Clearly there are other reasons for recruits per spawner estimates in a management context. While these comments here are not intended to reflect on any uses other than estimation of long-term trends and extinction risks, these comments are intended as caution concerning how age-of-return information is used. The example here illustrates that incorporating age-of-return data does not always lead improved estimates of trend. Another study, by Zabel and Levin (2002), on estimation of density-dependence illustrates that using mean age of return data in recruits per spawner models leads to mis-estimation of the nature of density-dependence in the population. Together these examples reiterate that methods for estimating population dynamics need to be thoroughly tested on simulated data incorporating the types of sampling error and age of return errors that one expects in the data at hand.

8-year geometric means of the natural cohort return rate

This metric uses the 8-year geometric mean of the spawner-to-spawner ratio for the natural spawning component of the population. Like ln(R/S), this another way to estimate λ. The reasons for not using this metric are the same as those for not using ln(R/S); see above discussion.

Smolt-to-adult ratios (SARs)

SARs, along with other measures of survivorship, are clearly important for analyzing how survivorship changes within a portion of the salmon life cycle. However this metric leaves out the adult-to-smolt portion of the life cycle. For the purpose of tracking the long-term trends, the entire life cycle, spawner-to-spawner, must be included since increases in smolt-to-adult survivorship could be offset by decreases in adult-to-smolt ratios. Thus, SARs are not used for estimating long-term trends.

Note also that SARs detailed types of data, which are not available for many stocks and makes their analysis regionally limited.

Simple regression of log natural abundance versus time

λ is the regression of log spawner counts versus time for an infinite (i.e. very long) time series. One way to estimate λ is to use the regression of log spawners versus time for the available, finite, time series. This method could have been used, but simulations indicated that it gives estimates that are essentially the same as the runsum estimates (Figure 23). Even if one did use a regression, one needs to use the methods in McClure et al (2003) to get the confidence intervals on λ. The confidence intervals on the regression cannot be used since this attributes all error in the data to observation error. This is incorrect; part of the error is process error and part of it is observation error, and one needs to use a statistical framework that properly apportions the error into these two types. In addition, one still needs to obtain the estimates of environmental variability, which are critical for estimates of the probability of crossing thresholds. The regression will not provide this since again a simple analysis of the variance of the residuals attributes error to observation not process error. Holmes (2004) reviews the currently available methods for parameter estimation for population processes with process and observation error.

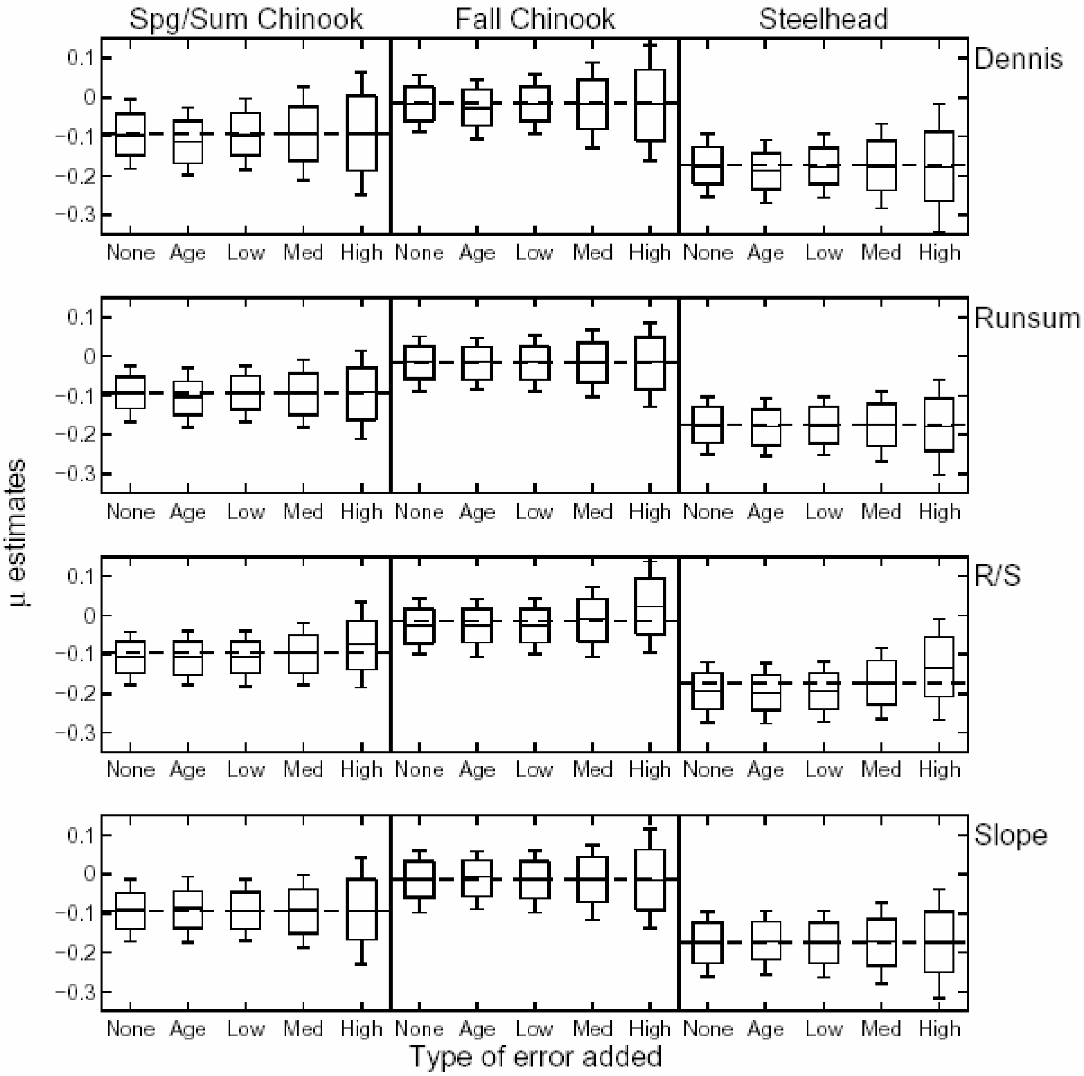

Figure 22. Distribution of log(λ) estimates using the dennis, runsum, recruits per spawner, and slope estimators from 1000 simulated time series from age-structured models of Snake River spring/summer chinook, Snake River fall chinook, and Upper Columbia steelhead. The models include density-dependent smolt survivorship. The "type of error added" refers to the type of errors in the spawner count data. None = no sampling error. Age = no sampling error but no age 0 fish at the start of the simulation. Low, Med, High = sampling error in spawner that is on the low, medium, or high ends of that estimated from Columbia River data. The models are decribed in detail in Holmes (2004).

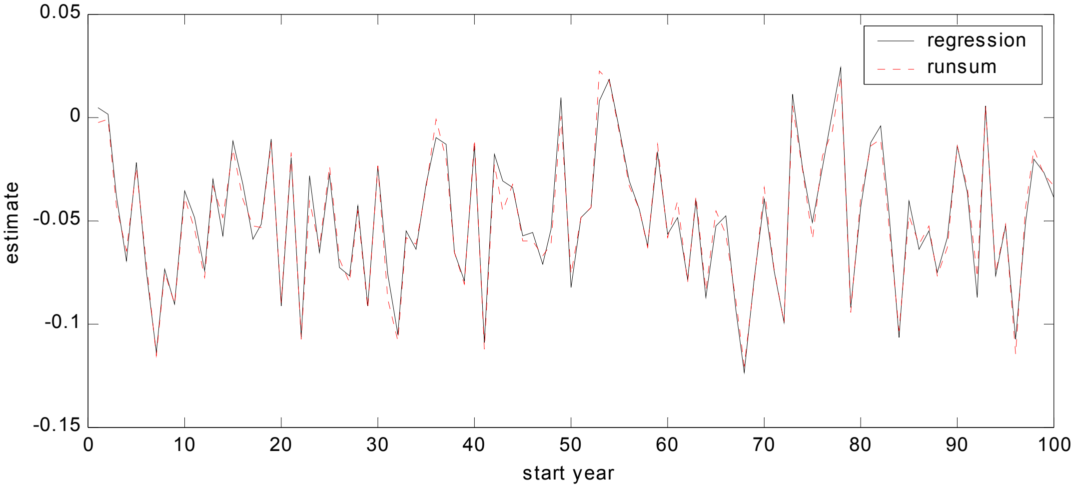

Figure 23. Example of stimated log(λ) from 20-year segments in a time series with $\mu = -0.05$ and $\sigma^2 = 0.02$. Segment 1 is year 1-20, segment 2 is year 2-21, segment 3 is year 3-22, etc. This shows how the estimates vary depending on the segment observed. The regression line (solid black) is from a regression of log counts versus time; the "runsum" method (red dashed) is that used in McClure et al. (2003) and for the Biological Opinion calculations.

Residuals from a stock/recruit relationship

Residuals for a stock/recruit relationship give information on how conditions in one year or cohort deviate from some longer trend. This can be useful for trying to determine if underlying changes for the long term trend has occurred, but is not useful for estimating the long term trend itself. Potentially these residuals could be used to estimate the environmental variability, although this is certainly not straight-forward. The variability in the residuals will be due not only to environmental variability but also variability due to density-dependence and the proclivity of salmon for "boom-bust" cycles. These latter types of variability are important for the short-term variability in population trajectories, but tend to dampen out with time and are less important for the long-term variability in population trajectories. Holmes (2004) gives an example of this using age-structured salmon models with density-dependence. Note also that residuals for a stock/recruit relationship also require age-structure data, which makes their analysis limited to stocks with that kind of detailed data.

Other methods

The methods used in McClure et al (2003) require very simple data, spawner time series, however, there are populations with much better and more detailed data, especially age-at-return and age-specific survivorship data. Incorporation of this data into the estimation of μ can increase the precision of the μ estimate, and consequently the λ estimate. Hinrichsen (2002) discusses estimation of λ using age-at-return information and shows how using this information increases precision although there is no change in bias relative to the λ estimate in McClure et al. (2003). The downside is that the methods in Hinrichsen (2002) are sensitive to high levels of observation error, for example, standard deviation of observation error greater than 0.7, which is certainly seen in redd count data (see discussion in Holmes 2004). More analytical work needs to be done to get around this sensitivity to observation error, but certainly this research suggests that more precise λ estimates can be obtained for those stocks with more extensive data. This is an area that is very promising, however, for regional analyses where we need to compare risks among stocks, some of which are data poor, we will have to continue to have and rely on methods that use only spawner time series for the sake of consistency.

Lindley (2003) presents state-space estimation for noisy time series and offers this as alternative to the estimation methods used in McClure et al. State-space estimation enables maximum-likelihood estimation of μ and σ^2 from noisy data (such as we have for salmon data). It has a strong statistical foundation. I have also been researching state-space estimation and tested Lindley's algorithm in Holmes (2004) and found that it gives much worse estimates of σ^2 than $\hat{\sigma}^2_{slp}$ given the particular characteristics and constraints we face with salmon data. The m estimates were similar to $\hat{\mu}_{run}$, however. I have also investigated a slightly different state-space algorithm for estimation and found similar results. State-space estimation is extremely promising, but a significant amount of research is still need to come up with algorithms that perform more robustly than the current methods in McClure et al.

Summary

While these commonly used metrics are useful for other questions, such as looking for survivorship changes in a particular habitat or life stage or understanding the contributions of particular age classes to recruitment to the population, they are limited in terms of estimating long-term trends, either because they look at just a segment of the population, lead to λ estimates that are more variable than the λ estimates used in the Biological Opinion, or require data that is not available across all populations. Furthermore these other methods do not lead us to an estimation of underlying variability in the population process (process error), which is essential for estimation of the probability of crossing critical population thresholds and for calculating the uncertainty in our risk estimates. The methods used for estimating λ and extinction metrics as described in McClure et al. (2003) have been extensively studied and validated with West Coast salmon time series (Holmes & Fagan 2002) and also salmon-specific simulations which include density dependence (Holmes 2004).

The time frame of ones data and λ estimates

Typically choices must be made about the data, specifically the years, to use to estimate λ. The point estimate of λ will depend on the time frame used, however keep in mind that in general the point estimate of λ should never be used alone since by itself the point estimate does not give an indication of the uncertainty in this estimate. One way present the uncertainty is to use confidence intervals, but confidence intervals are often misleading since they give the erroneous impression that the true value is equally likely within a large interval. Likelihood profiles or posterior probability distributions of λ are much more useful and give a rapid feel for the uncertainty in the estimate of λ. If one uses a posterior probability distribution, it becomes clear that the estimate of λ is not so sensitive to the time frame of the data or the addition of one extra year of data as would appear when only point estimates are presented.

This being said, selection of a reasonable time frame is very important. The following considerations should generally be kept in mind when selecting the time frame to use: a) more data is better, b) the time frame should be representative of historical trends, i.e. not be dominated by 'good' or 'bad' conditions and not dominated by an isolated perturbation and c) for the sake of uniformity and comparison, the time frame should be consistent across stocks. In McClure et al. (2003) the effect of using different time frames for estimation, specifically 1980-2000 versus 1960-2000, on risk metrics for the Columbia River ESUs is shown. The differences were not statistically significant nor in any consistent direction, i.e. for some stocks the 1980-2000 time period gave slightly more severe risk estimates and for others it gave less severe estimates.

From a management standpoint, λ estimates that vary widely depending on the exact starting year of the time series are problematic, and research showing that the estimates are statistically optimal while satisfying does not lessen this practical problem. There are a couple of strategies that I have proposed to deal with this:

1) Use robust estimators of the mean for the μ estimates. Currently in Eqn. 5, a straight mean is used, however a straight mean is highly sensitive to outliers. My preliminary tudies of the effect of different start years on λ estimates using Snake River spring/summer chinook time series indicated that a robust estimator of the mean eliminated much of the problem of λ estimates that vary widely depending on the start year. There are a variety of robust mean estimators; a trimmed mean is the simplest.
2) I examined the 1970s to present data throughout the Columbia River and found that the 1980-present data was affected by an especially unusual series of years between 1978-1982 or so. The estimates using the 1980-present time frame appeared to be more different that one would expect compared to estimates using any other time frame. My initial analysis suggested that 1976-present would generally be a better time frame to use, although this does suffer from dam effects in the early years for some stocks. The 1984-present data could also be used to avoid the 1978-82 period, however, a strong argument can be made that this overly emphasizes a period characterized by bad ocean conditions.

II. ACCOUNTING FOR UNCERTAINTY IN RISK ESTIMATES

Note the review by Drs. Deutschman and Heppell focused on the support for the DH method as discussed in Section I above. The DH method has undergone extensive review and has been used a variety of management documents. In contrast, this section represents an initial proposal for integrating uncertainty estimates into management decision-making. What method to use is still an area of active discussion.

A certain amount of variability in estimated parameters and risk metrics is an unavoidable aspect of the analysis of stochastic population processes, simply due to the nature of these processes. Accounting for uncertainty is important when using risks estimates for management decisions. To be able to do this, one first must have a good understanding of the uncertainty in the estimated risk metrics. One of the strengths of diffusion approximation methods is that the statistical distributions of the estimated parameters are known. As a result, the uncertainty in the estimated risks can be calculated. This is often not the case for other PVA approaches.

There are a variety of approaches that one might use to account for uncertainty. One standard method is to simply show confidence intervals, but straight confidence intervals give little indication of the level of data support across the interval. Approaches that use the whole likelihood functions the risk metrics, rather than simply the point estimates and confidence intervals, will help to clarify the level of data support over a range of true risk levels. Below a statistical decision theory (e.g. Berger 1985 is one of many texts on decision theory) framework based on Bayesian posterior probability distributions is presented for integrating estimates of the data support for different risk levels with the consequences of true risk levels. Wade (2000) and Dorazio and Johnson (2003) provide recent discussions of this Bayesian decision framework in conservation biology and resource management contexts.

The idea in a nutshell is to estimate the probability that the risk metric, for example λ, is within particular ranges that are important from a management perspective. For example:

Probability the true λ is in these ranges	< 0.9	0.15
	0.9 – 0.95	0.3
	0.95 – 1.0	0.5
	> 1.0	0.05

Table 1. Estimated probabilities that λ is within different ranges.

These probabilities are estimated using the posterior probability distribution that is estimated from the data. Figure 24 gives an example of the posterior probability distribution for λ estimated from a 38-year times series of spring chinook in the Upper Columbia River basin (data from T. Cooney, NMFS). The probability that λ is within the range *a* to *b* is calculated by integrating the posterior probability distribution between *a* and *b*.

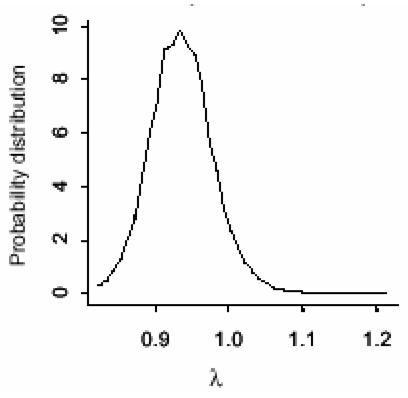

Figure 24. Estimated posterior probability distribution for λ for Upper Columbia spring chinook. From Holmes (2004).

There are a variety of ways these probabilities could be used. They might be used alone and qualitative thresholds set, such as if the probability that λ is less than 0.95 is greater than some threshold, then an such and such actions can (or cannot) occur. Note that the probability of a low λ can be high due to certainty that λ is low or due to uncertainty about λ. Thus, such a strategy leads to caution in the face of high uncertainty. A more quantitative, decision-theoretic, approach can be taken if the probability that actions will be ‘sufficient’ (however that might be defined) can be calculated given different true λs. For example,

		Probability of action being ‘sufficient’		
		Action A	Action B	Action C
λ range	< 0.9	0.3	0.1	0
	0.9 –0.95	0.5	0.3	0.2
	0.95-1.0	0.8	0.6	0.5
	> 1.0	1.0	1.0	0.8

Table 2. Estimated probabilities of action sufficiency given different true λ ranges.

These probabilities are multiplied by the probability of λ being within those ranges and then summed over all ranges to give the total probability that actions are ‘sufficient’. This probability incorporates the uncertainty in the estimated λ:

Probability of action being ‘sufficient’		
Action A	Action B	Action C
0.64	0.45	0.35

Table 3. Probabilities in Table 2 multiplied by those in Table 1 and summed over all λ ranges.

An example where the probabilities in Table 2 would be relatively easy to calculate is different harvest levels. Instead of giving a simple ‘yes/no’ answer as would be the case if using point estimates of λ, this approach quantifies the uncertainty in our estimate of λ and emphasizes that there is not a simple “100% or 0%” probability of an action being effective.

Probabilities of crossing thresholds are notoriously uncertain and variable, and analyzing the uncertainty connected with a proposed probability metric (e.g. ‘probability of extinction’) is especially critical when using these metrics. Figure 25 shows the estimated probability density distributions for the probability of 90% decline within 25, 50 or 100 years given a 20-year time series with an estimated λ of 1 or 0.93. The distributions when the estimated λ is 1 are fairly flat or U-shaped. This indicates that there is not much information about what the probability of 90% decline is. The estimation of the probability of 90% decline can be improved by using an informative prior on the process error. Twenty years of data is not sufficient for accurate process error estimates. If one argues that the variability driving long-term dynamics is similar across chinook throughout the basin, then one might use as an informative prior the distribution of process error estimates for a large number of stocks throughout the basin. Figure 26 shows how the estimation of the probability of 90% decline improves using an informative prior. Now it appears that estimation of the risk of 90% decline in 50 or 100 years is fairly informative for the stock with a low λ. For the stock with a λ equal to 1, 50 and 100-year probabilities are uncertain, but 25-year probability are much better.

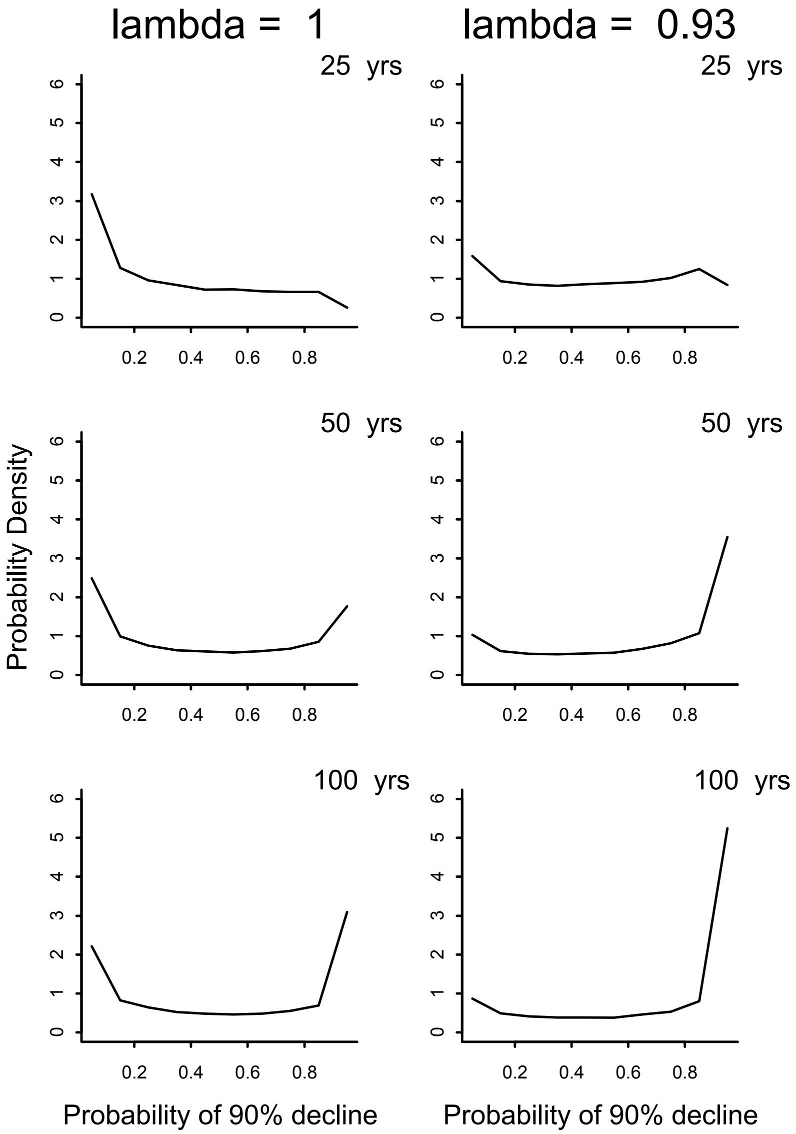

Figure 25. Estimated posterior probability distribution of the probability of 90% decline in 25, 50 and 100 years given a 20 year time series with estimated μ of 0 or –0.072 and an estimated process error of 0.08 and estimated non-process error of 0.71. A uniform prior on the process error was assumed.

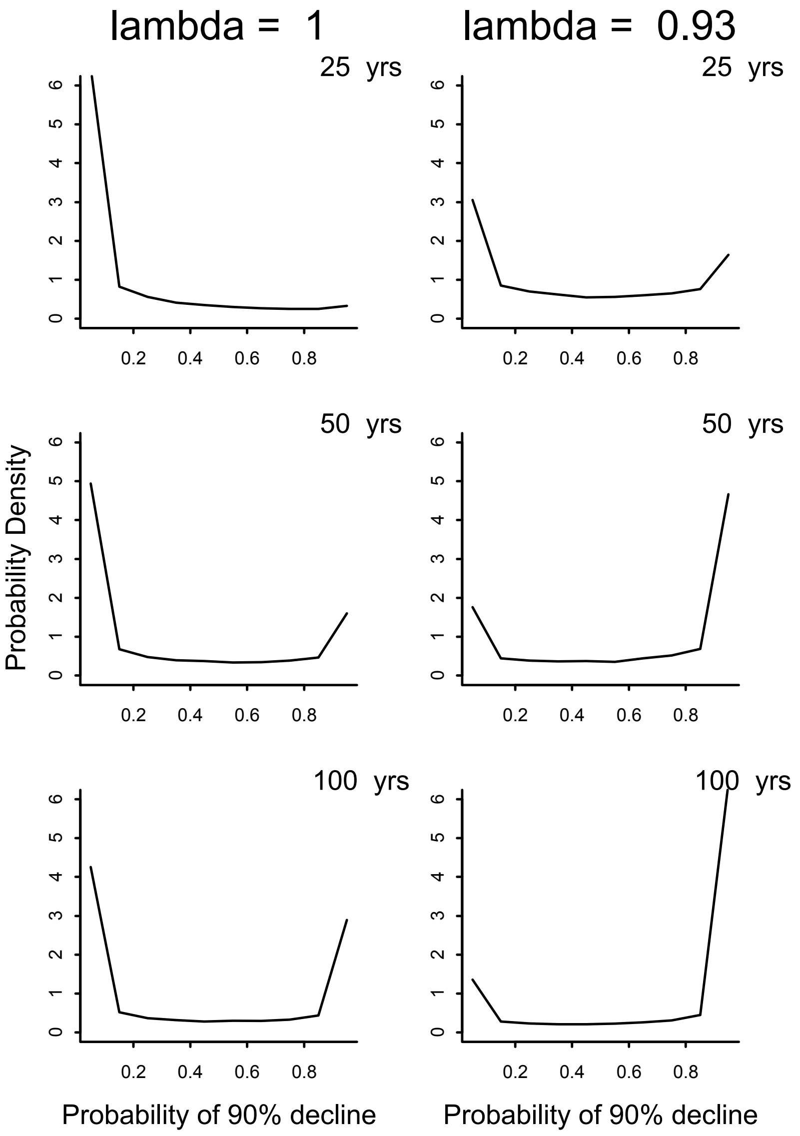

Figure 26. Estimated posterior probability distribution of the probability of 90% decline in 25, 50 and 100 years given a 20 year time series with estimated μ of 0 or –0.072 and an estimated process error of 0.08 and estimated non-process error of 0.71. A highly informative prior on the process error was assumed.

These last two figures focus on the probability of 90% decline in 25, 50 or 100 years. There are other ways to look at the probability of 90% decline that can be more informative. For example, here is an analysis of the probability of an eventual 90% decline for based on the Upper Columbia spring chinook time series (from Holmes 2004). The estimated probability of eventual 90% decline is almost 1.0, that is it is almost certain to occur (Figure 27, top), however there a great deal of uncertainty as to when this will occur (Figure 27, bottom) except that it is highly likely within 100 years. Figure 27, bottom panel, shows the expected probability of 90% decline within a given time frame. From the figure, on average there is a 70% probability of a 90% decline within 50 years for this population and an average 80% probability that the 90% decline occurs within 100 years.

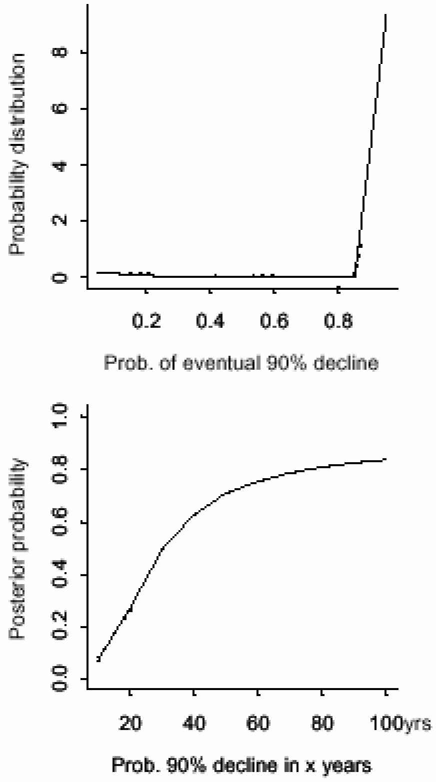

Figure 27. Estimated posterior probability distribution of eventual 90% decline in the Upper Columbia spring/summer chinook (top panel) and the probability that this decline has occurred within different time frames (bottom panel). From Holmes (2004).

The time to 90% decline can then be put into Table form similar to Table 1 for λ ranges (Table 4) which can then be combined with a Table similar to Table 2 for the probability that an action will be sufficient if 90% decline does not occur until after *x* years.

Probability 90% decline only after x years	> 25 yrs	0.7
	> 50 yrs	0.38
	> 75 yrs	0.2
	> 100 yrs	0.18

Table 4. Probability that 90% decline does not occur until after x years. From Figure 27, bottom panel.

Quasi-extinction threshold versus absolute extinction

Throughout this paper, the probability of 90% decline is discussed. The probability of 90% can be estimated for stocks for which we only have index data and not total spawner information. Thus it can be more widely applied. However, decline to specific critical population sizes are also of great importance in PVA analyses. Although estimating extinction to 1 individual is a popular risk metric, and unfortunately sometimes mandated, caution is required when using the diffusion approximation is to estimate extinction to very low numbers since factors that drive dynamics at very low population sizes (such as demographic stochasticity) and the catastrophes often associated with ultimate extinction will likely be poorly represented in a time series of a relatively larger population declining to low numbers. There have been a wide variety of papers published on this in the conservation biology literature. The general recommendation is to estimate the probability of decline to some critical population size (quite a bit greater than 1); this is termed a 'quasi-extinction' threshold. Fagan et al. (unpublished manuscript) studied a collection of actual time series of species that went extinct and compared diffusion approximations for quasi-extinction thresholds versus extinction to 1 individual. This analysis found that quasi-extinction estimates (to a size much greater than 1) using diffusion approximations fit the observed data, but that extinction estimates (to 1 individual) were very poor and underestimate the true risk.

With this in mind, one might ask why was the probability of decline to 1 individual estimated in the McClure et al. analysis. The reasons for this were four-fold. 1) The analysis was focused on estimating risks if current conditions continue as they appeared in the time series data. It was recognized that this would tend to underestimate risks if factors such as density-depensation occurred as the population got small, however baseline estimates of risks under current conditions are required in order to make meaningful statements about risks under hypothetical future scenarios, such as lower population growth rates as the population gets small. 2) The probability of 90% decline does not incorporate the actual population size. The implications of a 90% decline of a population of 10 individuals is quite different than a 90% decline for a population of 100,000 individuals. The probability of decline to 1 individual provided a risk metric that incorporated both the overall rate of decline of the population and the population size. Thus we could then compare ESUs in terms of a risk metric that integrates these two factors –regardless of whether this is an underestimate of the true probability of extinction. 3) Any other extinction threshold we could have specified would have been arbitrary – given the information we had on critical population size. Decline to 1 individual is meaningful for all populations. 4) The Fagan et al. analysis had one notable exception, i.e. one population where the probability of extinction would be properly estimated. That was the one population time series that followed an actual salmonid extinction (sockeye); the rest of the time series followed bird and reptile extinctions.

REFERENCES

Achord, S., P. S. Levin, and R. W. Zabel. 2003. Density-dependent mortality in Pacific salmon: the ghost of impacts past? Ecology Letters 6: 335-342.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer Verlag, NY, NY, USA.

Caswell, H. 2001. Matrix Population Models. Sinauer, Sunderland, Mass.

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

Dorazio, R. M., and F. A. Johnson. 2003. Bayesian inference and decision theory —a framework for decision making in natural resource management. Ecological Applications 13:556–563.

Fagan, W. F., J. Rango, A. Folarin, J. Sorensen, J. Lippe, and N. E. McIntyre. Cross-validation of quasi-extinction risks from real time series: an examination of diffusion approximation methods. Manuscript.

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.

Hinrichsen, R. A. 2002. The accuracy of alternative stochastic growth rate estimates for salmon populations. Canadian Journal Fisheries and Aquatic Sciences 59:1014–1023.

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Science USA 98:5072-77.

Holmes, E. E. and W. F. Fagan. 2002. Validating population viability analysis for corrupted data sets. Ecology 83: 2379-2386.

Lande, R. and S. H. Orzack. 1988. Extinction dynamics of age-structured populations in a fluctuating environment. Proceedings of the National Academy of Science USA 85:7418-7421.

Lande, R., S. Engen, and B. Saether. 2003. Stochastic population models in ecology and conservation: an introduction. Oxford University Press, Oxford, UK.

Lindley, S. T. 2003. Estimation of population growth and extinction parameters from noisy data. Ecological Applications 13: 806-813.

McClure, M. M., E. E. Holmes, B. L. Sanderson, and C. E. Jordan. 2003. A large-scale, multi-species risk assessment: anadromous salmonids in the Columbia River Basin. Ecological Applications 13:964-989.

Morris, W. F. and D. F. Doak. 2002. Quantitative conservation biology: theory and practice of population viability analysis. Sinauer Press, Sunderland, MA.

Morris, W., D. Doak, M. Groom, P. Kareiva, J. Fieberg, L. Gerber, P. Murphy, and D. Thomson. 1999. A practical handbook for population viability analysis. The Nature Conservancy.

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

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

Tuljapurkar, S. D. and S. H. Orzack. 1980. Population dynamics in variable environments. I. Long-run growth rates and extinction. Theoretical Population Biology 18:314-342.

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

Zabel, R. W. and P. S. Levin. 2002. Simple assumptions on age composition lead to erroneous conclusions on the nature of density dependence in age-structured populations. Oecologia 133: 349-355.