Building Models and Gathering Data: Can We Do This Better?

Wim Kimmerer, Bill Mitchell, and Andy Hamilton

Abstract

We are constructing a "second generation" model of chinook salmon for the Sacramento Basin to help investigate factors affecting salmon populations and the effects of management actions. We chose to build a new model rather than modify an older one to apply recent developments in computer interfaces and individual-based modeling and to incorporate a more detailed and flexible geographic representation. We also expected that substantial new knowledge had been developed that would enable us better to characterize the life cycle and influences on survival of chinook salmon. These expectations have not been met, and despite some recent progress we still find gaps between the knowledge available and that needed for successful modeling. Key examples of gaps in our knowledge include sublethal temperature effects, abundance of young fish, factors triggering migration, factors limiting rearing habitat, and survival of young salmon, particularly fry rearing in the mainstem or Delta reaches and early survival in the ocean. We believe these gaps arise for several reasons: (1) a mismatch in perceptions of what data are needed; (2) a lack of institutional commitment to long-term, broad-scale programs to provide knowledge useful in modeling; and (3) the fundamental difficulty of gathering information about environmental influences on fish populations.

Introduction

Models are representations of real-world objects or systems. Simulation models are formal mathematical representations of dynamic systems developed to examine the time course of system response to selected inputs. These models can be used as research or management tools or, if the underlying mathematics and parameters are known well enough, for prediction. Models of ecological systems are rarely suitable for prediction. Simulation models can be useful in investigating properties of a complex system, but are also useful as a framework for organizing knowledge and identifying knowledge gaps.

We are in the second phase of development of a chinook salmon simulation model for the Sacramento Basin. In the first phase we developed a conceptual model which has been distributed for review. In the second phase we are receiving and discussing comments on the conceptual model, while developing the code for the simulation model with the initial goal of producing a working prototype.

In this paper we briefly describe the model and some of its potential uses. We then discuss the more significant gaps in knowledge that have been identified during model development. Some of these gaps have been known for many years, yet little progress has so far been made to close them. We discuss some possible reasons for this and potential remedies that could lead to more effective allocation of effort devoted to research and monitoring on salmon biology and better understanding of the effects of human actions on salmon life histories.

Background

The Central Valley Project Improvement Act (CVPIA), an ambitious effort to increase production of chinook salmon in the Central Valley, mandates the development of "ecosystem models" to support understanding of potential measures needed to restore anadromous fisheries. The model described here is an element of the ecosystem modeling effort designed to assist with analyses and comparison of various alternatives for water and fisheries management. It is intended to build on both current understanding of the ecology of salmon and experiences of previous modeling efforts for chinook salmon in the Central Valley. These efforts include the following:

1. A simple stock-recruit model used to investigate effects of Delta conditions (Kelley and others 1986)

2. CPOP, a cohort simulation model of Sacramento Basin fall-run or winter-run chinook salmon (Kimmerer and others 1989), written in Fortran.

3. EACH, a simulation model for the San Joaquin Basin with similar structure to CPOP, written in Stella (EA Engineering, Science, and Technology 1991).

4. Two statistical models of the effects of Delta conditions on San Joaquin Basin chinook salmon (Speed 1993; Rein 1994: http://felix.vcu.edu/~srein/chinook.ASA/talk.html).

5. An individual-based simulation model of chinook salmon smolt production in the Tuolumne River (Jager and others 1996).

6. Statistical models of mark-recapture experiments using salmon smolts in the Delta (Kjelson and others 1982; Baker and others 1995; Rice and Newman 1997).

7. The CRiSP model of chinook salmon smolt passage, originally developed for the Columbia River by the University of Washington, modified for the Sacramento River in a student paper (http://www.cqs.washington.edu/papers/sacramento.html).

8. A survival model of winter-run salmon (Botsford and Brittnacher 1996).

We refer to our model as a "second-generation" model, because it builds on the results of previous modeling efforts. This model differs from previous models: it applies to all races in the Sacramento Basin; uses an individual-based approach; takes input from a variety of data sources, including flow and temperature data or model output; is designed in modules to simplify analyses of selected stages of the life cycle; will have a modern user interface so users can spend their time learning about the model rather than the program; and is being programmed in an object-oriented language that will make future modifications relatively straightforward.

The model is essentially a large combination of conditional statements about the salmon population. It contains various mathematical descriptions of attributes of habitat and individual fish, which determine responses of salmon to their environment. Many of the mathematical descriptions and the parameters and input variables used to develop numerical values for responses are based on limited data or expert opinion. Thus, it is extremely unlikely that all of them are accurate, so output of the model is not reliable as a prediction of future salmon population trends. Rather, the model will be most useful in a comparison among alternative scenarios. Provision will be made for varying important parameters and selecting alternative mathematical descriptions of functional relationships to determine the sensitivity of the conclusions based on model runs to the assumptions contained in the model.

Model Description

The model is capable of simulating the entire life cycle of all four races of chinook salmon in the Sacramento Basin (Figure 1). Conceptually the model can be divided into the four modules shown in the figure. Individual modules, corresponding to stages of the life cycle, can be run independently to simplify the model run for particular purposes. This will be a useful feature for investigating particular aspects of the life cycle such as spawning or ocean life.

We have chosen to use an individual-based modeling approach (DeAngelis and Gross 1995). Individual-based models (IBMs), also known as agent-based or multi-agent models, are a relatively recent development in modeling made possible by substantial advances in computer memory and speed. In an IBM, populations are represented by some number of individual entities, rather than by cohorts or other aggregates. Models written at the cohort or higher levels of aggregation have many advantages, but they do not accurately portray the population response to environmental change when the individuals in a cohort undergo different trajectories of growth or movement. This can happen when, for example, physical habitat is occupied at a small scale so that different fish experience different environments. A cohort model also suffers from the disadvantage that any nonlinear response of the fish to their environment distorts the statistical distribution of properties within the cohort (e.g., mean weight). Finally, some environmental influences act on individuals over a long period relative to the simulated time step; resolving variable temporal influences can be very complicated in a cohort or similar model.

In an IBM, there is no difficulty resolving whatever level of spatial or temporal resolution is of interest, and heterogeneity at the selected level of resolution is incorporated explicitly in the model. Any environmental influence requiring a "memory" of past conditions (e.g., thermal or toxic stress, feeding history) is easily represented. Nonlinearities in responses do not result in distortion of distributions of properties. Events occurring at the individual level, such as movement, growth, or death, are summed to arrive at the population response.

There are significant advantages in the individual-based approach: clarity and consistency of logic; unambiguous "currency" of the model (i.e., individual fish); ease of tracking movements and adding new features (e.g., energetic and genetic effects, interactions); ease of accumulating effects of past conditions (e.g., toxic body burden and condition factor); and straightforward simulation of responses to a spatially and temporally heterogeneous environment.

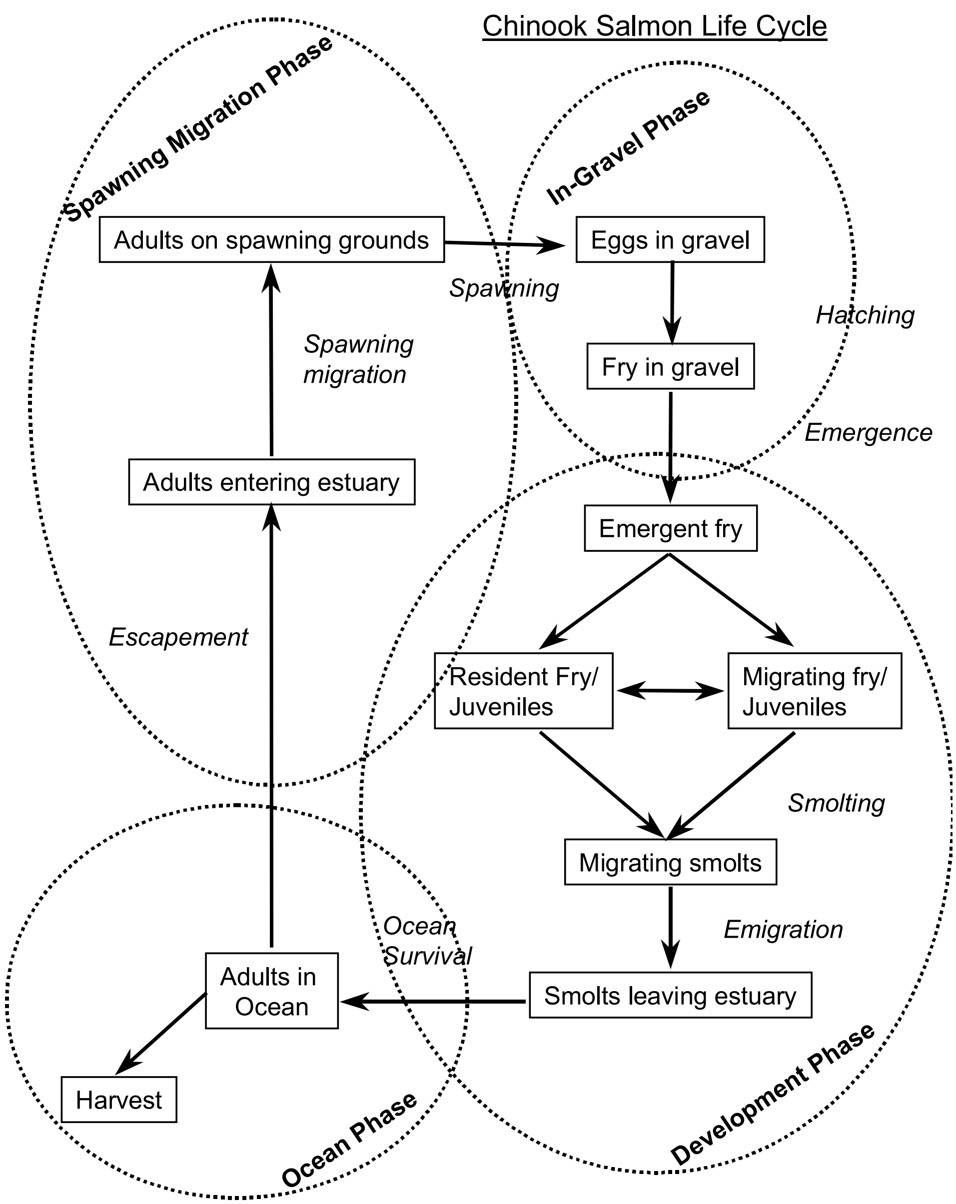

Figure 1 Key points in the life cycle of chinook salmon. The four oval areas represent the major life stages, represented by separate modules in the model. Arrows indicate a change of state of surviving salmon, with ocean harvest represented explicitly but other mortality not shown. Terms in italics indicate major life history events occurring in each stage.

The principal disadvantage of an IBM is that it is computationally intensive, and the computer power needed to run the model can be difficult to predict. Furthermore, simulating explicitly the hundreds of millions of fall-run juveniles in the Sacramento Basin would make the model unwieldy even with the fastest available computers. Therefore, the populations will be represented by a sample of the actual fish, and each model fish will be a "super-individual" representing some number of individuals (Scheffer and others 1995). This method, which is analogous to stratified sampling in opinion polls, should provide equivalent results to modeling every individual but at a manageable cost in computer time. It may be superior to the resampling method of Rose and others (1993), which can introduce bias if the number of model fish is too low (Scheffer and others 1995). It is also different from cohort modeling because sufficient numbers of sample fish are tracked to represent adequately the full range of variability in the population. The ratio of model to actual fish can be varied among life stages and races to keep the sample size large. Thus, abundant fall-run fry will be represented at a fairly small ratio of model to actual fish, while winter-run (and initially all) adults will be represented at a ratio of 1:1. Preliminary testing will ensure that the ratio selected does not bias the results. Clearly the selection of these ratios will represent a compromise between the speed at which the model runs and the amount of bias or error due to aggregation, and can change with type of model run and available computer power.

The individual-based approach lends itself directly to the use of object-oriented programming methods. In contrast to procedural languages (e.g., FORTRAN, C), an object-oriented language isolates elements of the program as "objects" which pass, receive, and respond to "messages." Objects may be any element of the program, but are most useful when they bear direct relationships to real objects, such as fish, river reaches, or computer windows. Thus, there is a direct correspondence between individuals in the model and objects in the program, making the transition from conceptual model to computer program as straightforward as possible. We have chosen to use the Swarm software package for multi-agent simulation of complex systems, developed by the Santa Fe Institute. This package comes with several ready-made objects and tools for input and analysis that will simplify coding and testing the model.

As noted previously, we have developed a draft conceptual model (Kimmerer and Jones & Stokes Associates 1998) and an annotated bibliography. We are proceeding on three parallel tracks in model development: (1) refining the conceptual model based on comments received, discussions with interested parties, and experience with submodels; (2) assessing the data available for model input; and (3) developing a model formulation in Swarm focusing initially on in-river life stages.

Principal Gaps in Knowledge

Significant advances have been made in understanding the biology of Sacramento Basin salmon since the previous population models were developed. However, our assessment of the available information gives little encouragement that the principal gaps have been filled. Although the model can be used to some degree to explore the consequences of different assumptions about these gaps, a lack of solid understanding may restrict use of the model for management purposes.

It is relatively easy to identify knowledge gaps, and several key ones are discussed below. However, a significant problem we have encountered in attempting to fill these gaps is that relatively few of the existing data are in the form of published reports or articles. Much of the information is either anecdotal, or has not yet been published or widely disseminated. Some data are presented in technical reports, but the data are not made available to the research community on a timely basis.

Thermal Effects Below Lethal Limits. When temperature exceeds lethal limits, mortality is expected to be rapid, but results of mark-recapture experiments in the Delta suggest effects at temperatures below these limits (Kjelson and Brandes 1989; Rice and Newman 1997). Although these effects could be artifacts from the use of hatchery smolts or other aspects of the experiments, it is also likely that similar effects apply to naturally-reared fish. If so, similar temperature effects should occur throughout the river system. They may arise through physiological changes that affect growth, disease resistance, predator avoidance, and smolting, through ecological effects such as increased predator activity or increased food requirement without an increase in supply, or through a combination of these effects. Since temperature in the system often varies within the range at which these effects seem to occur, these effects may be important influences on survival of young salmon. Available information on thermal effects, however, is largely confined to laboratory experiments on temperature above lethal limits, with abundant food (e.g., Brett 1952). The potential effects listed above remain virtually unexamined.

Abundance of Young Fish. There are reasonably good estimates of adult and redd abundance, although abundance of some adults has become more difficult to determine with the revised operation of the Red Bluff Diversion Dam, where dam gates are open during most of the upstream migration period. However, estimates of fry or smolt abundance in the rivers are uncommon, and although the data are available, estimates of abundance have not been made for the Delta. Many measurements of abundance in the river system provide only indices rather than actual abundance values. The problem is for many measures of abundance, no suitable method has been developed to calibrate the measures to the actual number of fish passing a point or residing in an area. Although these indices are adequate for comparing abundance data among years and investigating effects of local restoration actions, they fall short of the data needed to develop a comprehensive view of the salmon population. In particular, mortality values, essential for assessing population status, require accurate abundance estimates.

Availability of Rearing Habitat. Recent data suggest that most of the young salmon in some of the rivers leave their natal streams shortly after emergence (Snider and others 1997). Furthermore, beach-seining data show large numbers of salmon fry in the lower Sacramento River and the Delta (Kjelson and Raquel 1981; Brandes and McLain, this volume). This implies the existence of two very different life histories, that is, fish that rear largely in the natal streams and those that rear mostly downstream. The relative contribution to recruitment by these life histories needs to be assessed, and some effort needs to be made to determine the factors that induce the young salmon to migrate as early fry instead of rearing in the natal streams. This may relate to the carrying capacity of different parts of the system for rearing salmon, which may be a key element in density dependence and therefore population regulation (for example, Elliott 1989). However, existing data are insufficient to assess the importance of rearing in the natal river compared with the mainstem Sacramento River and Delta, the factors influencing the availability of rearing habitat, or the factors that stimulate movement of pre-smolt salmon. The principal issue is where and under what conditions the extent or quality of physical habitat limits the abundance or survival of rearing salmon. Although the model may be useful in testing the outcome of alternative conceptual models about rearing habitat, the ultimate answer to its importance must be obtained through hypothesis-driven field research. The importance of rearing habitat has obvious, large implications for the success of alternative restoration actions.

Survival of Young Salmon. A related issue about which little is known is survival during early life. Survival through hatching and emergence is at least qualitatively understood to be high except in cases of extreme changes in flow or high temperature. However, survival during rearing, seaward migration, and early ocean life is unknown, except for survival indices for smolts passing through the Delta. The location of rearing may have a big effect on survival: for example, density-dependent migration out of the natal stream combined with lower, density-independent survival in the Delta would result in density-dependent survival. Little is known about the influence of food supply on survival, nor is there good information on the abundance and activity of predators. Finally, the occurrence and locus of density dependence, a crucial ecological feedback to any biological population, is unknown; previous studies have shown evidence of density dependence in young salmonids both in streams (Neilson and Banford 1983; Elliott 1989) and in the ocean (Peterman 1984).

Filling these knowledge gaps will not be easy. Most of them would require a coordinated effort involving a variety of agencies and a long time frame. However, without this information the effects of restoration actions will be difficult to predict, and therefore the actions will be difficult to justify.

Filling the Knowledge Gaps

Why are these information gaps still present? We do not wish to understate the difficulty of gathering the kind of knowledge described above, nor to denigrate the efforts of the biologists investigating Central Valley salmon. Much of the difficulty lies with the complexity of the ecosystem and the populations to be investigated. Nevertheless, we believe there are some key impediments to filling these knowledge gaps, and removing or reducing these impediments may improve the rate at which the gaps are filled.

The first impediment is a mismatch in perception among modelers, fish biologists, and managers about what data are needed and how to use a model. Modelers tend to focus on the "big picture," with less attention to details and a tendency for excessive generalization. Fish biologists tend to have a deeper understanding of certain topics, but a narrower view, often constrained by their experience to certain aspects of geography or life history. Understandably, many fish biologists tend to view data needs in terms of their own research experience. Many managers prefer not to hear about uncertainty and tend to rely heavily on expert opinion or on well-presented (usually conceptual or statistical) models. Although managers often support status and trends monitoring, they may see little need for research aimed at fundamental questions, which can be expensive and risky. The perspectives of these three groups do not lend themselves to a coordinated attack on the key problems, because each group sees the key issues differently.

The second impediment is what we see as a lack of institutional commitment to resolving system-level uncertainties. Much of the work being done by fish biologists and other scientists in the system is focused on particular exigencies, mostly related to endangered-species protection. Thus, little time is available for consideration of larger issues. There is no agency whose mission is solely to investigate and understand the biology of salmon and the influences on it. Each of the resource agencies has significant other duties, particularly environmental or endangered-species protection, that may actually impede progress toward understanding at a system level. This impediment has been evident in the resistance of some agency biologists to adaptive management experiments designed to determine the effects of certain management actions on salmon populations when the experimental actions were seen as potentially (but not demonstrably) harmful.

There also seems to be a strong degree of territoriality in the Central Valley salmon biology field. Although the situation is improving (for example, with this Symposium), there is still a remarkable lack of collaboration among researchers. This situation is particularly alarming given the amount of work being done at public expense and the importance of salmon to the Central Valley's ecosystems and economy.

Several potential approaches may help to resolve these issues. The most direct is individual commitment by fish biologists to consider the "big picture" in what they do on a daily basis and to continually re-evaluate their contribution. Although such a commitment would seem consistent with the role and activities of scientists, it would be naive to expect individual scientists to deviate much from their immediate interests to the common good, at least without added incentive.

This indicates a need for institutional commitment to working toward answering large-scale questions. This commitment could be underwritten by one of the larger organizations (e.g., CALFED Bay-Delta Program, Comprehensive Monitoring, Assessment and Research Program, Interagency Ecological Program), but the individual agencies would still have to support the contributions of their own fish biologists to the larger view. This may be seen as contrary to the mission of resource agencies, which have immediate responsibilities for endangered-species protection and other activities that may preclude devoting adequate attention to large-scale issues. One mechanism for enlarging the view of agency biologists is to make publication in peer-reviewed journals a criterion for promotion. The process of preparing a paper and getting it through the review process is an excellent way of helping a researcher to put his or her work in a larger context.

An alternative method for filling gaps is to establish a small, dynamic research team whose sole mission would be to gather, analyze, and publish data specifically related to population-level issues. This team could be given the mandate to collect data from other researchers, and to initiate field research projects into areas outside of the interests of other agencies. Mechanisms would have to be established to ensure cooperation by agency biologists, and reciprocally to ensure partnerships between members of this team and agency biologists.

An additional aid to filling in knowledge gaps is to make data freely available. Although data are routinely published in annual and other reports, these data are not readily available to other researchers. Identifying and obtaining data has been one of the most time-consuming and frustrating activities in our modeling work. These data have been collected by public agencies with public funds, and the maximum possible use should be made of them. The prerogative of the investigators to publish their results can be upheld through a delay time of no more than one year from the date of collection to the date at which the data are made available on an Internet site. The salmon monitoring and research community would do well to follow the lead of the Interagency Ecological Program in terms of data dissemination and availability.

Regardless of the mechanism used, we urge managers and biologists to consider seriously the need for better use of the available information, better mechanisms for determining what information is gathered, and research targeted at a more comprehensive view of the biology and population dynamics of salmon.

In our model development to date, we have found it easy to identify significant gaps in the knowledge about salmon, as discussed above. No model runs were necessary to convince us that the gaps are serious impediments to understanding the complete life cycle of chinook salmon. As the simulation model is developed, we anticipate using sensitivity analysis to further delineate where significant gaps occur, and possibly to develop methods for filling the gaps. We hope that as this work progresses some of the impediments to knowledge discussed above can be removed, and progress can be made toward filling the gaps.

References

Baker PF, Speed TP, Ligon FK. 1995. Estimating the influence of temperature on the survival of chinook salmon smolts (*Oncorhynchus tshawytscha*) migrating through the Sacramento-San Joaquin River Delta of California. Can J Fish Aquat Sci 52:855–63.

Botsford LW, Brittnacher JG. 1996. Viability of a Pacific salmon population: the Sacramento River winter run chinook. Unpublished draft manuscript. Center for Population Biology, University of California, Davis.

Brandes P, McLain J. 2001. Juvenile chinook salmon abundance, distribution, and survival in the Sacramento-San Joaquin Estuary. In: Brown RL, editor. Fish Bulletin 179: Contributions to the biology of Central Valley salmonids. Volume 2. Sacramento (CA): California Department of Fish and Game.

Brett JR. 1952. Temperature tolerance of young Pacific salmon, genus *Oncorhynchus*. J Fish Res Board Can 9:265–309.

DeAngelis DL, Gross LJ, editors. 1995. Individual-based models and approaches in ecology. New York: Chapman.

EA Engineering, Science, and Technology. 1991. San Joaquin River system chinook salmon population model documentation. Prepared for Turlock and Modesto Irrigation Districts, California.

Elliott JM. 1989. Mechanisms responsible for population regulation in young migratory trout, *Salmo trutta*. I. The critical time for survival. Journal of Animal Ecology 58:987–1001.

Jager HI, Cardwell HE, Sale MJ, Bevelhimer MS, Coutant CC. 1996. Simulating the effects of instream flows on fall chinook salmon. Unpublished report. Oak Ridge (TN): Environmental Sciences Division, Oak Ridge National Laboratory.

Kimmerer W, Jones & Stokes Associates Inc. 1998. Conceptual framework for the Sacramento River chinook salmon individual-based model. Draft report prepared for the US Fish and Wildlife Service. Sacramento, California.

Kimmerer W, Hagar JM, Garcia J, Williams T. 1989. Chinook salmon population model for the Sacramento River basin, version CPOP-2. Tiburon (CA): Biosystems Analysis, Inc.

Kjelson MA, Raquel PF. 1981. The life history of fall run juvenile chinook salmon, *Oncorhynchus tshawytscha*, in the Sacramento-San Joaquin Estuary of California. Estuaries 4:285.

Kjelson MA, Raquel PF, Fisher FW. 1982. Life history of fall-run juvenile chinook salmon, *Oncorhynchus tshawytscha*, in the Sacramento-San Joaquin estuary, California. In: Kennedy VS, editor. Estuarine comparisons. New York: Academic Press. p 393–411.

Kjelson MA, Brandes P. 1989. The use of smolt survival estimates to quantify the effects of habitat changes on salmonid stocks in the Sacramento-San Joaquin Rivers, California. In: Levings C, Holtby L, Henderson M, editors. Proceedings of the National Workshop on Effects of Habitat Alteration on Salmonid Stocks. Can Spec Publ Fish Aquat Sci 105:100–15.

Neilson JD, Banford CE. 1983. Chinook salmon (*Oncorhynchus tshawytscha*) spawner characteristics in relation to redd physical features. Can J Zool 61:1524–31.

Peterman RM. 1984. Density-dependent growth in early ocean life of sockeye salmon (*Oncorhynchus nerka*). Can J Fish Aquat Sci 41:1825–9.

Rice J, Newman K. 1997. Statistical model for survival of chinook salmon smolts outmigrating through the lower Sacramento-San Joaquin system. Interagency Ecological Program Technical Report 59. Sacramento (CA): California Department of Water Resources.

Rose KA, Christensen SW, DeAngelis DL. 1993. Individual-based modeling of populations with high mortality—a new method based on following a fixed number of model individuals. Ecological Modelling 68:273–92.

Scheffer M, Baveco JM, DeAngelis DL, Rose KA, van Nes EH. 1995. Super-individuals, a simple solution for modelling large populations on an individual basis. Ecological Modelling 80:161–70.

Snider B, Titus RG, Payne BA. 1997. Lower American River Emigration Survey, October 1995 - September 1996. California Department of Fish and Game, Environmental Services Division, Stream Evaluation Program.

Speed T. 1993. Modelling and managing a salmon population. In: Barnett V, Turkman KF, editors. Statistics for the environment. New York: J Wiley & Sons. p 268–90.