A Strategy for Parameter Sensitivity and Uncertainty Analysis of Individual-based Models

Steven F. Railsback, Paul C. Cunningham, Roland H. Lamberson

Department of Mathematics, Humboldt State University, Arcata, CA 95521

Abstract

Parameter uncertainty and sensitivity analysis is especially important for large, complex individual-based models intended to support management decisions. Yet these models are difficult to analyze because they tend to have many parameters and long execution times. We define a three-phase analysis strategy. Phase 1 examines model sensitivity to each parameter by itself. Phase 2 identifies interactions in model response to a limited number of parameter pairs. Phase 3 examines how robust decision-support results are to parameter uncertainty: several management alternatives are defined and simulated, then the analysis looks at how often the model's ranking of the alternatives changes as a limited number of important parameters are perturbed. This strategy was applied to inSTREAM, an IBM that simulates effects of river management on trout populations. The analysis found no evidence of extreme sensitivity or "error propagation"; one parameter had effects that were stronger than anticipated but easily explained. Decision-support results of inSTREAM were highly robust to parameter uncertainty. Energetic parameters (for food intake and metabolism) were especially important, a result also found in other sensitivity analyses of large IBMs.

Key words: sensivity analysis, uncertainty analysis, robustness analysis, individual-based model, decision-support

1 Introduction

Analyzing the effects of parameter uncertainty on results is an important step in modeling, especially for large and complex models and for models used to make management decisions. The most complex models used for environmental management are now often individual-based models (IBMs); examples include the IBM of

Email addresses: LRA@Northcoast.com, PMC15@humboldt.edu, RHL1@humboldt.edu (Steven F. Railsback, Paul C. Cunningham, Roland H. Lamberson).
URL: www.humboldt.edu/~ecomodel (Steven F. Railsback, Paul C. Cunningham, Roland H. Lamberson).

habitat alteration effects on shorebirds by Goss-Custard et al. (2006), an IBM of how river flow fluctuations affect juvenile fish (Grand et al., 2006; this and other examples are also described in the online supplement to Grimm et al., 2006), and the trout IBM we use here. These large IBMs represent the physiology and behavior of individuals, and processes of the environment the individuals live in, using many equations and parameters. Consequently, potential clients of such models are naturally concerned about how robust results are to parameter uncertainty. The once-widespread belief that IBMs are inherently subject to “error propagation” (Mooij and DeAngelis 1999) also contributes to skepticism of their robustness to parameter values.

Sensitivity and uncertainty analysis can be thought of as having two major goals (Saltelli et al. 2000). First is providing understanding of the model: how are its outputs related to its assumptions, parameters, and inputs? Second is providing information on how robust model results are: given the uncertainties in its components, how much confidence should users have in model results? Parameter sensitivity and uncertainty analysis of complex simulation models is typically conducted by executing models many times while varying the parameter values (see, e.g., Rose 1989; Saltelli et al. 2000). Varying all parameters simultaneously allows analysis of model response to individual parameters and combinations of parameters.

Unfortunately, many of the characteristics that make IBMs useful for complex ecological and environmental management problems also make parameter analysis difficult (several of these characteristics were identified by Rose, 1989):

- Because they represent a variety of processes, IBMs typically have many parameters. Many of these are likely to have reliable values from laboratory research on individuals, field measurements of environmental processes, etc., but often some parameter values can only be reasonable estimates and others are highly uncertain.
- Many IBMs are computationally intensive so the feasible number of model runs is limited.
- IBMs can produce several different kinds of output that are each of interest (e.g., population abundance and biomass; size and age distributions; spatial distributions), and parameters can have different effects on different outputs.
- IBMs are usually stochastic, so effects of parameter values can be masked by “noise”.
- Model equations can be of any form, so model results cannot be assumed to vary linearly, or even continuously, with parameter values.
- In some IBMs, as in nature, different processes are important in different situations; e.g., a physiological process such as temperature stress may be very important when environmental conditions are stressful and completely unimportant in other situations. Hence, a parameter’s importance can be highly context-dependent.

As a consequence of these characteristics, standard parameter analysis strategies can be infeasible or incomplete for complex IBMs. Even if computation was not a limitation, a complete, traditional parameter sensitivity analysis could produce more information than is practical to analyze and understand. Few parameter sensitivity analyses of complex IBMs have been published, and those we found (Stillman et al. 2000; Shirley et al. 2003; Amano et al. 2006) only analyzed model response to each parameter by itself and did not analyze interactions among parameters or effects of parameter uncertainty on conclusions drawn from the model.

In this paper we define a general strategy for parameter sensitivity and robustness analysis of large, management-oriented IBMs. The strategy (described in Sect. 2) includes objectives and analysis methods that make a useful tradeoff between what we would like to know about a model and what is feasible to learn, and follows conventional sensitivity analysis approaches to the extent possible. We illustrate the strategy (Sect. 3) by applying it to inSTREAM, an IBM of stream trout designed to support river management decisions (Railsback and Harvey, 2001, 2002).

While we focus only on sensitivity and uncertainty of parameters (equation coefficients), methods similar to those we describe could also be used to analyze effects of model inputs (initial conditions, time-series habitat data, etc.). Analysis of structural uncertainty in IBMs (the effects of key assumptions) is discussed by Grimm et al. (2005) and in Ch. 9 of Grimm and Railsback (2005).

2 The Strategy

Our strategy for analyzing effects of parameter uncertainty on complex IBMs has three phases. While phases 2 and 3 use the results of previous phases, each phase has a distinct objective. The strategy is intended to provide a general understanding of how sensitive the model is to parameter values, identify individual parameters and parameter combinations that results are most sensitive to, and estimate how robust management-related conclusions drawn from the model are to parameter uncertainty.

We assume that, prior to Phase 1, all parameters have values estimated from the best available information (which can include, for some parameters, calibration of the model to observations). We refer to these as the “standard” parameter values.

2.1 Phase 1: Individual parameter sensitivity

The objectives of Phase 1 are to (1) determine how sensitive key model results are to each parameter, over the parameter’s full range of feasible values; (2) develop a general understanding of how robust model output is to parameter values; and (3) identify the parameters most important for further analysis in phases 2 and 3. In conventional approaches to sensitivity analysis of simulation models, the first two of these objectives are addressed (along with objectives of our Phase 2) by running the model many times while varying all parameters over wide ranges (Rose, 1989). Our Phase 1 is used to avoid the computational and analysis burdens of this conventional approach (Sect. 2.2); it evaluates sensitivity to each parameter separately so less-important parameters can be excluded from later phases.

The Phase 1 steps are:

1) Identify one (or a few) most-important model outputs to analyze. For management-oriented IBMs, these outputs are likely to be population-level summary statistics that are relevant to management questions such as population viability or production; an example is the abundance of reproductive adults, averaged over the entire simulation period from "census" data taken from the model once per simulated year. Output from early in the simulation can be excluded to keep the model's initial conditions from hiding effects of parameter values.

2) For each parameter, determine a range of feasible values. This step is critical and challenging. Analysis results will be highly dependent on the minimum and maximum feasible values selected here, and thought and judgment are required to identify useful values. Our experience indicates that each parameter should be examined carefully by people familiar with the information used to develop its standard value.

Simply varying all parameters over a consistent range (e.g., $\pm 50\%$ of the standard value) (e.g., Amano et al. 2006) seems straightforward and unbiased (Rose, 1989), but fails in at least two situations. First, some parameters are closely based on reliable data (e.g., from laboratory experiments on individual organisms), and the data can provide a much better estimate of the parameter's feasible range. For example, the data may show that the parameter value is very unlikely to be outside 5% of the standard value and values beyond 5% may produce absurd results (an example is in Sect. 3.1). Conversely, the data may show that the value is highly uncertain and the feasible range very large. The second situation is when parameter values are logically constrained. Survival probability parameters for risks such as predation are an example: the daily survival probability cannot possibly be greater than 1.0, and often is unlikely to be less than 0.95 (in which case half the population would be killed within 14 days). Hence, the feasible range of such a parameter is constrained to much less than $\pm 50\%$. (Ignoring this constraint resulted in a well-known example of absurd sensitivity analysis results, discussed by Mooij and DeAngelis, 1999.)

3) For a parameter, identify a limited number of values, spaced systematically over the parameter's range of feasible values. The same number of values are used for

all parameters, and should be high enough to keep analysis results from being dominated by stochastic noise, but not unnecessarily high because the required number of model runs for Phase 1 is equal to this number times the number of parameters analyzed. These parameter values could be spaced evenly over the parameter's range, but even spacing may not represent the distribution of values well if the parameter's standard value is not at the center of its range (e.g., if feasible ranges are defined as -50% to +100% of the standard value).

For each of these parameter values, also determine its value scaled to a range of 0–1, where 0.0 corresponds to the low end and 1.0 to the high end of the parameter's range of feasible values. For example, if five values are chosen for all parameters, and a parameter's selected values are evenly spaced over a range of 20 to 100, the scaled values for the parameter are 0, 0.25, 0.5, 0.75, and 1.0.

4) Execute the IBM once for each value identified in step 3. All other parameters are held at their standard value.

5) Calculate a sensitivity index for the parameter. This index is the slope of the model's output variable (from Step 1) with respect to the *scaled* parameter values determined in Step 3, determined using linear regression. Because parameter values are scaled, this sensitivity index can be compared across parameters.

However, it is also important to graph and visually inspect how the model output varied with the parameter's values to see if the relationship is nonlinear. For example, model output could peak at an intermediate parameter value, in which case the sensitivity index could be evaluated as the mean slope of the relation (a) below and (b) above the peak.

6) Repeat steps 3–5 for all parameters, and examine the sensitivity indexes for each to address the Phase 1 objectives. Of special importance is identifying any parameters with unexpectedly strong effects on model results. Such high-sensitivity parameters may indicate model equations or processes that are more important than anticipated; or they may indicate that the range of feasible values needs to be revised because it includes regions that produce absurd results.

2.2 Phase 2: Parameter interactions

The objective of Phase 2 is to investigate the frequency and strength of parameter interactions. "Parameter interactions" occur when a model's sensitivity to one parameter depends on the value of another parameter (Rose, 1989). To our knowledge, little if anything has been published on parameter interactions in IBMs, most likely because of the computational burden of conventional analysis approaches. Latin hypercube sampling (LHS; Rose 1989; Saltelli et al. 2000) makes this factorial approach more efficient, but even the analysis of parameter interaction results becomes a large project when the number of parameters is high: the number of potential pairwise interactions is $\frac{n(n-1)}{2}$ where n is the number of parameters, so even with only 20 parameters there are 190 potential interactions to analyze.

We developed an analysis approach that takes advantage of the individual-parameter sensitivity information generated in Phase 1 to limit the computational demand. First, the Phase 1 information is used to select only a small number of parameters with high sensitivity index values to investigate for interactions. Then each pairwise combination of these Phase 2 parameters is examined for interactions. In the absence of interactions, when two parameters are varied the model results fall approximately on a plane (for parameters to which the model responds approximately linearly). The slope S_E of this plane can be estimated from Phase 1 results: if I_a and I_b are the Phase 1 sensitivity indexes for parameters a and b, then $S_E = \sqrt{I_a + I_b}$. If there is an interaction, model results will no longer fall on a plane when two parameters are varied simultaneously. Hence, the presence of interactions between two parameters can be detected by any statistic that indicates the model response is non-planar with respect to the parameters. We used a somewhat arbitrary but simple and conservative (unlikely to detect interactions when they do not occur) measure: an interaction was assumed to occur if the model response slope (using linear regression on scaled parameter values from Phase 1), from simulations in which both parameters are perturbed simultaneously, differs from S_E by more than a specified amount.

The specific methods we used for each pair of Phase 2 parameters are:

1) Calculate S_E.

2) Select three values for each parameter: the standard value and the low and high ends of the range of feasible values from Phase 1.

3) Run the model for all nine combinations of values for the two parameters; and then replicate this factorial experiment at least two additional times (by using different random number seeds). (One of the nine combinations will actually be the standard value of all parameters so need not be re-executed for each parameter pair.)

4) Using linear regression, estimate the observed slope S_O of the model output's response plane with respect to the scaled parameter values: $S_O = \sqrt{S_a + S_b}$ where S_a and S_b are the regression coefficients for parameters a and b from the nine simulations. Calculate S_O separately for each replicate of the factorial experiment, and determine the mean and standard deviation in S_O among the replicates. [1]

5) Define an interaction among the parameters as occurring if S_E is outside the confidence interval defined by the mean $\pm$ two standard deviations of S_O.

[1] Paul–verify whether this is actually exactly what you did.

This approach is obviously not appropriate for parameters that the model responds to in a strongly nonlinear way. In such cases, alternatives could include using a linearizing transformation of results or simply examining how the model's response to the parameter with nonlinear effects differs among several discrete values of the other parameter.

2.3 Phase 3: Robustness of decision-support results

The objective of Phase 3 is to evaluate the effect of parameter uncertainty and sensitivity on the ultimate use of management IBMs: comparing alternative management actions. The motivation for Phase 3 is a problem discussed by Drechsler (1998): that conventional parameter sensitivity analyses do not tell us how parameter values affect such decision-support applications of models. Even if a model is highly sensitive to an uncertain parameter, it is not clear that this uncertainty affects the *relative* model results when management alternatives are simulated. To address this objective, we use a robustness analysis approach (see Ch. 9 of Grimm and Railsback 2005), asking how robust decision-support results from the IBM are to parameter uncertainty.

Our Phase 3 methods were modified from the approaches of Drechsler (2000), who addressed effects of parameter uncertainty on management alternatives in models that represent these alternatives via different sets of parameter values. We assume instead that, in complex IBMs, alternative management scenarios are represented as alternative sets of input data (e.g., initial population characteristics, spatial input describing habitat conditions, or time series input of managed variables such as river flow or harvest levels), while parameter values remain unchanged across scenarios. The general approach is to simultaneously vary a small number of important parameters using LHS, and examine how the IBM's ranking of several management scenarios is affected. Phase 3 uses the following steps.

1) Define the management scenarios and develop a set of input representing each. If this analysis is being conducted for an actual management application of the IBM, then real management alternatives can be used. Otherwise, hypothetical but realistic scenarios can be developed. The number of model runs required for the analysis increases linearly with the number of scenarios (s), so not many should be used; but hypothetical scenarios should reflect the range of inputs (and kinds of inputs that could vary) in real applications. This step also includes defining the IBM output(s) used to rank the management scenarios. The IBM should be run for several replicates of each scenario, using standard parameter values, to determine how much the selected output differs among the scenarios and how much stochastic noise there is.

2) Select the parameters to be analyzed. Because we use LHS, the number of model

runs required for Phase 3 does not necessarily increase directly with the number of parameters varied (Rose, 1989). However, including more Phase 3 parameters does increase the number of model runs needed to provide confidence that any strong effects that one parameter are not swamped and that important parameter combinations have not been missed.

Judgment is important in selecting the Phase 3 parameters. A primary consideration is the individual-parameter sensitivity results of Phase 1: the parameters with the highest sensitivity index values from Phase 1 deserve consideration for Phase 3, although such parameters may be excluded if their values are relatively certain (e.g., from laboratory studies). Parameters commonly used to calibrate the IBM should also be included in Phase 3. One way we kept the number of parameters low was to include only one parameter for a particular equation or process in the IBM, even if several of its parameters had high sensitivity values from Phase 1.

3) Definite a distribution (treated as a probability density function, PDF) for the value of each parameter. Triangular and rectangular distributions are useful because they provide distinct lower and upper bounds. We used triangular distributions with the peak at the parameter's standard value and the ends at the lower and upper bounds determined in Phase 1.

4) Divide each parameter's distribution into k intervals of equal probability, from which samples will be drawn during LHS. The value of k should be at least three, but there seems to be little reason for it to be much higher than perhaps four.

5) Conduct the LHS to determine which interval values are drawn from for each parameter, for a block of model runs (see, e.g., Sect. 2.2 of Rose, 1989). A "block" is k model runs, with values for each parameter chosen so each run's value is from a different interval. In our example below, we use $k = 3$, so each parameter's distribution is broken into 3 intervals (low, medium, and high; L, M, and H). For a block of 3 model runs, these 3 intervals are randomly shuffled for each parameter: the first parameter might have values from interval M in run 1, L in run 2, and H in run 3; the second parameter might have values from H, M, and then L; the third parameter from H, L, M, etc. (The same interval is never used twice for the same parameter in the same block of runs.)

6) Draw values of each parameter randomly from within its LHS interval, for each model run. To do so, we treated each interval of the parameter distributions (L, M, and H) as a separate PDF, so values with higher probability density are more likely to be drawn. For each parameter i and model run k, determine the parameter values and their associated likelihood (over the parameter's total distribution) $p_{i,k}$.

7) Execute the block of model runs. For each of the k parameter sets in the block, the IBM is run for each of the s management scenarios.

8) Determine the expected value E_s for each management scenario s, for the block

of model runs:

$$E_s = \frac{\Sigma_{j=1}^{k} P_j O_j}{\Sigma_{j=1}^{k} P_j}$$

where O is the output from each model run and P is the total likelihood of a model run, calculated by multiplying together the values of $p_{i,k}$ for each parameter. Determine the rank of each scenario: the scenario with rank 1 has the highest value of E_s, etc., up to rank s, which has the lowest E_s.

9) Repeat steps 5–8 for additional blocks, looking at the management scenario rankings for each block of model runs. Stop after it is sufficiently clear how much the rankings vary among blocks. One way to determine when enough blocks have been executed is to calculate, after each new block is executed, the value of E_s of each scenario over all the completed blocks. When scenario rankings from these cumulative values of $E_s$no longer change as more blocks are executed, the analysis can stop.

This approach weights the results for each set of parameter values by the likelihood of those values: results from runs with parameter values farther, on average, from the standard values are given less weight in the analysis. Some may feel that this approach underestimates effects of parameter uncertainty, or is simply too hard to explain. An alternative is, in step 8, to look at the unweighted rankings from each parameter set (O_1–O_k instead of E_s).

3 Example: Parameter Sensitivity of inSTREAM

We illustrate the sensitivity analysis strategy via an application to inSTREAM, an IBM designed to predict effects of river management (e.g., changes in daily flow, temperature, or turbidity) on trout populations (Railsback and Harvey, 2001, 2002; `www.humboldt.edu/~ecomodel/instream.htm`). In this IBM, site characteristics and management alternatives are represented via input data such as habitat cell characteristics and daily flow, temperature, and turbidity values. Only eight habitat parameters are used, mainly to determine daily food availability from hydraulic conditions in each cell. Many more parameters are used as coefficients in algorithms representing trout behaviors (e.g., feeding, habitat selection, spawning), physiological processes (e.g., growth, reproduction), and a variety of mortality risks. We analyzed a total of 90 parameters. These range in uncertainty from those with fairly well-known values determined from extensive data (e.g., lab experiments on feeding and bioenergetics; field measurements of fecundity), to those representing processes that are extremely difficult to observe (e.g., how predation risk varies with water depth or velocity).

For all of the analyses we focused on only one of the many outputs produced by inSTREAM: the biomass of adult trout (age 1 or older) as censused once per simulated year in mid-October, averaged over 11 simulated years. (Results from the first three of 14 simulated years were ignored as potentially influenced by initial conditions.)

3.1 Phase 1: Individual parameter sensitivity

To develop the Phase 1 individual-parameter sensitivity indexes, we used seven values for each parameter, with the first value at the low end of the range, the fourth value being the parameter's standard value, and the seventh value at the high end of the range. Parameter values were scaled to a range of 0–1, so the seven values of each parameter had scaled values of 0.0, 0.167, 0.333, 0.5, 0.667, 0.833, and 1.0.

Feasible ranges of parameters were defined by the authors of inSTREAM, who considered the parameter's meaning and the information used to estimate its standard value. In one instance, preliminary results led us to go back and reconsider the ranges selected for parameters. Two parameters control the length-weight relation in the simulated trout: as trout accumulate weight, their length is updated using the (inverted) empirical relationship:

$$fishWeight = fishWeightParamA \times fishLength^{fishWeightParamB}.$$

Initially we simply assumed *fishWeightParamA* and *fishWeightParamB* had feasible ranges of $\pm 5\%$ [2] ; however, results were absurd for parameter values at the extremes of this range (e.g., the model produced trout weighing a few grams but many meters long). A more careful review of the data (measured lengths and weights of real trout) showed that the feasible ranges of these parameters were much smaller.

The Phase 1 results produced no major surprises and no indication of extreme sensitivity to parameter values, but they were highly informative. The model exhibited low sensitivity to a large majority of parameters (Fig. 1): 60% of parameters have sensitivity index less than 500. A few parameters had high sensitivity index values: 11% of parameters had sensitivity above 2000, and two had values above 3000. The parameters we most expected to have strong effects on inSTREAM results did in fact have high sensitivity indexes: two parameters we use for calibration (controlling food availability and risk of predation by terrestrial animals) had sensitivity indexes of 1800 and 3300. However, two other parameters we use to calibrate juvenile trout size and abundance (representing a second food source and risk of predation by other fish) had relatively little effect on the adult trout predictions

[2] Paul needs to corroborate this.

(sensitivity values less than 1000). The parameter with highest sensitivity represents how the risk from terrestrial predators varies with water depth; this result was not anticipated but in retrospect makes sense: these predators are the dominant cause of mortality for simulated adults and depth (a) strongly affects the risk and (b) varies widely over space.

The prevalence of parameters with low sensitivity values does not mean that many of inSTREAM's parameters are unnecessary because they have little effect on results. Some of these parameters are necessary to represent one end of a function (e.g., the logistic curve for how predation risk varies with depth) that the model is sensitive to the other end of. Other parameters represent processes that are not important at the study site we used but likely would be important at other sites; for example, several parameters represent effects of extreme temperatures, which do not occur at the site used in this analysis.

Only one parameter produced a clearly non-linear and peaked response. This parameter is the time horizon over which trout make risk–growth tradeoffs in selecting their habitat cell; Railsback et al. 1999. Low values give most emphasis to avoiding predation and high values give most emphasis to avoiding starvation; adult trout biomass was highest at intermediate values.

3.2 Phase 2: Parameter interactions

The ten parameters with Phase 1 sensitivity values above 2000 were selected for Phase 2, so there were 45 pairwise interaction analyses. Using the methods described in Sect. 2.2 with three replicate runs for each parameter value combination, we found interactions in 42 of these 45 analyses. In some cases the interactions were quite strong: the mean value of S_O over three replicates was as much as 23 times greater than S_E; in 11 parameter pairs, S_O was over 5 times greater than S_E. All the parameters in these interactions control food intake or metabolic processes.

It is not clear how unique our finding of widespread parameter interactions is, as we found no similar analyses of complex IBMs. These results indicate that attempting to calibrate inSTREAM by varying one parameter at a time could be frustrating. (Instead, we execute factorial calibration experiments varying the 2-3 calibration parameters simultaneously).

3.3 Phase 3: Robustness of decision-support results

For Phase 3 we further reduced the number of analyzed parameters to seven. We used a triangular PDF to describe value ranges for each parameter; the PDF had its peak at the parameter's standard value and a range matching the range of values used in Phase 1. With $k = 3$ ranges for LHS, we broke each parameter's full range into 99 evenly spaced values, and calculated the likelihood for each such that the sum of likelihoods over the 99 values equals 1.0. The boundaries between the three parameter ranges (L, M, H) occur where the sum of likelihoods for values to the left equal 0.33 and 0.67. For parameters with their standard value in the center of their distribution, the low parameter range includes the first 40 of the 99 equally spaced values; the middle range includes the middle 19 values; and the high range includes the upper 40 values.

The decision-support results we analyzed are predicted trout biomass under four hypothetical stream management scenarios. These scenarios represent alternative management measures for a water diversion and timber harvest (both imaginary) on a mid-sized stream. The water diversion would reduce stream flow; flow affects the area of habitat and the amount of food for trout, and water depths and velocities. The timber harvest is assumed to increase turbidity (cloudiness of the water), which reduces feeding success. The scenarios (Table 1) differ in the minimum flow required to remain in the stream and the extent to which turbidity is increased. In simulations using standard parameter values, inSTREAM predicted scenarios 1-3 to produce trout biomass averaging 52, 76, and 72% of the baseline scenario 4. Scenarios 2 and 3 produce quite similar results; in fact the results in Fig. 2 for these two scenarios are not significantly different (one-way analysis of variance with Bonferroni comparison of means, p=0.05, n=10).

Even though absolute results from inSTREAM varied strongly among the different parameter sets, parameter variation had little effect on the relative rank of the four management scenarios. The likelihood-weighted average trout biomass values E_s produced exactly the same ranking of the scenarios as we increased the number of three-parameter-set blocks from one to 15 (Fig. 3), and the values of E_s stabilized after 5 blocks were executed. The baseline (scenario 4) produced highest trout biomass, followed in rank of descending biomass by scenarios 2, 3, and 1. In fact, all blocks, examined individually, produced the same likelihood-weighted rankings, even for the very similar scenarios 2 and 3. This consistency occurred even though the predicted trout biomass varied widely: some model runs produced complete extinction of the population and others produced biomass as much as 20 times that predicted with standard parameters.

Interestingly, we found the values of E_s to be much higher than the results obtained with standard parameter values (compare Figs. 2 and 3). This discrepancy occurs because, in the LHS analysis, parameter combinations that negatively affect simulated populations can never reduce trout biomass to less than zero but there is no limit on how much biomass can increase under parameter combinations with positive effects. Hence, simulated biomass could be only 100% lower than the biomass with standard parameter values but was as much as 2,000% higher.

The unweighted results are also quite consistent. When we simply averaged the simulated trout biomass for each scenario over the three model runs in each LHS block, we obtained the correct ranking in 13 of 15 blocks. The best and worst management scenarios were correctly identified in all 15 blocks.

4 Conclusions

Scientists develop and use complex models and IBMs because they are more like the real systems we study and, therefore, let us address more complex aspects of those systems. But one unfortunate consequence of being more like real systems is that complex models are harder to analyze and understand (Grimm and Railsback, 2005). Traditional uncertainty and sensitivity analysis methods cannot provide a complete picture of how complex IBMs respond to parameter variation because these models typically have many parameters, produce a variety of results, take a long time to execute, are stochastic, and are nonlinear in many ways. Yet their complexity makes parameter analysis especially important for these models.

The three-phase strategy we developed appears to be a useful compromise between what modelers need to know about parameter sensitivity of complex IBMs and what is computationally feasible. Phase 1 is especially important for identifying parameters most deserving attention in calibration and in research to reduce uncertainties. Phase 3 seems especially important for giving a model's clients an indication of how robust conclusions drawn from the model are. While Phase 2 results may be less urgent for model development or application, its analysis of parameter interactions seems important for developing a solid understanding of how an IBM behaves.

Even though our analysis strategy is a compromise, it still requires significant computational resources. In our example analysis we report results of 1660 runs of inSTREAM, which each take one half to several hours to execute on a desktop computer (the execution time varies widely as it depends on the number of trout "alive" during the run). However, the strategy is flexible and adaptable: in applications to other models, users can control the computational effort by altering the number of values for each parameter in Phase 1, the number of parameters included in phases 2 and 3, the number of replicate simulations used in Phase 2, and the value of k in Phase 3. On the other hand, we only conducted our analysis for one study site (the process could be completely repeated for additional sites) and focused only on one particularly important output of the IBM.

Our example analysis of inSTREAM confirmed some of our expectations about which parameters have strong effects. But the analysis also indicated that some parameters we use for calibration have only moderate effects on key outputs and identified one parameter—for how predation risk varies with depth, which is unfortunately difficult to measure—that has greater importance than we expected. At the same time, the analysis provided evidence that management support results from the model are quite robust to parameter uncertainty.

Our sensitivity analysis of inSTREAM found results generally more sensitive to parameters for food availability and metabolic processes than to behavior-related parameters, as did sensitivity analyses of at least two other large IBMs (Amano et al., 2006; Stillman et al., 2000). While representing behavior is undoubtedly critical for the accuracy of these IBMs, the consistent importance of food and metabolic parameters indicates that energetic processes are also very important and deserve careful attention in model development and testing. In fact, behavior in these three models is strongly determined by energetic processes, likely one reason why food and metabolic parameters are so important.

References

Amano, T., Ushiyama, K., Moriguchi, S., Fujita, G., Higuchi, H., 2006. Decision-making in group foragers with incomplete information: test of individual-based model in geese. Ecological Monographs 76, 601–616.

Drechsler, M., 1998. Sensitivity analysis of complex models. Biological Conservation 86, 401–412.

Drechsler, M., 2000. A model-based decision aid for species protection under uncertainty. Biological Conservation 94, 23–30.

Goss-Custard, J., Burton, N. H. K., Clark, N. A., Ferns, P. N., McGrorty, S., Reading, C. J., Rehfisch, M. M., Stillman, R. A., Townend, I., West, A. D., Worrall, D. H., 2006. Test of a behavior-based individual-based model: response of shorebird mortality to habitat loss. Ecological Applications 16 (6), 2215–2222.

Grand, T. C., Railsback, S. F., Hayes, J. W., LaGory, K., 2006. A physical habitat model for predicting the effects of flow fluctuations in nursery habitats of the endangered Colorado pikeminnow (*Ptychocheilus lucius*). River Research and Applications 22, 1125–1142.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., DeAngelis, D. L., 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198, 115–296.

Grimm, V., Railsback, S. F., 2005. Individual-based modeling and ecology. Princeton Series in Theoretical and Computational Biology. Princeton University Press, Princeton, New Jersey.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D. L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991.

Mooij, W. M., DeAngelis, D. L., 1999. Error propagation in spatially explicit population models: a reassessment. Conservation Biology 13, 930–933.

Railsback, S., Harvey, B., 2001. Individual-based model formulation for cutthroat trout, Little Jones Creek, California. Tech. Rep. General Technical Report PSW-GTR-182, Pacific Southwest Research Station, Forest Service, U. S. Department of Agriculture.

Railsback, S. F., Harvey, B. C., 2002. Analysis of habitat selection rules using an individual-based model. Ecology 83, 1817–1830.

Railsback, S. F., Lamberson, R. H., Harvey, B. C., Duffy, W. E., 1999. Movement rules for spatially explicit individual-based models of stream fish. Ecological Modelling 123 (2-3), 73–89.

Rose, K. A., 1989. Sensitivity analysis in ecological simulation models. In: Singh, M. (Ed.), Systems and Control Encyclopedia. Pergamon Press, New York, pp. 4230–4234.

Saltelli, A., Chan, K., Scott, E. M. (Eds.), 2000. Sensitivity analysis. Wiley series in probability and statistics. John Wiley and Sons, New York.

Shirley, M. D., Rushton, S. P., Smith, G. C., South, A. B., Lurz, P. W., 2003. Investigating the spatial dynamics of bovine tuberculosis in badger populations: evaluating an individual-based simulation model. Ecological Modelling 167, 139–157.

Stillman, R. A., Goss-Custard, J. D., West, A. D., Durell, S. E. A. V. l. d., Caldow, R. W. G., McGrorty, S., Clarke, R. T., 2000. Predicting mortality in novel environments: tests and sensitivity of a behaviour-based model. Journal of Applied Ecology 37, 564–588.

Table 1
Hypothetical management scenarios used in the Phase 3 analysis of inSTREAM.

Scenario	Minimum flow (cubic meters per second)	Turbidity (increase from baseline)
1 (no mitigation)	0.3	60%
2 (mitigated flow)	0.5	60%
3 (mitigated turbidity)	0.3	20%
4 (baseline)	—	—

Figure Captions

Figure 1: Phase 1 parameter sensitivity index distribution for inSTREAM. There is one dot for each parameter analyzed; its X value is the parameter's sensitivity index value and its Y value is the percent of parameters with sensitivity values less than or equal to the parameter's.

Figure 2: Results of 10 replicate simulations for the four management scenarios considered in the Phase 3 analysis of inSTREAM, using standard parameter values.

Figure 3: Expected trout biomass E_s under the four alternative management scenarios of the Phase 3 analysis of inSTREAM, calculated over one to 15 LHS blocks. The X axis is the number of blocks executed; the Y axis is the value of E_s calculated over those blocks. After five blocks were executed, additional blocks had little effect on expected biomass.

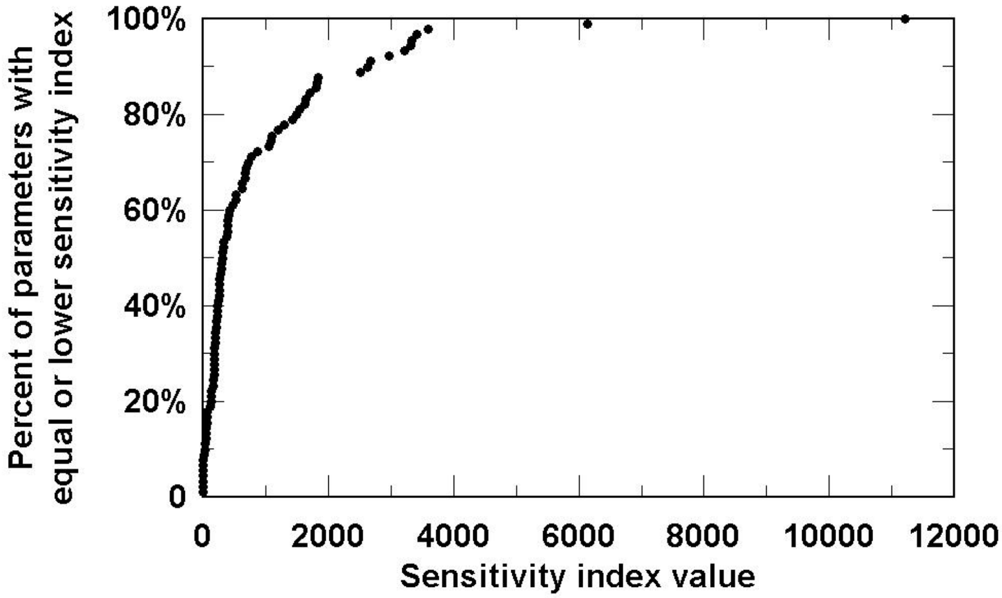
Percent of parameters with equal or lower sensitivity index
100%
80%
60%
40%
20%
0
0
2000
4000
6000
8000
10000
12000
Sensitivity index value

Figure 1.

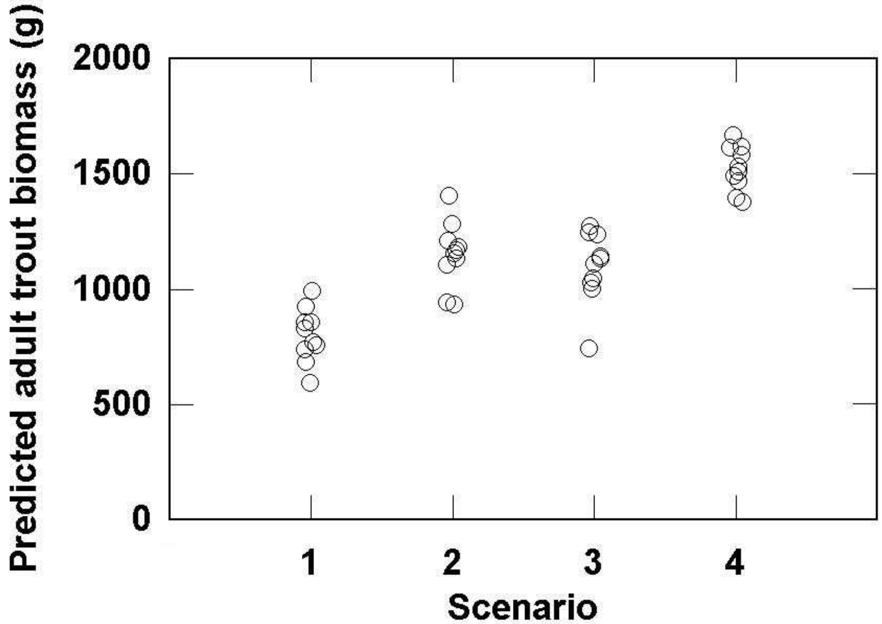
Predicted adult trout biomass (g)
2000
1500
1000
500
0
1
2
3
4
Scenario

Figure 2.

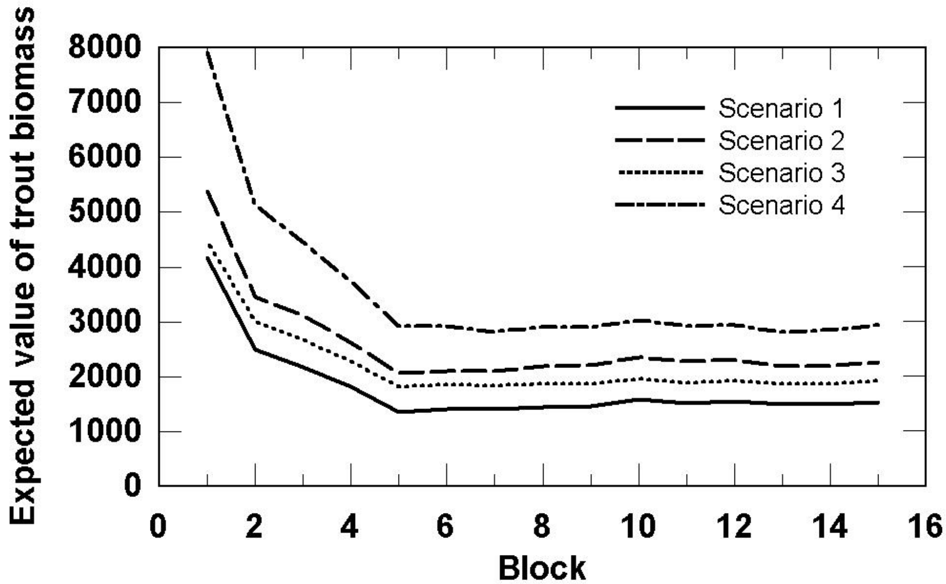
Expected value of trout biomass
8000
7000
6000
5000
4000
3000
2000
1000
0
0
2
4
6
8
10
12
14
16
Block
Scenario 1
Scenario 2
Scenario 3
Scenario 4

Figure 3.