

Dealing with uncertainty in ecosystem models: lessons from a complex salmon model

PAUL McELHANY,^{1,4} E. ASHLEY STEEL,¹ KAREN AVERY,¹ NAOMI YODER,¹ CRAIG BUSACK,² AND BRAD THOMPSON³

¹*National Oceanic and Atmospheric Administration, Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, Washington 98112 USA*

²*Washington Department of Fish and Wildlife, Natural Resources Building, 1111 Washington Street Southeast, Olympia, Washington 98501 USA*

³*U.S. Fish and Wildlife Service, Western Washington Office, 510 Desmond Drive Southeast, Suite 102, Lacey, Washington 98503 USA*

Abstract. Ecosystem models have been developed for assessment and management in a wide variety of environments. As model complexity increases, it becomes more difficult to trace how imperfect knowledge of internal model parameters, data inputs, or relationships among parameters might impact model results, affecting predictions and subsequent management decisions. Sensitivity analysis is an essential component of model evaluation, particularly when models are used to make management decisions. Results should be expressed as probabilities and should realistically account for uncertainty. When models are particularly complex, this can be difficult to do and to present in ways that do not obfuscate essential results. We conducted a sensitivity analysis of the Ecosystem Diagnosis and Treatment (EDT) model, which predicts salmon productivity and capacity as a function of ecosystem conditions. We used a novel “structured sensitivity analysis” approach that is particularly useful for very complex models or those with an abundance of interconnected parameters. We identified small, medium, and large plausible ranges for both input data and model parameters. Using a Monte Carlo approach, we explored the variation in output, prediction intervals, and sensitivity indices, given these plausible input distributions. The analyses indicated that, as a consequence of internal parameter uncertainty, EDT productivity and capacity predictions lack the precision needed for many management applications. However, EDT prioritization of reaches for preservation or restoration was more robust to given input uncertainties, indicating that EDT may be more useful as a relative measure of fish performance than as an absolute measure. Like all large models, if EDT output is to be used as input to other models or management tools it is important to explicitly incorporate the uncertainty and sensitivity analyses into such secondary analyses. Sensitivity analyses should become standard operating procedure for evaluation of ecosystem models.

Key words: *Ecosystem Diagnosis and Treatment (EDT) model; ecosystem model; Pacific Northwest, USA; salmon; sensitivity analysis; Sobol index; uncertainty.*

INTRODUCTION

Scientists and managers have increasingly recognized that conservation concerns, including endangered species management, need to be addressed in the context of entire ecosystems (Beechie et al. 2003, DEFRA 2005, Harssan et al. 2005). Concurrently, technological advances have enabled the development of increasingly complex mathematical and statistical models for describing these ecosystems and aiding conservation planning. In the realm of aquatic ecosystems, the EcoPath/EcoSim framework alone has been applied to several hundred marine, estuarine, and freshwater food webs (Pauly et al. 2000). A number of terrestrial ecosystem models have been developed to examine

forest dynamics (Chertov et al. 2006, Johnson et al. 2007). At an even larger scale, global climate change models (23 models in the latest report from the Intergovernmental Panel on Climate Change [IPCC] [Randall et al. 2007]) are a major research and policy focus with broad implications. All of these ecosystem models are diverse in terms of scope and approach, but share the general feature of a large number of parameters with complex interactions. These models are necessarily built with imperfect information. Model inputs such as environmental data or population abundance estimates are often fraught with errors both known and unknown. Given these inevitable uncertainties, large and complex ecosystem models must be evaluated through sensitivity analyses before their output can be effectively applied to conservation problems (Hilborn and Mangel 1997, Saltelli et al. 2000b, Regan et al. 2002, Clark 2003, Harwood and Stokes 2003, Pielke and Conant 2003, Tang et al. 2006).

Manuscript received 1 April 2008; revised 1 April 2009; accepted 20 April 2009; final version received 16 June 2009.
Corresponding Editor: S. S. Heppell.

⁴ E-mail: paul.mcelhany@noaa.gov

Sensitivity analysis is a systematic exploration of how model results change in response to changes in the model input and parameters. It can improve the interpretation of modeled output and aid in reducing model complexity. Knowing which parameters have little influence on model output and whether parameter interactions are additive (i.e., parameters are independent), synergistic (i.e., parameters combine for an exaggerated effect on output), or compensatory (i.e., parameters “cancel each other out” for reduced effect on output) can help identify which parameters can be removed from the model without changing its predictive capabilities and where to focus additional field sampling resources for improved model precision. Since model inputs are themselves only estimates, we need to understand how model output, and therefore management recommendations, might change as a result of alternative reasonable estimates.

Despite the need for sensitivity analyses, tools and technologies for evaluating large ecosystem models have lagged behind model development and usage. Technical challenges have certainly played a role in this lag, but an additional factor may be reluctance by both managers and modelers to fully incorporate uncertainty into analyses, for fear that wide prediction intervals will render a model less valuable as a management tool. We argue that information about prediction intervals and model sensitivity will make a model more valuable and sensitivity analysis should become standard operating procedure. In our experience, managers often have no idea what to do with uncertainty estimates, so it is important that modelers communicate uncertainty results clearly and with examples of how uncertainty may affect decision making.

In this paper, we used a novel, structured approach to conduct a full sensitivity analysis of a large, complex, and widely used model employed for managing Pacific salmon populations. The Ecosystem Diagnosis and Treatment (EDT) model (Lichatowich et al. 1995, Moberand et al. 1997) is used to estimate salmon abundance and productivity and as an approach for identifying and prioritizing the manner in which recovery funds should be allocated to specific habitat restoration and preservation actions. The model contains a large number of parameters characterizing both the environment and relationships between the environment and salmon, most of which are estimated with considerable uncertainty. For example, the model includes parameters describing the level of total pollutants in the water column and the sublethal effects of those pollutants on salmon; neither the amount nor effect of pollution can be estimated with much precision. Although there have been some other efforts at partial sensitivity analysis of EDT (Steel et al. 2009), the uncertainty in input parameters has not been explicitly taken into account and the EDT model currently provides only point estimate predictions. We argue that

current model output is unsatisfactory for making informed management decisions.

As model complexity increases, it becomes more difficult to trace how imperfect knowledge of internal model parameters or data inputs might impact model predictions, causing many modelers to give the problem short shrift. A thorough sensitivity analysis often requires a Monte Carlo approach in which the model is run many times, with each run using random samples drawn from the input parameter probability distributions. This process can be extremely time-consuming for complex model applications; our approach reduces this limitation. We provide prediction intervals for key model outputs, such as estimated salmon capacity, and compare the sensitivity of modeled output to groups of different types of parameters. Our goals in this analysis were twofold. First, we wanted to develop and implement a method for evaluating the EDT model as a means for improving endangered salmon population management. Second, we wanted to use the EDT analysis as the basis for a general evaluation of how uncertainty analysis can be applied to large ecosystem models to make better management decisions.

Pacific salmon

Pacific salmon are an ideal species for building and evaluating ecosystem models due to their complex life history strategies and known dependency on habitat quality, which is the focus of most recovery efforts (Quinn 2004). There are six species of Pacific salmonids in the Pacific Northwest (PNW; Washington, Oregon, Idaho) of the United States that are predominantly anadromous and semelparous, *Oncorhynchus tshawytscha* (chinook), *O. kisutch* (coho), *O. mykiss* (steelhead), *O. nerka* (sockeye), *O. keta* (chum), and *O. gorbuscha* (pink) (Groot and Margolis 1991), and two other species, *O. clarki* (cutthroat) and *Salvelinus confluentus* (bull trout), that are less frequently anadromous (Quinn and Myers 2007). Pacific salmon are of considerable cultural and economic importance to communities in the PNW, particularly native communities. Because of habitat degradation, dam construction, overfishing, hatchery production, and other threats, many salmon populations in the PNW and in California are currently listed as threatened or endangered under the U.S. Endangered Species Act (Good et al. 2007).

As a consequence, extensive efforts are underway to recover threatened populations and protect existing healthy populations. In response to salmon population declines, a great deal of money has been and will likely continue to be spent on actions to restore and protect critical freshwater habitat for salmonids. Between 2000 and 2006, the federal Pacific Coastal Salmon Recovery Fund alone allocated over US\$500 million to salmon recovery (NOAA 2007), and millions more from state, local, and private sources have been spent to recover salmon. The types of restoration and management actions that are prescribed for salmon recovery are

often very precise and on small scales, which means that models to predict the effects of those actions or prioritize them must be on similar spatiotemporal scales (and hence, very complex).

The EDT model

The Ecosystem Diagnosis and Treatment model was developed by Mobrand Biometrics (now Mobrand, Jones and Stokes) to predict salmon performance (primarily abundance and productivity) as a function of ecosystem conditions (primarily freshwater habitat) in order to identify environmental features for protection or restoration. The EDT model also provides estimates of population characteristics for cases in which little or no empirical fish data exist (Lichatowich et al. 1995, Mobrand et al. 1997, Lestelle et al. 2004, Mobrand Biometrics 2005). The model is proprietary and we did not have access to the code used to implement the mathematical algorithms. However, in cooperation with Mobrand Biometrics, we obtained a compiled copy of the program, which allowed us to manipulate the internal parameters.

The EDT model has been used in over 100 watersheds in the PNW to evaluate hundreds of individual salmonid populations. The evaluations include an estimate of population abundance, productivity, capacity, and diversity under four different environmental scenarios. The evaluations also include an assessment of which freshwater habitat attributes in which reaches are most important for either restoration or preservation. This reach/attribute prioritization might be expressed, for example, as “The greatest increase in chinook salmon abundance would likely occur if stream substrate embeddedness were reduced to historical levels between river km 7 and river km 9.” Both the fish performance and reach/attribute prioritization predictions have been widely incorporated into PNW salmon and ecosystem management plans (e.g., Lower Columbia Fish Recovery Board 2004, Shared Strategy for Puget Sound 2007). The predictions of fish performance have been used to inform recovery goals and harvest management decisions and the reach/attribute prioritizations have been used to help guide the funding of recovery actions. Recently, the EDT model results have been widely used as inputs to the All H Analyzer (AHA) salmon life cycle model (Hatchery Scientific Review Group 2009a). The AHA model evaluates the impact of habitat, hatcheries, harvest, and hydrosystem operations (the “H’s”) on salmon populations. Through this AHA model, EDT is anticipated to have additional influence on the prioritization of management actions for threatened and endangered salmon (Hatchery Scientific Review Group 2009b).

A recent paper by Blair et al. (2009) provides an overview description of the EDT model and its conceptual underpinnings. An electronic library of EDT documentation is available from Mobrand, Jones

and Stokes (MJS) (*available online*).⁵ Here, we provide a brief overview of the model, focusing on the mechanics of the equations that are needed to understand the sensitivity analysis. A summary of parameter groups in the EDT model is provided in Table 1.

The EDT model is an ecosystem model taken from the perspective of a single salmonid species. The species life cycle is broken into a number of life stages, and survival and capacity of each stage are characterized by a Beverton-Holt function in which the number of recruits (R) is a function of the number of spawners (S), the intrinsic productivity (a), and the maximum number of recruits (b):

$$R = \frac{aS}{1 + \frac{a}{b}S}. \quad (1)$$

These life-stage-specific Beverton-Holt functions are coupled together to estimate the overall productivity and capacity for the population (Moussalli and Hilborn 1986). The life stage productivity and capacity estimates are generated in EDT as a function of the condition of the ecosystem. The EDT model describes the condition of the ecosystem by first partitioning the freshwater habitat into stream reaches, which are then characterized by the status of 43 specific attributes, such as maximum temperature or percentage of pools. To estimate the survival of a particular life stage in a particular reach, the survival of that life stage under ideal conditions (the “benchmark” parameter) is decremented based on a function describing the extent to which existing habitat conditions reduce productivity from the ideal (these are the productivity “rule” parameters). A somewhat similar process is used to estimate river-reach specific life stage capacities in which the capacity is reduced from ideal benchmark densities based on habitat condition. Marine survival is modeled assuming density-independent mortality.

For a given population, there is usually a large amount of variability in spawning location, travel speed, and life stage transition timing (Groot and Margolis 1991, Quinn 2004). The EDT model describes the different potential life history paths through the landscape (i.e., the amount of time a population stays in a particular river reach at a particular life stage) as “trajectories.” To estimate the impact of habitat quality on the total population in the watershed, the EDT model estimates an average productivity and capacity estimate across a large number (about 500) of randomly selected trajectories. A number of model parameters define temporal windows for life stage transitions and ranges of travel time that are used in a stochastic process to produce the sample of possible trajectories. The population abundance and capacity for each trajectory are calculated assuming that the trajectory describes a single population. To obtain the overall population

⁵ (<http://www.mobrand.com/MBI/library.html>)

TABLE 1. Parameters in the Ecosystem Diagnosis and Treatment (EDT) model.

| Parameter and subparameter groups | User access | No. individual parameters |
|--|-------------|----------------------------|
| Habitat attributes | | |
| Habitat attributes | accessible | thousands |
| Habitat types | accessible | hundreds to thousands |
| Gradient | accessible | tens to low hundreds |
| Off-channel factor | accessible | tens to low hundreds |
| Channel length | accessible | tens to low hundreds |
| Width minimum and maximum | accessible | tens to hundreds |
| Benchmarks | | |
| Productivity | Mobrand | ~10 |
| Density | Mobrand | ~10 |
| Life stage duration | Mobrand | low tens |
| Growth factor | Mobrand | low tens |
| Rules | | |
| Key habitat weights | Mobrand | high hundreds to thousands |
| g value | Mobrand | 1 |
| Factor | Mobrand | thousands |
| Exponent | Mobrand | thousands |
| Month weights | accessible | hundreds to thousands |
| Food multiplier curve | Mobrand | low tens |
| Fraction of life history types | accessible | <10 |
| Reach data global | | |
| Length | database | tens |
| Gradient | database | tens |
| Month weights | database | tens to low hundreds |
| Width minimum and maximum | database | tens |
| Adult age | | |
| Ocean survival multiplier | database | <10 |
| Ocean age | database | <10 |
| Fecundity | database | <10 |
| Sex ratio | database | <10 |
| Juvenile age (steelhead only) | | |
| Marine survival multiplier | database | <10 |
| Proportion smolt | database | <10 |
| Life stage pattern and population description | | |
| Spawn window | accessible | tens |
| Life stage window | accessible | tens to low hundreds |
| Life stage duration | accessible | tens to low hundreds |
| Smolt migration | Mobrand | <10 |
| Transitional season | Mobrand | 1 |
| Integration method | | |
| Productivity | database | 2 |
| Capacity | database | 2 |
| Trajectory seed | Mobrand | 1 |

Notes: The groups and subgroup categories were the units for sensitivity analysis. Parameters with “accessible” user access could be readily modified by users via a graphical interface, those with “database” access could be modified by the model developer within the program database at user request, and those with “Mobrand” access were set by the model developer. The number of individual parameters in each of the subparameter groups for a single population application is presented as a range because parameter number varies by watershed and species.

productivity and capacity, all of the individual trajectory populations are combined using one of four methods (or “integration types”).

The model is typically run under four environmental scenarios: (1) current habitat condition with ocean harvest, (2) current habitat condition without ocean harvest, (3) historical habitat condition, based on habitat reconstruction, and (4) a “degraded” habitat condition. These scenarios differ in the reach-specific habitat attribute parameter values and whether a

density-independent harvest mortality is included in adult survival. The model identifies highest priority reaches for restoration as those reaches that would result in the greatest improvement in population abundance and productivity if changed from current conditions to historical conditions, while all other reaches remain at current conditions. High-priority reaches for preservation are identified as those reaches that would cause the greatest reduction in population abundance and productivity if changed from current conditions to degraded

conditions, while holding all other reaches at current conditions. In a similar comparison of current, historic, and degraded scenarios, the model is also used to prioritize specific habitat attributes (e.g., maximum temperature) in specific reaches for restoration or preservation, but we have not conducted sensitivity analyses on these attribute prioritizations.

The EDT model has three classes of input parameters based on the whether they can be modified by model users (Table 1). The “user modifiable via interface” parameters can be readily entered and edited by EDT users through a graphical interface. These parameters include the reach-specific habitat attributes. The “user modifiable in database” parameters may be adjusted at user request but must be manually changed by Mobrاند Biometrics in the program database. Examples of this type of parameter are the density-independent adult ocean survivals. The “Mobrاند-defined” parameters are based on estimates from Mobrاند Biometrics and are not modifiable by the model users. These parameters include the benchmarks and productivity rules. This distinction becomes important when considering model sensitivity and whether model precision is best improved by reducing uncertainty in user-modifiable or user-inaccessible parameters.

METHODS

The goals of the uncertainty and sensitivity analyses were to understand the extent to which the model output might vary given the uncertainties in the inputs and parameters and, further, which of the input or parameter uncertainties causes the most uncertainty in model predictions. We used prediction intervals to describe the uncertainty in model output under a range of realistic input uncertainty scenarios and sensitivity indices to describe which parameters were responsible for the greatest prediction uncertainty. For our analysis, we looked only at the uncertainty that arose when we did not have exact information on the model inputs or parameters, i.e., precision. We were not testing whether or not the model is accurate (i.e., whether the model makes predictions that are true). This is a critical issue; prediction intervals and sensitivity indices describe the behavior of the model, not whether the model correctly describes the behavior of the real world.

To generate the large number of input data sets and conduct the Monte Carlo models runs described here, we developed a Java computer program that interacted with the compiled EDT program from Mobrاند Biometrics.

Input distributions

Our analyses required estimating plausible input parameter error distributions. This was a challenging step in the evaluation process. Example input distributions used for our analyses and a description of all the input parameter distributions used for the analyses are found in the Appendix. The EDT input parameters

came from a number of sources. Ideally, the inputs would all have been based on measurement techniques with quantitative estimates of precision (e.g., sample counts with precision $\pm X\%$). More commonly, inputs were derived from a synthesis of literature values, extrapolation from data collected in neighboring regions, techniques with unknown precision, or simply expert opinion. For estimating input distributions for the reach-specific habitat attribute parameters, we utilized the results of a survey of eight biologists with EDT experience (Busack and Thompson 2006). The respondents were asked to provide a maximum and minimum habitat score for a particular attribute, given its point estimate value and the attribute’s “level of proof.” In the EDT model, a level of proof value can be assigned to each habitat input parameter where level of proof describes the general method used to estimate the parameter (e.g., empirical data or expert opinion). The “level of proof” is a qualitative description of the input precision. The survey was used to convert this qualitative description into a quantitative estimate precision for habitat attribute parameters. The survey included 1520 questions, and the results were used to generate triangular distributions around the point estimate values for reach-specific habitat attribute parameters in our analysis.

No “level of proof” ratings were available for the non-habitat parameters. To estimate input distributions for these attributes, we relied on the expertise of a group of biologists with experience in developing and using EDT and in salmon ecology (see author list of Steel et al. [2009]). Since there was little, if any, quantitative analysis of precision for many of the parameters (e.g., the productivity rules), we often used a simple uniform distribution with (\pm) some percentage error, where the percentage error value used was deemed a plausible level of error by the group. To explore the manner in which the analysis results would differ with different input distribution assumptions, we considered our best estimate as a medium error distribution and also explored smaller and larger errors. A small error was half the range of the medium error and a large error was twice the range of the medium error.

For the Monte Carlo runs in our analysis, parameters were independently drawn from the distributions of the individual parameters. However, there was some filtering that removed parameter combinations for which EDT would not generate productivity or capacity estimates. These included parameter combinations with impossible life history trajectories or capacity-survival characteristics. Examples of excluded trajectories were scenarios in which random draws from migration timing windows placed a life stage in a location deemed by the model structure to be impossible.

Prediction intervals

We generated prediction intervals by drawing random values from the plausible input distributions to create a

large number (hundreds) of alternative plausible data sets that included input data and internal model parameters (hereafter referred to as inputs). We ran each of these data sets through the EDT model and derived the prediction interval from the resulting distribution of model outputs (hereafter referred to as outputs). For example, a 90% prediction interval on population abundance would be the abundance range containing 90% of the output model predictions. For generating output prediction intervals, we included probability distributions for all of the available EDT input parameters (see Table 1 and Appendix).

Structured sensitivity analysis

The purpose of the sensitivity analysis was to determine how much uncertainty in each of the input parameters contributed to the overall uncertainty in outputs. Given the large number of parameters in the EDT model it would not have been practical or particularly informative to evaluate the sensitivity of every single parameter. Therefore, we took a structured approach to the sensitivity analysis (Fig. 1). A number of parameters were initially grouped together and analyses were conducted to determine the sensitivities of each of these groups. Once the sensitivities to these initial groups were determined, we then conducted another analysis in which the most sensitive groups were broken up into smaller subgroups to determine the most sensitive components. The initial groups that showed less sensitivity were either pooled into a new single group or fixed at point estimate values. This process of grouping and regrouping was repeated several times to identify individual parameters or groups containing only a small number of parameters that had the most influence on model precision.

For the initial analysis, the input parameters were divided into 13 groups (Table 1). These 13 groupings were based on both logical groups and on the internal structure of the EDT program. Consequently, some of the groups, such as habitat attributes and productivity rules, contained many individual parameters, while other groups, such as food multiplier or trajectory seed, contained only one or a very few parameters. Defining all EDT parameters is beyond the scope of this paper, and the reader is referred to the EDT documentation (see Blair et al. 2009) for more details.

One-at-a-time (OAT) analysis

In an OAT analysis, the focal parameter, or group of parameters, is varied while all other parameters and groups are held at their point estimate values (Saltelli et al. 2000a). The OAT analyses are local sensitivity analyses because they explore sensitivities locally around the point estimate inputs. Since our analyses were conducted on groups of parameters rather than individual parameters, many of the quantitative local OAT indices methods, such as simple regression, were not

available. Therefore, we present OAT results simply in the form of prediction intervals and graphs of distributions generated by varying only the focal input parameter group.

Global variance partitioning sensitivity analysis

To identify the input parameters with greatest effect on output precision, we applied a global sensitivity analysis method based on the work of Sobol (Sobol 1993, Saltelli et al. 2000a, Fieberg and Jenkins 2005, Tang et al. 2006). With a global sensitivity analysis, the variability in the model output is partitioned among groups of input parameters in an approach analogous to an analysis of variance (ANOVA) in experimental design (Table 2). A global sensitivity analysis recognizes that all of the parameters in a model are estimated with error. The sensitivity index for a focal parameter is calculated by integrating the variation in model output caused by variation in model input across the distribution of all possible parameter values for the other parameters. The analysis is “global” in that the sensitivity indices consider the entire joint parameter space. As with ANOVA, the total variability in the model prediction can be broken down into main effects and interaction effects among the parameters. The Sobol method allows identification of “main effects” and “total effects,” in which total effects include the main effect for a given parameter and all the interaction terms of which that parameter is a part. If the model is additive, the main effect and total effect indices will be equal. Otherwise, the difference in the indices is a measure of the interaction between the focal parameter and all other parameters. The main effect sensitivity index of a parameter can be interpreted as the expected proportion that the model output variance could be reduced if that parameter alone were known without error. The total effect sensitivity index for a particular parameter can be interpreted as the total proportion of variance that would remain if all the parameters except the focal parameter were known without error. In other words, the main effect describes how much one could improve the prediction precision if one were to know only the focal parameter, and the complement of the total effects describes the best improvement in precision that would be possible if one doesn’t know the focal parameter. Knowing both of these is important for deciding how to prioritize monitoring for improved model precision since they may differ from one another if there are significant interactions among parameters.

Applying the method of Sobol requires a relatively large number of model runs, on the order of $N(2K + 2)$, where N is the number of model runs needed to estimate the output variance (several hundred) and K is the number of model parameters or parameter groups (Saltelli 2002). As noted above, a given application of EDT has thousands of parameters. A single run of the model takes between two and 100 minutes, depending

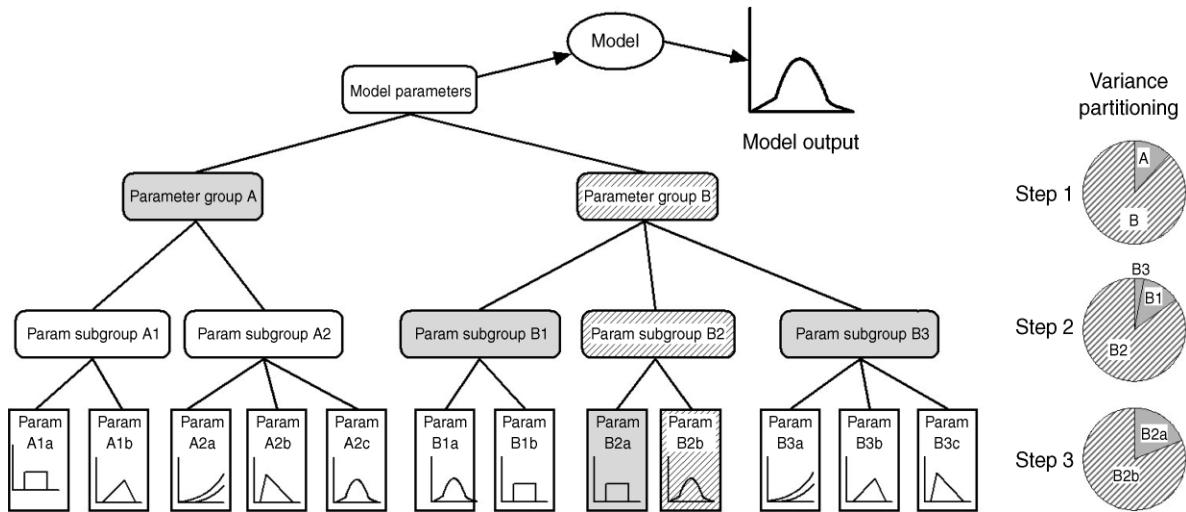


FIG. 1. Conceptual diagram of the structured variance partitioning approach, which is used for sensitivity analysis of models with a large number of parameters. Model parameters are first organized hierarchally. Variance partitioning sensitivity analysis is then applied, starting at the highest level in the hierarchy. The parameter group that has the greatest influence on output variance at a given level (pie charts) is examined in the next stage of the hierarchy. This is repeated at all stages to drill down to the individual parameters having the most influence on model output variability. The most influential parameter group and parameter in this example is indicated with hatching. Other parameter groups and parameters used in sensitivity analysis are shown with solid gray shading. Each of the individual parameters has an estimate input distribution, as shown in the thumbnail graphs.

upon the watershed. The structured sequential approach using parameter groups allows us to reduce the computational requirements to a (barely) plausible level.

Populations evaluated

For this analysis, we explored uncertainty and sensitivity of the EDT model as applied to three salmon populations; East Fork Lewis River fall chinook, Germany Creek coho, and West Fork Washougal River mainstem steelhead. These populations are located in the Lower Columbia River region of Washington State. The baseline, or point estimate, parameterization of the EDT model for these populations was conducted by the Washington Department of Fish and Wildlife (Lower Columbia Fish Recovery Board 2004). These populations were selected to provide a diversity of species and watersheds in the analysis. The populations were also

selected because they have a relatively small number of EDT reaches and therefore required less computer time.

RESULTS

Prediction intervals

When all input parameters were varied based on plausible distributions, the EDT fish performance outputs were highly variable and generated wide prediction intervals (Table 3). The most likely output values differed substantially from the EDT point estimate (Fig. 2). There was some variation in reach prioritization caused by input uncertainty, but in general, there was less variability in the reach prioritization results than for the performance measures. For the East Fork Lewis chinook, high priority preservation reaches tended to be within one or two rank values of one another and corresponded well to the point estimate

TABLE 2. Example of global variance partitioning sensitivity analysis, where *V* is variance and *A*, *B*, and *C* are model input parameters.

| Global sensitivity analysis term | Equation |
|--|--|
| Total model output variance | $V_{total} = V_A + V_B + V_C + V_{AB} + V_{AC} + V_{BC} + V_{ABC}$ |
| Main effects sensitivity index for parameter <i>A</i> | $S_A = \frac{V_A}{V_{total}}$ |
| Total effects sensitivity index for parameter <i>A</i> | $S_{A_total} = \frac{V_A + V_{AB} + V_{AC} + V_{ABC}}{V_{total}}$ |

Notes: Variance in the model output is decomposed using the Sobol method into the variance contribution from each of the input parameters and their interactions. The main effects sensitivity index is the fraction of the variability attributable to the focal parameter alone, and the total effects sensitivity index is the fraction of variability attributable to the focal parameter and all of the interactions involving the focal parameter.

TABLE 3. Prediction intervals for Ecosystem Diagnosis and Treatment (EDT) output with different populations, input errors, and parameter groups varied.

| Population | Input error | Parameters | Productivity (recruits/ spawner at low abundance) | | | Capacity (no. fish) | | | Equilibrium abundance (no. fish) | | |
|-------------------------|-------------|------------------------|--|--------|------------------|---------------------|--------|------------------|-------------------------------------|--------|-----------------|
| | | | Mean | Median | 80% interval | Mean | Median | 80% interval | Mean | Median | 80% interval |
| Germany coho | small | all | 5.47 | 5.24 | 2.55–8.89 | 955 | 877 | 386–1560 | 771 | 707 | 238–1338 |
| Germany coho | medium | all | 3.64 | 3.19 | 1.45–6.42 | 326 | 288 | 110–585 | 225 | 185 | 35–877 |
| Germany coho | large | all | 2.09 | 1.17 | 0.00–5.63 | 125 | 37 | 1–379 | 77 | 2 | 0–253 |
| East Fork Lewis chinook | small | all | 2.43 | 2.04 | 1.30–4.22 | 819 | 708 | 265–1554 | 498 | 341 | 70–1164 |
| East Fork Lewis chinook | medium | all | 4.14 | 3.21 | 1.17–8.47 | 1281 | 983 | 245–2666 | 941 | 635 | 35–2274 |
| East Fork Lewis chinook | large | all | 2.06 | 0.25 | 0.00–6.79 | 185 | 28 | 1–319 | 114 | 0 | 0–159 |
| East Fork Lewis chinook | medium | adult age | 4.79 | 4.30 | 1.98–7.95 | 3237 | 3073 | 1204–5407 | 2564 | 2376 | 591–4737 |
| East Fork Lewis chinook | medium | benchmark | 3.11 | 3.08 | 2.29–3.96 | 2214 | 2221 | 1577–2818 | 1501 | 1499 | 896–2093 |
| East Fork Lewis chinook | medium | food multiplier | 3.32 | ... | 3.32–3.32 | 2338 | ... | 2106–2548 | 1633 | ... | 1471–1780 |
| East Fork Lewis chinook | medium | habitat | 3.31 | 3.30 | 2.97–3.67 | 2267 | 2264 | 2127–2418 | 1578 | 1580 | 1436–1736 |
| East Fork Lewis chinook | medium | integration type | 3.11 | 2.91 | 2.91–3.32 | 2205 | 2333 | 2054–2333 | 1492 | 1531 | 1348–1630 |
| East Fork Lewis chinook | medium | life history | 3.32 | ... | 3.26–3.35 | 2338 | 2333 | 2296–2377 | 1633 | 1630 | 1609–1661 |
| East Fork Lewis chinook | medium | month pattern | 3.32 | ... | 3.26–3.38 | 2294 | 2307 | 2196–2369 | 1603 | 1618 | 1533–1646 |
| East Fork Lewis chinook | medium | pattern life stage | 3.31 | 3.27 | 2.72–3.89 | 2013 | 1950 | 1542–2620 | 1406 | 1353 | 988–1921 |
| East Fork Lewis chinook | medium | population description | 3.32 | ... | 3.08–3.56 | 2301 | 2319 | 2214–2353 | 1606 | 1628 | 1495–1692 |
| East Fork Lewis chinook | medium | reach data global | 3.30 | ... | 3.27–3.32 | 1373 | 1455 | 694–1791 | 957 | 1014 | 483–1250 |
| East Fork Lewis chinook | medium | rules | 3.40 | 3.36 | 1.97–4.87 | 1820 | 1774 | 1375–2291 | 1248 | 1228 | 737–1788 |
| East Fork Lewis chinook | medium | trajectory seed | 3.49 | 3.48 | 3.18–3.81 | 1920 | 1918 | 1692–2140 | 1370 | 1367 | 1181–1570 |
| Washougal steelhead | small | all | 2.92 | 2.58 | 0.92–4.92 | 139 | 125 | 78–234 | 83 | 72 | 1–149 |
| Washougal steelhead | medium | all | 2.58 | 1.93 | 0.23–5.60 | 81 | 61 | 21–153 | 46 | 25 | 0–115 |
| Washougal steelhead | large | all | 1.10 | 0.01 | 0.00–3.60 | 24 | 2 | 0–54 | 16 | 0 | 0–23 |

Notes: The productivity, capacity, and equilibrium are as defined for the Beverton-Holt function. In analyses highlighted in boldface, the range of the 80% prediction interval is larger than the mean value. The three salmon populations were Germany Creek coho, East Fork Lewis River fall chinook, and West Fork Washougal River mainstem winter steelhead.

rankings for the highest priority reaches (Fig. 3). The East Fork Lewis chinook restoration prioritizations were less consentient, but the top three priority reaches could be identified. The Germany coho and Washougal steelhead ranking outputs showed a bit more variability in prioritizations than the East Fork Lewis chinook, and there was some switching of ranks among the top four reaches (see Appendix).

Note that the term “prediction interval,” as used here, was a measure of model precision, not model accuracy. We simply do not know from this analysis whether the true value (e.g., abundance) was likely to be contained by the prediction interval because we have not evaluated model accuracy by comparing predictions to empirical data.

Prediction intervals and distribution graphs can be developed when the outputs are a series of single values (e.g., population performance values such as adult abundance or juvenile productivity). However, for the stream reach prioritization (restoration and preservation scenarios) an output was a list of ranks, not a single value. To display the output distribution from the reach rankings, we developed bubble graphs (Fig. 3) in which the size of the bubble indicates the number of output values for a reach at that rank.

The OAT results

There was considerable variation in the output distributions among parameter groups (Fig. 4). Parameter groups that tended to show relatively narrow output

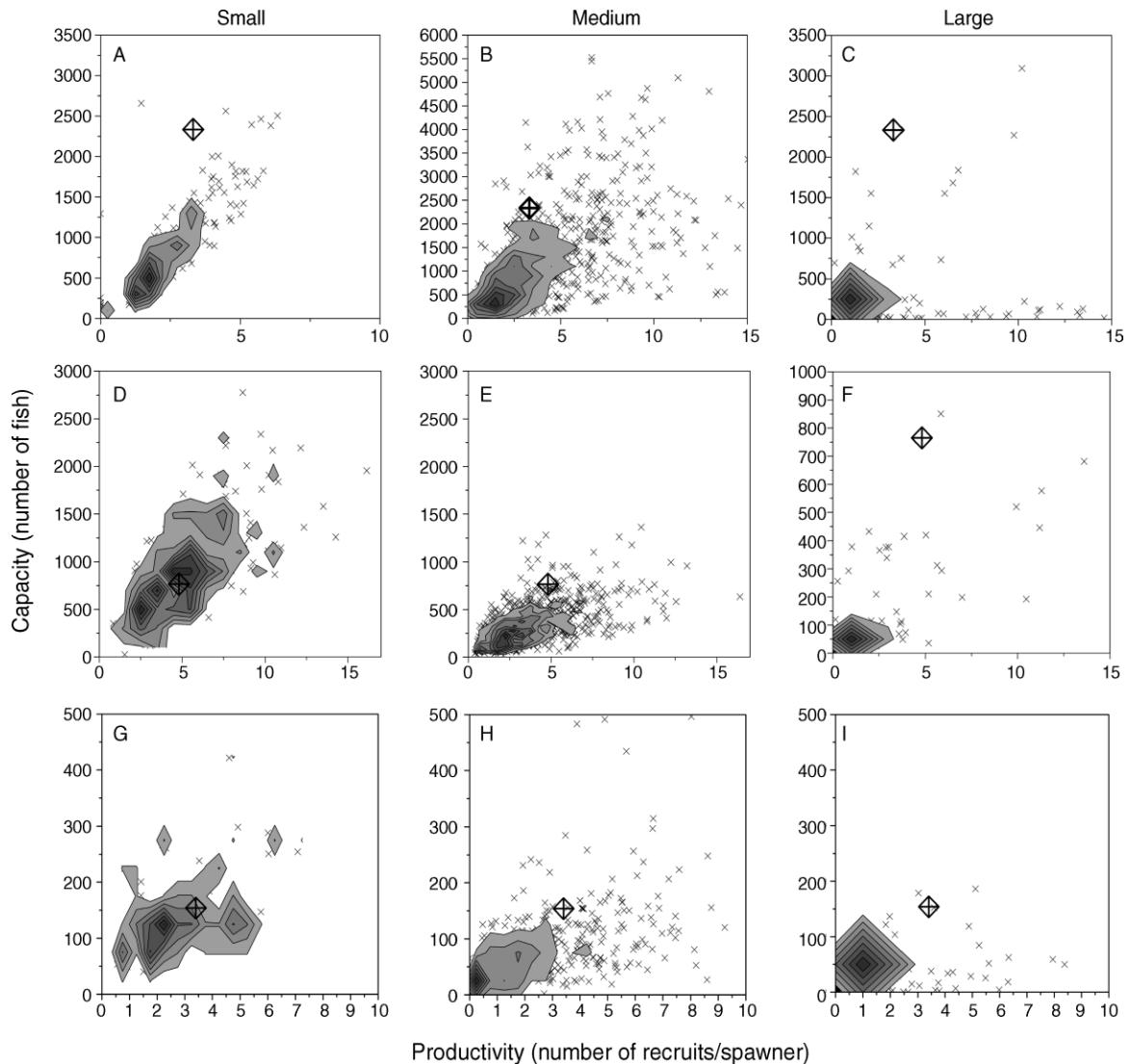


FIG. 2. Monte Carlo results for Ecosystem Diagnosis and Treatment (EDT) capacity and productivity output (at low abundance, current without harvest) for three salmon populations with small, medium, and large input errors for all model parameters: (A, B, and C) East Fork Lewis River fall chinook salmon; (D, E, and F) Germany Creek coho salmon; (G, H, and I) West Fork Washougal River mainstem winter steelhead salmon. Capacity and productivity are based on the Beverton-Holt function definitions as used in EDT. The 'x's are the individual outputs from the model runs and show the range of outputs and specific outliers, but point overlap in the area of most frequent occurrence prevents clear visualization of the output distribution. The gray-scale contour shows the frequency of capacity and productivity output from the simulations, with darker areas showing the most common capacity and productivity combinations. The diamonds are the EDT point estimates generated using the point estimate inputs. Some of the data points of the EDT models run are omitted from view in the figures because the results were outside the range of the axes presented. The Monte Carlo sample sizes for panels A–I are: (A) 247; (B) 894, 14 data points omitted from view; (C) 218, three data points omitted from view; (D) 198; (E) 967; (F) 84; (G) 80; (H) 476, nine data points omitted from view; (I) 209, three data points omitted from view.

distributions included the trajectory seed, habitat attributes, food multiplier, monthly pattern, life history percentage, and global reach data. Parameter groups that tended to show higher levels of variation included the adult parameters, productivity rules, life stage pattern, benchmarks, and integration type. The populations differed in which parameter group produced the greatest ranges in output.

Global sensitivity analysis

The results of the global sensitivity analysis are displayed in a pie chart in which the relative contribution of each parameter is indicated by the size of the wedge (see Saltelli et al. 2000b). The global sensitivity indices for the parameter groups in the three populations are shown in Fig. 5. For the East Fork Lewis River fall chinook population, the primary (>10%) main effect

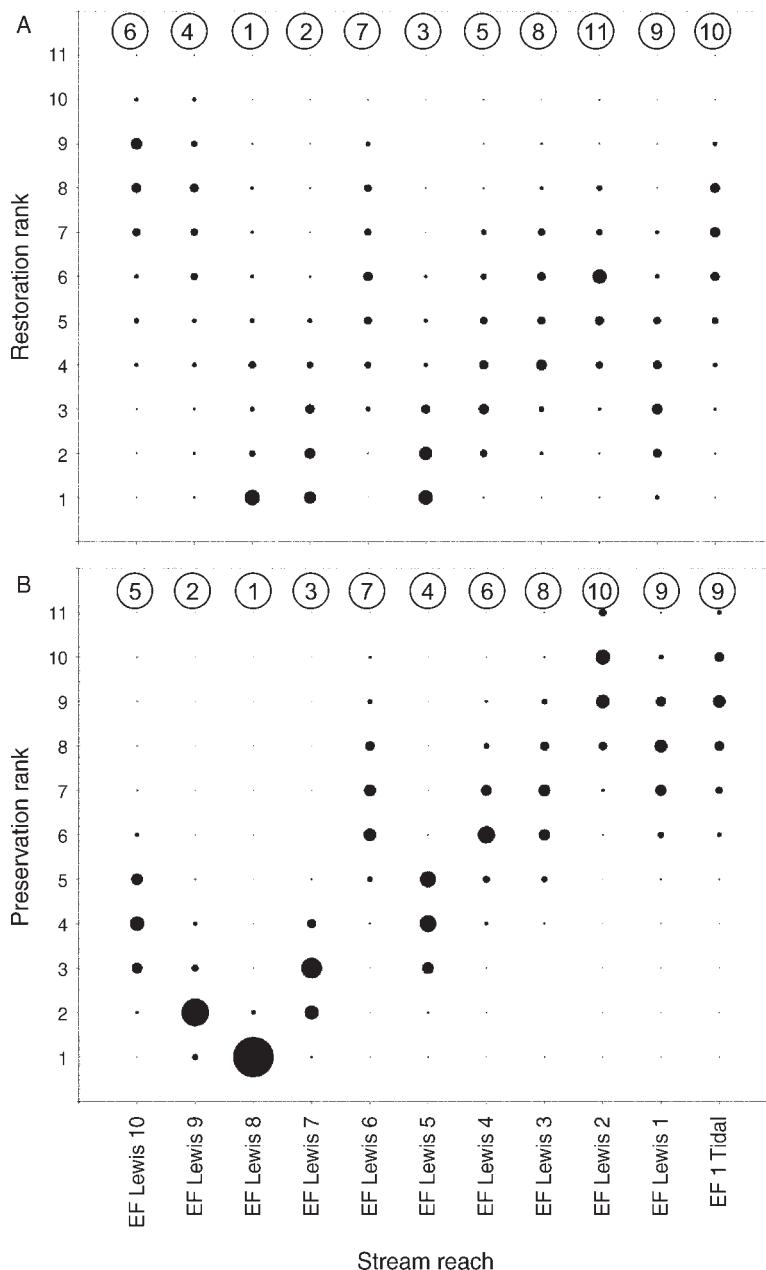


FIG. 3. Bubble graphs of Ecosystem Diagnosis and Treatment (EDT) (A) reach restoration and (B) preservation prioritizations for East Fork Lewis River fall chinook salmon using medium input error. This output is for the “combined” EDT rank metric, which is the mean of the diversity index, productivity, and abundance ranks. For a given reach, the bubbles show the relative number of simulations in each rank category. Rank “1” is the highest priority for restoration or preservation. If a reach has a single large bubble at a particular rank, then all of the simulations gave that reach the same rank. If a reach has a number of smaller bubbles spread out over many ranks, then the reach was ranked in many different categories in different simulations. The numbers in the circles indicate the EDT prioritizations based on point estimate inputs. The analysis was based on 476 Monte Carlo simulations.

parameter groups contributing to output variance in productivity were the adult parameters, benchmarks, and rules and for capacity output variance, the primary contributors were adult parameters, benchmarks, and the global reach data. For the Germany Creek coho population, the primary parameter groups contributing

to output variance in productivity were adult pattern life stage and rules and for capacity output variance, the primary contributors were pattern life stage and global reach data. For the West Fork Washougal River mainstem steelhead population, the primary parameter groups contributing to output variance in productivity

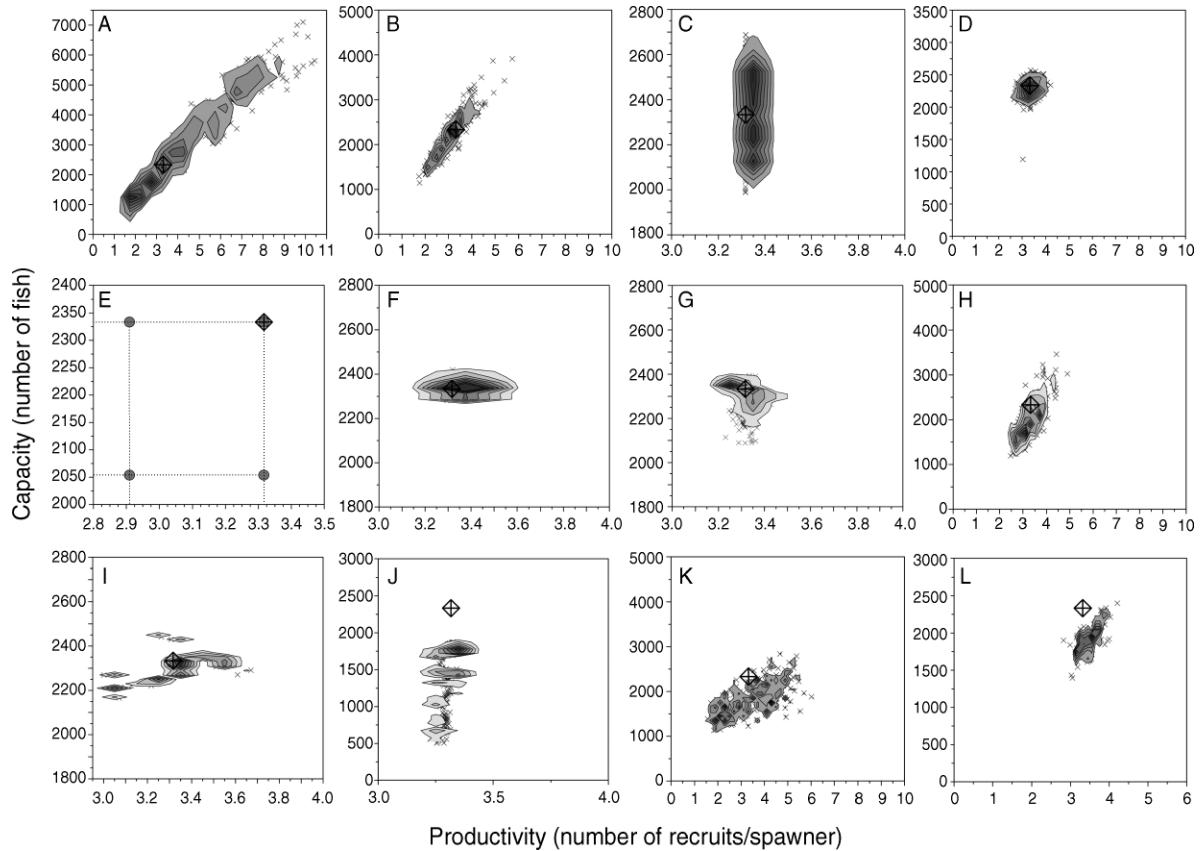


FIG. 4. Monte Carlo one-at-a-time (OAT) results for Ecosystem Diagnosis and Treatment (EDT) capacity and productivity output (at low abundance, current without harvest) for the East Fork Lewis River fall chinook salmon with medium input errors. Only one parameter group at a time was varied in each panel, with all other parameters held at their point estimates. The parameter groups are (with sample sizes in parentheses): (A) adult parameters (250); (B) benchmarks (250); (C) food multiplier (250); (D) habitat attributes (250); (E) integration type (250); (F) life history (250); (G) monthly pattern (250); (H) life stage pattern (214); (I) spawning window (250, one data point omitted from view); (J) global reach data (250); (K) productivity rules (250); and (L) trajectory seed (151). The \times 's are the individual outputs from the model runs and show the range of outputs and specific outliers, but point overlap in the area of most frequent occurrence prevents clear visualization of the output distribution. The gray-scale contour shows the frequency of capacity and productivity output from the simulations, with darker areas showing the most common capacity and productivity combinations. The diamonds are the EDT point estimates generated using the point estimate inputs.

were the juvenile parameters, life history parameter, and rules. For capacity output variance, the primary contributors were integration type, juvenile parameters, and life history parameters. In the three populations, the parameter groups food multiplier, habitat, monthly pattern data, population description, and trajectory seed were never major contributors to the output variance for either productivity or capacity.

Fig. 6 shows variance partitioning results for sub-parameters of the groups that were the primary contributors to variance for the East Fork Lewis chinook population: adult parameters, benchmarks, and rules. Within the adult parameter group, the ocean survival parameters had the greatest influence on EDT productivity estimate variability. For the benchmark group, the benchmark productivity parameters had the greatest influence. In the rule parameter group, the most

influential parameter was the “*g* value.” The *g* value is a single number, 0.37, that controls the amount of “synergy” as different habitat attributes are combined to affect survival. Fig. 6 indicates some interaction among the parameter groups, as shown by the difference between the main effects and total effects results. Interactions appeared particularly strong among the rule parameters.

DISCUSSION

Our EDT sensitivity analysis provides nonintuitive information for managers about appropriate and inappropriate uses of the model, biases in mean model predictions, and where increased data collection could most improve the model design and precision. It is possible (using a structured analysis) and necessary for

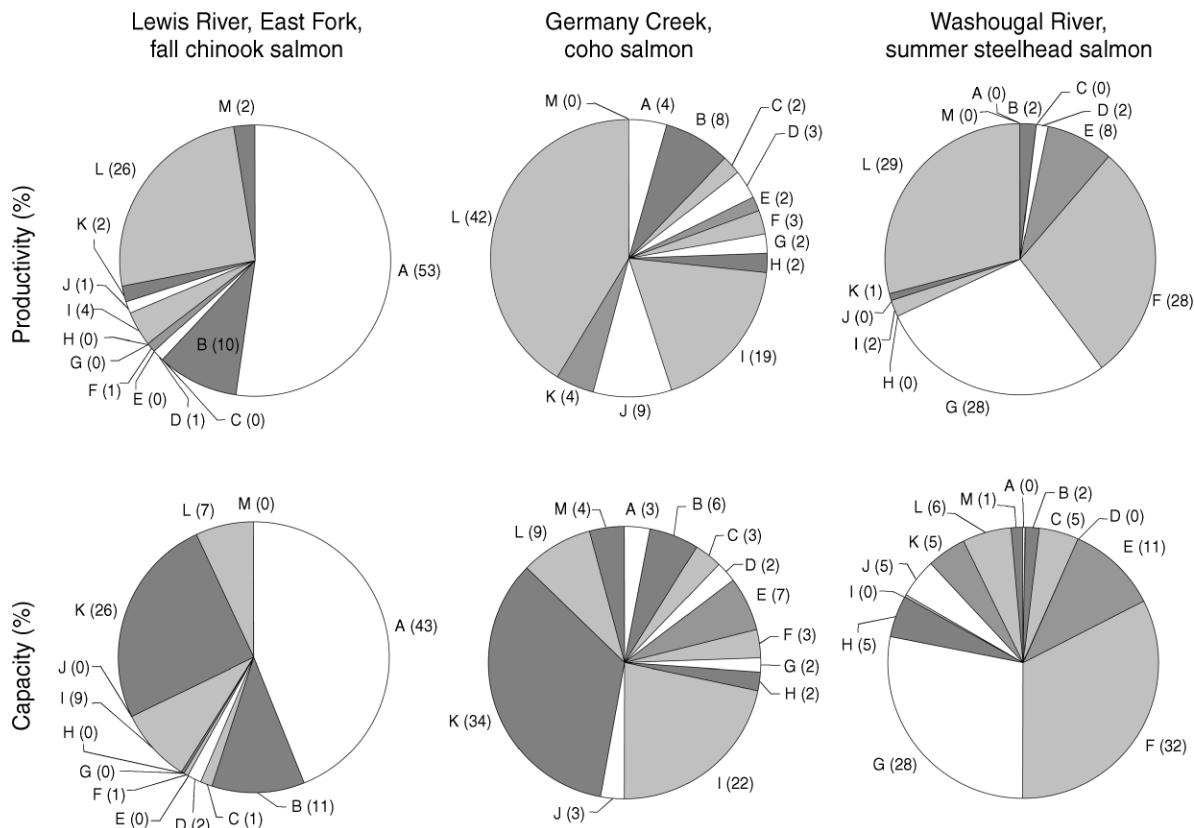


FIG. 5. Variance partitioning global sensitivity analysis of Ecosystem Diagnosis and Treatment (EDT) productivity output (current conditions without harvest) for three salmon populations: East Fork Lewis River fall chinook, Germany Creek coho, and West Fork Washougal River mainstem steelhead. Model input parameters were drawn from the “medium” error distributions. The size of the pie slice indicates the fraction of the output variance attributable to uncertainty in the input parameter group. This figure shows the main effects of each parameter group (i.e., no interaction effects). In the pie charts, the sensitivity indices are normalized to give relative percentage of contribution of each of the input parameter groups. This analysis used 13 parameter groups labeled A, adult age; B, benchmarks; C, food multiplier; D, habitat; E, integration type; F, juvenile age; G, life history; H, monthly pattern data; I, pattern life stage; J, population description; K, reach data global; L, productivity rules; and M, trajectory seed.

this type of information to be regularly generated as part of ecosystem model development.

Understanding the EDT prediction intervals

The prediction intervals for the fish performance measures (productivity, capacity, and abundance) show a relatively large range in model output. This may not be surprising given the levels of uncertainty in many of the input parameters. The output distributions also show a marked shift in the mean relative to the EDT output based on point estimates, which may not be as intuitive. The likely reasons for this are twofold. The first reason is nonlinearity between inputs and EDT predictions. Jensen’s inequality states that a nonlinear function (convex or concave) evaluated at the expected value of the input distributions will not necessarily equal the expected value of the function evaluated over the input distribution (Jensen 1906). In other words, the shift could arise from the nonlinearity of EDT. Although a shift in the mean due to nonlinearity is not surprising,

the magnitude of the effect observed here was not anticipated.

The second process contributing to the shift in mean could be asymmetries in the input distributions. The asymmetries in input distribution arise because many of the input parameters are bounded. For example, the percentage of a reach that is pools must be between 0 and 100%, and if a reach is estimated at 100% pools, any error could only include the possibility that the plausible percentage of pools value is lower, not higher. This would produce an asymmetry in the input error distribution that could shift the output distribution in a particular direction. The population could be at a lower bound for some parameters and the upper bound for others. These effects could tend to cancel one another out and not necessarily lead to a directional shift in the output, but there may be an overall pattern that affects the output mean.

The shift in mean is greatest for the model runs in which all of the input parameters are varied compared to

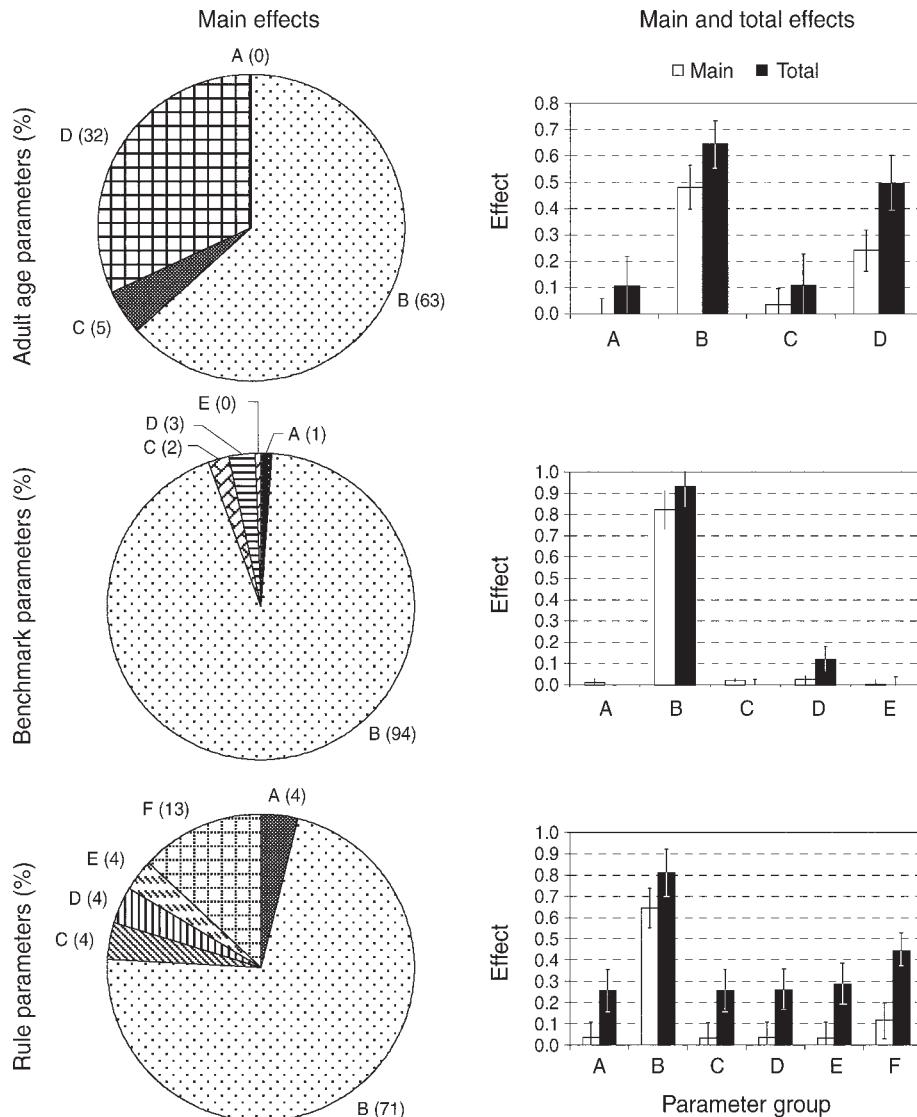


FIG. 6. Variance partitioning global sensitivity analysis of Ecosystem Diagnosis and Treatment (EDT) productivity output (current conditions without harvest) for East Fork Lewis River chinook salmon. This figure shows result for subparameters within three of the larger parameter groups (adult parameters, benchmarks, and rules). Model input parameters were drawn from the “medium” error distributions. The size of the pie slice indicates the fraction of the output variance attributable to uncertainty in the input parameter group. In each pie chart, the largest wedge belongs to the parameters not in the group being analyzed, so we only discuss the remaining subparameters in the paper. This figure shows both main effects and total effects (i.e., main parameter effects plus interactions). The bar graphs show the actual sensitivity index estimate and 90% bootstrap intervals on the index. In the pie charts, the sensitivity indices are normalized to give relative percentage of contribution of each of the input parameter groups. For adult parameters: A, eggs; B, non-adult parameters; C, ocean age; and D, ocean survival. For benchmark parameters: A, density; B, non-benchmark parameters; C, life stage duration; D, productivity benchmark; and E, space. For rule parameters: A, constant; B, non-rule parameters; C, exponent; D, factor 1; E, factor 2; and F, *g* value (a single number, 0.37, that controls the amount of “synergy” as different habitat attributes are combined to affect survival).

model runs in which only one parameter group at a time is varied as in the OAT analysis. Potentially, there is an interaction among the parameter groups that is increasing the shift in mean. This interaction is not measured by the global variance partitioning. The variance partitioning is directed at parsing the variance observed in the outputs generated when all the parameters are varied,

not in understanding the shift caused by Jensen’s inequality.

So, which value, the point estimate or the Monte Carlo output mean, is the “best” to present as the EDT prediction for most likely fish performance? If the input distributions truly reflect the possible values of the input parameters, the Monte Carlo output is likely a better

reflection of the expected results of the model. However, we are not confident that we fully understand the input distributions, so it remains an open question as to how best to characterize the expected value of the model prediction.

In matters that require a relatively high degree of confidence, such as harvest management or endangered species goal setting, we urge caution in the use of EDT abundance and productivity predictions. Also, if EDT is to be used as input to other models, such as AHA, it is important to explicitly incorporate the uncertainty in EDT productivity and capacity into the secondary analysis. The productivity and abundance output may be more useful as a relative measure than as an absolute measure. We suspect that the EDT productivity and capacity predictions may be useful as a metric for comparing relative potential among populations across basins, but we did not evaluate this possibility.

In contrast to the absolute fish performance metrics, the reach prioritizations for preservation and restoration seem relatively robust to the input uncertainty. Reach priorities tended not to shift by more than a couple of ranks and the most important reaches could be consistently identified. The restoration priorities are determined by comparing the abundance and productivity of a population under current habitat conditions to the abundance and productivity of a population under historical habitat conditions. Thus, the only parameters that vary in making the prioritization are the habitat parameters. The habitat parameters are estimated with relatively higher precision than many of the other parameters. Differences in reach habitat values for current vs. historical conditions persist despite the input uncertainty.

Information on uncertainty in reach prioritizations can be used to improve management decisions. A major task in managing endangered species is deciding how to allocate money for restoration and preservation. Funds are generally scarce and allocation often seems to be driven more by opportunity than by rigorous analysis. For example, funds may tend to be allocated preferentially to locations with willing land owners regardless of biological importance (Roni 2004). However, there is some discretion for managers, and EDT has been a primary tool for prioritization of salmon restoration activities in many PNW watersheds. Our analysis indicates that the EDT model provides relatively consistent prioritizations, even in the face of sizable input uncertainties. However, these prioritizations are not absolute. For example, based on the EDT model, the top three reaches to restore for recovery of fall chinook in the East Fork Lewis River would be reaches 5, 7, and 8, though there is a possibility that the top three would include reach 1 (Fig. 3). None of these top three reaches is clearly the highest priority, so within this set of reaches managers may want to give higher weight in prioritization to other factors, such as cost or political acceptability. This is in contrast to the results for

preservation of Lewis River fall chinook, where reach 8 is considered the highest priority under any plausible parameterization of the model (Fig. 3). This suggests that, based on EDT analysis, managers would have less room for discretion: to protect fall chinook, a method for safeguarding reach 8 must be found. Using the information from the uncertainty analysis, a formal decision framework could be developed that explicitly considers the probability of biological response as predicted by the model and the expected cost of proposed actions. Steel et al. (2008) provide an example of this type of decision analysis that also incorporates consideration of multiple models. Although evaluating precision of a single model is a vast improvement over the use of point estimates, the possibility that the model construct itself may be inaccurate must also be taken into account. The question of model accuracy can be addressed by comparing multiple models (as was done in Steel et al. [2008]) and, more directly, by model calibration.

Implications of the prediction intervals for model calibration

During development, the EDT model was not statistically calibrated in a formal fitting procedure. Instead, point estimate parameters were considered adequate by Moberg if EDT abundance and productivity predictions were “reasonably close” to empirical estimates of abundance and productivity derived from fish count data. Given the size and dimensionality of the potential parameter space, it would be possible to find many plausible parameters sets that would produce predictions considered “reasonably close” to the fish count predictions. It is a problem of fitting many (thousands of) parameters to only a few data points (fish count estimates); there is no unique solution. With this large plausible parameter space, it is quite possible to get a qualitatively determined reasonable fit with the wrong combination of input parameters.

The analyses we conducted bear on the potential for misparameterizing the model. Preliminary comparisons of the EDT output distributions to fish count data indicate that the EDT predictions based on plausible input likely encompass both reasonable and unreasonable fits to the fish count data. Because the plausible input space generates such a wide range of possible outputs, the model could be credibly parameterized in a way that was consistently biased high, biased low, or gave inconsistent results relative to the empirical fish count data, all the while being within the loosely defined bounds of “reasonable.” This is especially true because the fish count estimates themselves are often very imprecise (McElhany et al. 2007).

The range constituting reasonably close to empirical fish count data seems sufficiently broad that EDT could predict values with significantly different management implications and still be considered adequately parameterized using these standards. Depending upon how

model results are used, relatively small differences in predictions could have important management implications. For example, the Beverton-Holt equilibrium maximum sustainable harvest rate for a population with productivity of 2 is 29% and the harvest rate for a population with productivity 3 is 42%. When considering parameter uncertainty, the different estimates (i.e., productivity of 2 or 3) are well within the range of possible EDT outputs but would imply quite different harvest policies. To our knowledge, no one is directly using EDT to set harvest policy in this way, but some applications of EDT do seem to rely on this level of precision for making policy recommendations regarding populations for which no empirical fish count data are available (e.g., AHA applications; Hatchery Scientific Review Group 2009b).

Parameter sensitivity

In addition to the evaluation of prediction intervals, it is useful to know which of the input parameters are responsible for the majority of the output variability. If we know, for example, that the majority of the uncertainty in the output is caused by uncertainty in the effect of toxic chemicals on survival, we can focus efforts to improve our estimation of that parameter in order to reduce the range of the prediction interval. In some cases, it may not be practical to reduce the uncertainty in a sensitive parameter, indicating that model prediction uncertainty cannot be reduced. The identification of parameters that are insensitive is also useful. If reducing the uncertainty in an input parameter does not have much effect on the distribution of the output, there is little to be gained (from the perspective of reducing EDT model prediction uncertainty) by improving the precision of the parameter estimate. Identification of an insensitive parameter may also raise questions about the general behavior of the model if there is an expectation that fish performance should respond to changes in the parameter. For example, if the distribution of model output is little affected by the juvenile age structure, but it is reasonable to expect that the length of time fish spend in fresh water is important, further investigation of the model behavior is needed before applying model results to management decisions expected to affect freshwater residency time. Insensitivity may also occur when some other quantity or quality limits model output. In this case, the sensitivity analysis also informs our understanding of how the model works.

The input parameters having the greatest effect on the precision of abundance and productivity predictions were typically those that served as general scalars to the output (e.g., adult survival parameters, benchmarks, and the productivity rules' "g" parameter). These parameters describe ideal theoretical generic conditions (e.g., benchmarks), poorly estimated and highly variable values (e.g., adult survival), or somewhat arbitrary constants (e.g., "g") and, as a consequence, there is a

relatively high level of uncertainty in these input parameters. Discouragingly, it is not immediately clear how to improve the precision of these input parameters.

Habitat conditions had little effect on model output precision. This is largely a consequence of the relatively high precision with which habitat attributes are believed to be estimated relative to the other parameters in the model. Our result does not suggest that habitat is unimportant to salmon, nor does it suggest that model output is unaffected by the habitat parameters; fish need habitat, and if you make changes in habitat input, the model output generally changes. The results do suggest that improvements in the precision of the habitat attribute estimation are unlikely to do very much for improving the precision of the EDT abundance and productivity predictions unless precision in some of the other input parameters is substantially improved. It should be noted that improvement in EDT precision is obviously not the only motivation for increased collection of habitat data and we do not want to suggest that current habitat monitoring efforts are sufficient. The lack of sensitivity to habitat parameters is an interesting result because previous criticisms of EDT have focused on the model's reliance on expert opinion for habitat parameter estimates as a major weakness (e.g., Ruckelshaus et al. 2002), but our analysis indicates that habitat estimates are not the primary source of model imprecision for commonly used output.

Limitations of the analysis

As described in *Methods*, a key challenge in the analysis was estimating the input parameter distributions. Errors in estimating input distributions could affect the sensitivity analysis results. One particularly difficult issue to address was interactions among parameters. For our analysis, all input parameters were sampled independently, which is probably appropriate for most parameters. For example, the error in estimating stream temperature is likely to be independent of the error in estimating stream heavy metals contamination. However, the error distributions of some parameters may depend on which value is sampled from a different parameter. In other words, our independent sampling assumption may not define the appropriate input parameter space. Saltelli and Tarantola (2002) evaluated strategies for assessing nonadditive models with correlated inputs and show that modeling the correlation structure can impact the results of a sensitivity analysis. We suspect that this issue is minor in our analysis given the relatively modest errors assigned to many of the input parameters, but we did not evaluate this. We also note that the issue is somewhat managed because we deleted combinations of input parameters for which the model failed to compute productivity and abundance estimates. These failed runs served to trim the multidimensional parameter space.

The computational time required for the model runs severely limited the analysis. Even deploying 16 CPUs full time on the project, it has taken thousands of hours of computer time for these analyses. This constraint has prevented drilling down with more detail into specific parameters using the Sobol method. Including a correlation structure in the sampling distribution would have substantially increased computational times, even when incorporating shortcuts such as those suggested by Saltelli and Tarantola (2002). The Sobol method is very powerful, but it is computationally intensive.

Understanding the behavior of the model is a crucial step in understanding whether the model has the potential to provide useful insights into the workings of the real world. Knowing how the model behaves is crucial, but an equally crucial step is determining whether the model provides an accurate (i.e., true) description of the processes being modeled. Model validation is the process of evaluating whether model predictions match empirical data. Because of their complexity and the spatial scales involved, it is often even more difficult to validate an ecosystem model than it is to evaluate its sensitivity. The EDT model is no exception to this pattern.

EDT analysis summary

With slightly different yet plausible inputs, the EDT model could produce quite different results for two commonly used model outputs: capacity and productivity. However, the model was reasonably robust with respect to ranking stream reaches for preservation and restoration. With different plausible inputs, the model produced similar reach rankings, indicating the model may be more appropriate for relative prediction as compared to absolute predictions. Our confidence in these results is increased because we examined three populations of different species in different watersheds. It is important to note again that we are testing precision, not accuracy. The model predicts the same priority reaches given plausible input, but we have not evaluated whether those are the “right” reaches from a biological perspective. The model capacity and productivity output seems most sensitive to uncertainties in input from the “user-modifiable-in-database” (e.g., adult parameters) and “Mobrand-defined” parameters (e.g., rules), rather than the user-modified-via-interface parameters (e.g., habitat attributes). Knowing many of the user-modified parameters perfectly, such as the habitat attributes, would likely not greatly improve the precision of the capacity and productivity predictions. The analysis also indicated that many parameter groups (e.g., food multiplier) have little influence on model output and may potentially be removed from the model to reduce complexity without reducing predictive capability.

Lessons for evaluating complex environmental models

Many environmental questions can only be addressed using complex models involving hundreds or thousands

of parameters. Environmental models may need to include so many parameters because there are many interacting components or because the models need to integrate information over large spatial scales. As an example of a model with a large number of parameters both because of spatial scale and complex interactions, the Geophysical Fluid Dynamics Laboratory Coupled Model, version 2 (GFDL CM2) breaks the world in to a grid with 12 960 cells and contains many parameters associated with each cell (Delworth et al. 2006). Despite the inherent difficulties in structuring and parameterizing such models, the complexity is often necessary to understand the consequences of important interactions among different parts of the environment. However, the interactions make it challenging to predict model behavior over the potentially vast, yet plausible parameter space. If a model becomes so complex that it is impossible to determine how the results would change with different inputs, we have lost the ability to understand how the model works and how it might (or might not) reflect the actual ecosystem.

Major impediments to conducting sensitivity analysis on ecosystem models include the lack of information on input uncertainties and the sheer computational effort required for the analysis, given that a single model run may take hours or days. These limitations can be overcome using our structured approach, and our analyses demonstrate the value in addressing model sensitivity. Many of the input distributions are based on expert opinion rather than quantitative analysis, but even these crude approximations provide useful information for describing model precision. Estimating input distributions has the added advantage of making the state of knowledge of the system explicit.

Computational limitations are being continually reduced by faster computers, but much can also be done by building sensitivity analysis into the model design from the start. The widespread availability of distributional functions for most modeling environments (e.g., R, Matlab, C++, Java, Excel using @risk, etc.) makes it possible to add uncertainty analysis to most projects. In this analysis, we conducted a post hoc evaluation of an existing model that was not originally developed with sensitivity analysis in mind. A more efficient approach is to include sensitivity analysis from the initial conception of the model. The EDT model is by no means alone in not including sensitivity analysis; for example, many existing marine ecosystem models do not explicitly consider uncertainty, and sensitivity analysis must be evaluated post hoc (Plaganyi 2007). Models are often used to provide evidence for a specific thesis or the results of specific actions. Under this interpretation, a sensitivity analysis serves to provide transparency to the analysis and should be included as an important element of evidence building (Saltelli et al. 2000b).

Uncertainty and sensitivity analyses are a crucial step in model development (Saltelli 2002) as well as in model

calibration and validation. The uncertainty analysis requires an explicit description of input parameter error distributions. The uncertainty analysis results (i.e., prediction intervals) are also necessary for formal model validation. There is uncertainty in both sides of the validation process (i.e., in both the model and the empirical data), and the appropriate statistical evaluation is a comparison of the two distributions, not point estimates (for EDT example, see Rawding 2004). The analysis can be used to identify sensitive parameters in which improved precision really improves the model output precision (e.g., EDT adult survival parameters), and it can be used to identify parameters that could be held at a fixed value or perhaps eliminated from the model altogether (e.g., EDT food multiplier).

Evaluation of ecosystem models using input distributions rather than point estimates can reveal bias in the mean model predictions that arise from Jensen's inequality. This is important if managers are basing their decisions on what they perceive to be the most likely outcome predicted by the model. It is doubtful that a model run only with point estimate inputs will provide the best estimate of the mean model prediction.

The precision interval outputs from a sensitivity analysis provide users with quantitative information about model predictions that can inform how the model output is used in making decisions. These precision intervals can also enable model output to be honestly funneled into secondary models. Ultimately, to affect management decisions, ecosystem model results need to be communicated to policy makers. Even for complex models, developers can provide and managers should expect model results that include an assessment of prediction uncertainties and sensitivities. There is some concern on the part of both model builders and users alike that acknowledging uncertainty could result in decision-making paralysis. However, there is a need for complete transparency of the modeling process and recognition that the false sense of precision that models can generate may lead management down the wrong course. Realistic incorporation of uncertainty through sensitivity analysis can contribute greatly to our confidence in model results by indicating the most likely output values, rather than point estimates that imply a precision that does not exist.

ACKNOWLEDGMENTS

Thanks to the EDT collaboration group, Moberand, Jones and Stokes, NWFSC reviewers (especially Dave Jensen), and the Java programmers (Ray Crawford, Susan Demner, Bao Huynh, and Jerry Frost).

LITERATURE CITED

- Beechie, T. J., P. Roni, A. Steel, and E. Quimby. 2003. Ecosystem recovery planning for listed salmon: an integrated assessment approach for salmon habitat. Technical Memorandum NMFS-NWFSC-58. National Oceanic and Atmospheric Administration, Seattle, Washington, USA.
- Blair, G. R., L. C. Lestelle, and L. E. Moberand. 2009. The ecosystem diagnosis and treatment model: a tool for assessing salmonid performance potential based on habitat conditions. Pages 289–309 in E. Knudson, editor. Pacific salmon environment and life history models: advancing science for sustainable salmon in the future. American Fisheries Society, Bethesda, Maryland, USA.
- Busack, C., and B. E. Thompson. 2006. Improving salmon habitat productivity models for setting escapement goals and prioritizing habitat recovery funding. Washington Department of Fish and Wildlife, Olympia, Washington, USA.
- Chertov, O., A. Komarov, A. Loutkianov, A. Mikhailov, M. Nadporozhskaya, and E. Zubkova. 2006. The use of the forest ecosystem model EFIMOD for research and practical implementation at forest stand, local and regional levels. *Ecological Modeling* 194:227–232.
- Clark, J. S. 2003. Uncertainty in ecological inference and forecasting. *Ecology* 84:1349–1350.
- Delworth, T. L., et al. 2006. GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. *Journal of Climate* 19:643–674.
- Department for Environment, Food and Rural Affairs. 2005. Charting progress: an integrated assessment of the state of UK seas. Department for Environment, Food and Rural Affairs, London, UK.
- Fieberg, J., and K. J. Jenkins. 2005. Assessing uncertainty in ecological systems using global sensitivity analyses: a case example of simulated wolf reintroduction effects on elk. *Ecological Modeling* 187:259–280.
- Good, T. P., T. J. Beechie, P. McElhany, M. McClure, and M. H. Ruckelshaus. 2007. Recovery planning for endangered species act listed Pacific salmon: using science to inform goals and strategies. *Fisheries* 32:426–440.
- Groot, G., and L. Margolis. 1991. Pacific salmon: life histories. University of British Columbia Press, Vancouver, British Columbia, Canada.
- Harsan, R. M., R. Scholes, and N. Ash. 2005. Ecosystems and human well-being current state and trends: findings of the Condition and Trends Working Group of the Millennium Ecosystem Assessment. Island Press, Washington, D.C., USA.
- Harwood, J., and K. Stokes. 2003. Coping with uncertainty in ecological advice: lessons from fisheries. *Trends in Ecology and Evolution* 18:617–622.
- Hatchery Scientific Review Group. 2009a. Columbia River hatchery reform system-wide report: Appendix D. (http://www.hatcheryreform.us/hrp_downloads/reports/columbia_river/system-wide/4_appendix_d_user_guide.pdf)
- Hatchery Scientific Review Group. 2009b. Report to Congress on Columbia River Basin hatchery reform. (http://www.hatcheryreform.us/hrp_downloads/reports/columbia_river/report_to_congress/hsrg_report_12.pdf)
- Hilborn, R., and M. Mangel. 1997. The ecological detective. Princeton University Press, Princeton, New Jersey, USA.
- Jensen, J. L. W. V. 1906. Sur les fonctions convexes et les inegalites entre les valeurs moyennes. *Acta Mathematica* 30: 175–193.
- Johnson, K. N., S. Duncan, and T. A. Spies. 2007. Regional policy models for forest biodiversity analysis: lessons from coastal Oregon. *Ecological Applications* 17:81–90.
- Lestelle, L. C., L. E. Moberand, and W. E. McConaha. 2004. Information structure of Ecosystem Diagnosis and Treatment (EDT) and habitat rating rules for chinook salmon, coho salmon, and steelhead trout. Moberand Biometrics, Vashon Island, Washington, USA.
- Lichtowich, J., L. Moberand, L. Lestelle, and T. Vogel. 1995. An approach to the diagnosis and treatment of depleted Pacific salmon populations in Pacific Northwest watersheds. *Fisheries* 20:10–18.
- Lower Columbia Fish Recovery Board. 2004. Lower Columbia salmon recovery and fish and wildlife subbasin plan. Lower Columbia Fish Recovery Board, Longview, Washington, USA.

- McElhany, P., M. Chilcote, J. M. Myers, and R. Beamesderfer. 2007. Viability status of Oregon salmon and steelhead populations in the Willamette and Lower Columbia Basins. National Oceanic and Atmospheric Administration, Northwest Fisheries Science Center, Seattle, Washington, USA.
- Mobrand Biometrics. 2005. Documentation of the mathematical algorithms within EDT. Mobrand Biometrics, Vashon Island, Washington, USA.
- Mobrand, L., J. Lichatowich, L. Lestelle, and T. Vogel. 1997. An approach to describing ecosystem performance "through the eyes of salmon." *Canadian Journal of Fisheries and Aquatic Sciences* 54:2964–2973.
- Moussalli, E., and R. Hilborn. 1986. Optimal stock size and harvest rate in multistage life history models. *Canadian Journal of Fisheries and Aquatic Sciences* 43:135–141.
- NOAA [National Oceanic and Atmospheric Administration]. 2007. 2007 Report to Congress: Pacific coastal salmon recovery fund. National Oceanic and Atmospheric Administration, Silver Springs, Maryland, USA.
- Pauly, D., V. Christensen, and C. J. Walters. 2000. Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. *ICES Journal of Marine Science* 57:697–706.
- Pielke, R. A., and R. T. Conant. 2003. Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. *Ecology* 84:1351–1358.
- Plaganyi, E. E. 2007. Models for an ecosystem approach to fisheries. Food and Agricultural Organization Fisheries Technical Paper 477. Food and Agricultural Organization, Rome, Italy.
- Quinn, T. P. 2004. The behavior and ecology of Pacific salmon and trout. University of Washington Press, Seattle, Washington, USA.
- Quinn, T. P., and K. W. Myers. 2007. Patterns of anadromy and migrations of Pacific salmon and trout at sea. Pages 9–24 in N. M. G. Harris, editor. *Sea trout: biology, conservation and management*. Blackwell, Malden, Massachusetts, USA.
- Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichet, J. Fyfe, V. Kattsov, J. Shukla, R. J. Srinivasan, A. Sumi, and K. E. Taylor. 2007. Climate models and their evaluation. Pages 591–662 in S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. *Climate change 2007: the physical science basis*. Contributions of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.
- Rawding, D. 2004. Comparison of spawner-recruit data with estimates of ecosystem diagnosis and treatment (EDT) spawner-recruit performance. Washington Department of Fish and Wildlife, Vancouver, Washington, USA.
- Regan, H. M., M. Colyvan, and M. A. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12:618–628.
- Roni, P., editor. 2004. *Monitoring stream and watershed restoration*. American Fisheries Society, Bethesda, Maryland, USA.
- Ruckelshaus, M. H., P. Levin, J. B. Johnson, and P. M. Kareiva. 2002. The Pacific salmon wars: what science brings to the challenge of recovering species. *Annual Review of Ecology and Systematics* 33:665–706.
- Saltelli, A. 2002. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* 145:280–297.
- Saltelli, A., K. Chan, and E. M. Scott, editors. 2000a. *Sensitivity analysis*. John Wiley and Sons, Chichester, UK.
- Saltelli, A., and S. Tarantola. 2002. On the importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of the American Statistical Association* 97:702–709.
- Saltelli, A., S. Tarantola, and F. Campolongo. 2000b. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15:377–395.
- Shared Strategy for Puget Sound. 2007. Puget Sound salmon recovery plan. Shared Strategy for Puget Sound, Seattle, Washington, USA.
- Sobol, I. M. 1993. Sensitivity analysis for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments* 1:407–414.
- Steel, E. A., A. Fullerton, Y. Caras, M. Sheer, P. Olson, D. Jensen, J. Burke, M. Maher, and P. McElhany. 2008. A spatially explicit decision support system for watershed-scale management of salmon. *Ecology and Society* 13:50.
- Steel, E. A., et al. 2009. Making the best use of modeled data: multiple approaches to sensitivity analysis of a fish-habitat model. *Fisheries* 34:330–339.
- Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven. 2006. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences Discussions* 3:3333–3395.

APPENDIX

Details of input and results for sensitivity analysis of the Ecosystem Diagnosis and Treatment (EDT) model (*Ecological Archives* A020-013-A1).