Ecological Applications, 5(2), 1995, pp. 411–420

ACCEPTANCE CRITERIA FOR USING INDIVIDUAL-BASED MODELS TO MAKE MANAGEMENT DECISIONS[1]

JONATHAN BART
Ohio Cooperative Research Unit, Department of Zoology, Ohio State University, Columbus, Ohio 43210 USA

Abstract. Individual-based population models are beginning to have substantial influence in conservation biology. One of the dangers of such models is that their reliability may be poorly understood, and generally over-estimated. To help avoid this problem a set of guidelines is suggested here for evaluating individual-based models. Major components of the evaluation include a description of the model and estimates of the reliability of its predictions. The description should include detailed explanations of the model's purpose(s), its structure, and the assumptions it makes. Analyses of reliability at four levels may be useful: structural assumptions, parameter values, secondary predictions of the model, and primary predictions of the model. The guidelines also recommend that "best-" and "worst-case" scenarios, intended to span the range of plausible outcomes, be prepared using the model, rather than presenting single predictions or conclusions. I argue that results from such an evaluation should be prepared and peer reviewed before the model is used to make or defend management decisions.

Key words: individual-based model; management decisions; model evaluation; modelling guidelines; population model; simulation; Spotted Owl; validation.

INTRODUCTION

The use of individual-based models to make or defend management decisions is becoming more common in conservation biology (DeAngelis and Gross 1992, Conroy et al. 1995, Dunning et al. 1995, Holt et al. 1995, Turner et al. 1995). Such models had a significant influence in the Northern Spotted Owl controversy (Thomas et al. 1990, Bart et al. 1993, Turner et al. 1995); they were used in the California Spotted Owl analysis (Verner et al. 1992); and they are currently being developed or considered for the desert tortoise, the Kirtland's warbler, and other species. These models are valuable because they analyze the effects of management proposals at a level of objectivity, detail, and realism that is often unachievable using either analytical models or professional judgement. As with any model, however, a danger exists that more credence may be given to the output of the model than is warranted (Thomas 1986). It is thus important that models be evaluated before they are used to help solve management problems.

This article describes a set of guidelines for evaluating individual-based models. The recommendations grew out of experiences on the Northern Spotted Owl Recovery Team. Output from an individual-based model was available to the Team, but Team members had difficulty deciding how much confidence should be placed in the results. The Team therefore asked that a set of guidelines be developed and that an evaluation conforming to these guidelines be carried out before results obtained using the model be accepted by the Team. This report is the result of that request. The guidelines are presented here both to help guide future work on Northern Spotted Owls, and to help other groups that must decide how much confidence to place in the results of individual-based models.

Many authors have stressed that ecological models should be evaluated before they are used to make or defend management decisions. For example, Gentil and Blake (1981:23), discussing ecosystem models, stated that "validation is absolutely necessary if the model is to be used as a practical tool." Russell (1975:9) listed six steps in model development and testing identified by Orlob (1975) and noted that "All these steps, of course, he sees as prior to application, the presumed goal." In some wildlife studies, models have not been tested as thoroughly as desirable. For example Berry (1986:4) stated that "validation, invalidation, and verification of models are critical processes and must receive more attention." Similar advice is contained in Overton (1977), Farmer et al. (1982), Marcot et al. (1983), Dedon et al. (1986), Raphael and Marcot (1986), and Laymon and Barrett (1986:87) who stated "We strongly discourage the use of untested models because they lack credibility."

Additional evidence of the need to evaluate models before using them is provided by authors who have tested predictive models and found that their performance was not as good as expected (Conroy et al. 1995). For example, Rotenberry (1986:217) developed models for bird abundance in semiarid shrubsteppe environments. The models performed well on the study plots but "failed to predict adequately the densities of sparrows on five of the original plots on which sampling continued for four more years." Similarly, when

[1] Manuscript received 3 January 1994; accepted 12 May 1994; final version received 1 September 1994.

Cole and Smith (1983:374) studied Habitat Suitability Index (HSI) models, they found that several were "almost useless." Laymon and Barrett (1986:87) obtained "poor results for the three (HSI) models even though they were based on what was believed to be good information." Other similar outcomes and recommendations that wildlife models be tested before being used are contained in Verner et al. (1986).

This brief summary of the pitfalls of not evaluating models seems sufficient to justify the following general principle:

> *Models should not be used to make or defend management decisions until they have been thoroughly evaluated and the results of the evaluation have been subjected to peer review. The peer-reviewed model evaluation should be clearly presented and included with the model when it is given to the managers.*

This principle should not be interpreted as meaning that models should play no role in decision-making until their reliability has been completely established (because then they would never be used); only that a substantial effort at evaluation should be made.

A number of authors have suggested general principles and approaches for model evaluation. Caswell (1976) and Conroy et al. (1995) distinguished between gaining insight into system behavior and predicting a specific outcome. Caswell noted that the objective of gaining insight into system behavior is an example of scientific inquiry in general and that validation (i.e., "proof") is not possible, whereas quantitative acceptance criteria (e.g., "95% of the predictions should be within 10% of the true value") can be developed when the objective is predicting a specific outcome. Swartzman and Kaluzny (1987) added that other objectives, such as identifying new research priorities may also be identified as reasonable goals of modelling. They distinguished three underlying objectives for model evaluations: ascertaining (1) how accurate the assumptions made in building and running the model are, (2) how realistic the behavior resulting from those assumptions is, and (3) how sensitive model behavior is to changes in these assumptions.

Other authors have provided guidelines for evaluating models in agriculture (DeWitt and Goudrian 1974), economics (Naylor and Finger 1971), ecosystems analysis (Garrett 1975, Mankin et al. 1975, Hall and Day 1977, Costanza 1985, Straskraba and Gnauck 1985), hydrology (James and Burges 1981, Reckhow and Chapra 1983), and wildlife (Grant 1986, Schamberger and O'Neil 1986, Starfield and Bleloch 1986).

The approach suggested below is derived from these papers and from the experiences of the Northern Spotted Owl Recovery Team. Team members needed—but did not have—a detailed, accepted set of guidelines describing the analyses that should be completed before model outputs were used to make management decisions. The descriptions and analyses below are intended as a first step towards establishing such a protocol.

Components of a Model Evaluation

Model evaluation is used here to mean "the determination of the usefulness and accuracy of model predictions . . ." (Marcot et al. 1983:317). The components of a comprehensive model evaluation are divided into four categories: objectives of the model, description of the model, analyses of model reliability, and synthesis. Each of these is discussed below.

Objectives of the model

Models cannot be evaluated rigorously without a detailed description of the objectives of the model and the degree of reliability the model is believed to provide for each prediction (Overton 1977, Reckhow and Chapra 1983, Swartzman and Kaluzny 1987, Conroy et al. 1995). Failure to provide detailed descriptions of model objectives has been a frequent problem. Overton (1977: 51), for example, comments that "model objectives are seldom stressed sufficiently. It is often difficult to identify objectives from model documentation, and the great majority of criticisms of models relate to a capacity for which the model was not designed in the first place." Each of the predictions made by the model should therefore be described, and the degree of reliability needed, or claimed to be provided by the model, should be specified.

Description of the model

A detailed description is needed of the model's general structure and organization and the sequence of steps it carries out to make its predictions. Consideration should be given to providing an appendix with a detailed description of the computer program. The main description could include the basis for classifying (i.e., subdividing) the environment; number of age and sex classes; behavior of the organism during reproduction, dispersal, and subsequent stages of the life cycle; and so on. Identifying assumptions implied by the model structure is also valuable. For example, if two age classes are defined, then individuals within each age class are implicitly assumed to have similar (average) demographic rates and types of behavior. If home ranges are delineated and ranked by the proportion of them covered by suitable habitat, then there is an implicit assumption that the average quality of habitats classified as "suitable" will not change through time.

Analyses of model reliability

This section follows the organization proposed by Schamberger and O'Neil (1986) who pointed out that analyses of model reliability can be carried out for any of four "levels": structure, parameter values, secondary predictions, and primary predictions. The analyses may focus on the predictions of greatest interest, but should include some information on the reliability of each of the predictions because any of them may be of interest (presumably they would not be provided by the model if they were of no interest). At the least, if the reliability of some predictions has not been investigated, this should be stated clearly.

Structure of the model.—The realism of the assumptions about behavior, habitat relationships, demography, and so on should be assessed by means of a detailed literature review and description of any original analyses that have been carried out. The basis for assumed long-term environmental trends should also be documented carefully. One does not expect a model to be fully realistic of course (Krebs 1980), but the degree of realism should be described, and estimates should be provided of how the model predictions are affected by departures from realism. Such discussions will be brief in many cases. For example, the model may assume that births and dispersals occur at the same time in all birds even though this is clearly not true. This simplification may seem unlikely to have a major impact on predictions by the model. In other cases, the effect of unrealistic, or possibly unrealistic, assumptions may be much harder to estimate. For example, the effect of assuming no senescence in productivity or adult survivorship or that females do (or do not) leave their territories following death of their mate may be much harder to assess, and performing a few model runs varying these traits or behavior may be worthwhile. Even if such analyses are not carried out, these issues should at least be identified and a candid assessment of how the lack of realism might affect the model's accuracy should be provided.

Parameter values.—Studies that provide the "best guess" about parameter values, and reasonable ranges for them, should be summarized or at least identified by reference. Where new analyses are carried out to estimate parameter values, they should be described in detail, perhaps in appendices. Standard statistical methods should be used to describe reasonable ranges for each parameter whenever possible. Even if formal statistical methods cannot be used, some indication of the range of possible values should be given.

The effect on model predictions of changing each parameter within its "confidence interval" should then be analyzed. A general qualitative discussion will suffice to make many important points. For example, predictions about the persistence of long-lived species are strongly affected by adult survival rates and apparently by settling rates (K. S. McKelvey, *personal communication*), whereas they are much less affected by small changes in productivity or juvenile survivorship.

Sensitivity analyses should be carried out to assess the effect of uncertainty about variables that have a substantial effect on model outputs, or that are poorly known. The basic approach involves varying each important parameter across the range of plausible values that it might take and recording the resulting change (from a baseline) in model predictions. Unfortunately, the number of combinations that might usefully be evaluated is usually very large, and a rationale for identifying a tractable number of combinations must be developed. Overton (1977), Swartzman and Kaluzny (1987) and Reckhow and Chapra (1983) discussed this problem and provided references that may be of some assistance. Much judgement, however, must be left to the modeler.

The phrase "sensitivity analysis" has been used by mathematical modelers to mean an examination of the partial derivatives (i.e., rates of change) of the structural equations, evaluated with respect to one of the parameters (e.g., Tomovic 1963, Meyer 1971, Tomovic and Vukobratovic 1972, Miller et al. 1973). These analyses provide explicit descriptions of how much the model output changes in response to small changes in the input values. Current population models, however, are generally simulation, rather than mathematical, models, and this approach may not be appropriate (Gardner et al. 1981, Reckhow and Chapra 1983). Even if it is, the models are usually far too complex (and often are not even explicit) for the purely mathematical techniques to be practical. As a result, sensitivity analyses of ecological models are now usually carried out by simulation rather than by mathematical analysis (Swartzman and Kaluzny 1987, Dunning et al. 1995). This approach involves varying one factor while holding all others constant to determine how sensitive model outputs are to the factor being studied. It is generally not feasible to vary all factors in this manner, but the ones thought to be most important can be studied. Dunning et al. (1995) discuss advantages of this approach.

When data from the species in question are unavailable or unreliable, then data from other species may be of considerable value in estimating the reliability of model assumptions. Jared Verner (in Thomas et al. 1990) used this approach in analyzing how many pairs of Spotted Owls each habitat patch should contain to reduce the probability of local extinction to an acceptable level.

Secondary predictions of the model.—The phrase "secondary predictions" is used here to mean intermediate outputs of the model that are not provided as standard output or would usually not be used in making management decisions but that can be used to assess the reliability of the model. In an avian population dynamics model, examples might include the pathways followed by dispersing juveniles, distribution of ages at first breeding, locations of individuals across a landscape, and distribution of the amount of suitable habitat within territories. Each of these outputs is a function of two or more input variables but might not be the model output of primary interest. They deserve evaluation, however, because comparing them to empirical data, or to data for other similar species, helps identify how well the model simulates actual population dynamics (Caswell 1976, Overton 1977). Sensitivity analyses may also be helpful at this level. For example, the pathways followed by dispersing birds, as depicted by the model, might be consistently different from the paths taken by actual birds, but this might not affect accuracy of the output by the model if final settling places were similar to those observed empirically. Overton (1977) notes that this phase of the analysis is greatly facilitated by structuring the model so that its individual components correspond to natural processes and events for which empirical data are available.

Primary predictions of the model.—Primary predictions of the model are the outputs of greatest interest; they provide, or summarize, the information that will be used in making management decisions. Assessments of these predictions basically involve "reality checks." For example, predicted population trends and distributions can be compared to observed trends and distributions. Perhaps the best evaluation at this level involves starting the model at some time in the past when conditions can be inferred, letting the environment change in ways that appear realistic, and then determining whether the "predicted" distribution of individuals roughly matches the actual distribution. This is a strong test of the model: it should be carried out with any population model when the required data are available, and the results should play a major role in the overall assessment of model reliability.

As noted by numerous authors (e.g., Mankin et al. 1975, Caswell 1976, Overton 1977, Marcot et al. 1983), many of the results from analyses of model reliability will have been used to further improve the model and thus do not constitute fully independent evaluations of the model. The extent to which the analyses are independent of model development should thus be made clear.

Synthesis

Results from the evaluation should be integrated and presented in a form that provides a realistic description of the reliability of the predictions of the model. One way of doing this is to develop two sets of input parameters: one that represents a "worst" (or minimum) case and one that represents a "best" (or maximum) case. These alternatives can then be used to establish the range of outcomes that appears reasonable based on the model evaluation. Formal statistical methods for establishing this range will seldom be available because the relationships between the variables and the process used to generate the predictions are too complex. If distributions can be specified for the variables, and if the structure of the model is not in significant doubt, then simulations can be used to generate the best- and worst-case scenarios. Formal procedures for this approach are described by Miller (1979) and Swartzman and Kaluzny (1987:220–234). Alternatively, a more qualitative rationale may be developed. Regardless of what methods are used to produce the best- and worst-case scenarios, their development should be given high priority, because without them there may be little way for a reader to interpret the results of the model evaluation.

Explicit consideration should also be given to the question "Does the model improve our ability to make decisions?" For example, one possible purpose of a population model is to rank alternative management plans, such as those considered under an Environmental Impact Statement. The alternatives, however, usually vary so greatly in the degree of protection they give to wildlife, that no formal modelling is needed to rank them. Indeed, if the model showed a different ranking than the one that seemed obvious, the users would probably conclude that the model was incorrect. In such a case, carrying out a modelling exercise is unnecessary and gives an undue appearance of sophistication in the analysis of options.

Consideration should be given to how over-reliance on the model can be avoided. As noted above, experience shows that people often overlook statements that the results of complex computer models are of poor reliability (Thomas 1986). For example, the report (Thomas et al. 1990) of the Interagency Scientific Committee (ISC) stated clearly that the modelling effort was of secondary importance in developing the conservation strategy. The results of the modelling were used primarily to defend the reasonableness of the claim that the ISC conservation strategy would provide adequate assurance of protecting owls. Nonetheless, subsequent court cases and other controversies over the ISC proposals have concentrated on the modelling work as the primary basis for the ISC's strategy (B. R. Noon, *personal communication*).

An Example: A Model for the Northern Spotted Owl

This section presents an evaluation, following the guidelines above, of a model, MOSAIC, for the Northern Spotted Owl. MOSAIC, being developed in my laboratory, is spatially explicit and individual-based and was inspired by the model OWL developed by McKelvey, Noon, and Lamberson (McKelvey et al. 1992). MOSAIC is designed to project owl numbers and distribution across large landscapes. The evaluation presented below is much abbreviated, but enough information is provided to describe the content of a complete evaluation.

Objectives of the model

The general objective being evaluated in this section is carrying out certain analyses required by the Recovery Plan for the Northern Spotted Owl (Bart et al. 1993) before the owl could be considered for delisting (i.e., removing the owl from the list of "Threatened Species" established under the Endangered Species Act of 1973 [Public Law 93-205, and as amended]). The model might be used to evaluate many different recovery proposals, but I concentrate below on the proposal contained in the Recovery Plan.

Four conditions were required before delisting should be considered. The first three emphasized existence of reliable monitoring data, stable or increasing population trends, and effective regulatory protection. The fourth criterion was that "The population is unlikely to need protection under the Endangered Species Act during the foreseeable future." The Plan included the following explanation of the analyses required to demonstrate that this criterion had been met:

Populations that are temporarily stable but likely to decline again in the foreseeable future cannot be considered recovered and should not be delisted. Detailed analyses of the likelihood that the population will remain stable or increase must be carried out before delisting. The analyses should include observed and anticipated effects of a) fluctuations in abundance, fecundity, and survivorship; b) movements by birds within the area and to or from surrounding areas; c) changes in habitat including ones due to catastrophic events; d) loss of genetic diversity; and e) any other threats to the population whose effects might be significant. These analyses are particularly important for small populations.

MOSAIC includes the effects listed above. It projects owl numbers in space and time and thus provides the information needed to estimate the likelihood that populations will "remain stable or increase." This phrase implies that owl abundance exceeds some minimum level and that population trend (annual rate of change in population size) exceeds some threshold (such as 1.0). The goal of the modelling exercise is to describe the likelihood that these population performance criteria would be met if a proposed recovery program was carried out. The Recovery Plan stresses that analyses of this sort would be needed for any area considered "significant" under the Endangered Species Act. The entire area being considered for delisting (the whole range or some large portion of it) would thus be divided into "domains," and a separate analysis would be carried out for each domain. The modelling exercise would be summarized by statements like "In all eight domains, populations persisted and were stable or increasing with probabilities exceeding 90%," or "In seven of eight domains, populations persisted and were stable or increasing with probabilities exceeding 95%, but in one domain this probability was only 30%."

The Endangered Species Act directs agencies to give the "benefit of the doubt to the listed species" in carrying out the Act. Thus, where doubt about model assumptions or parameter values exists, the model should be constructed to present a plausible but worst-case scenario.

The specific primary objective for the modelling exercise might be determined in the following stages: (1) the user of the model results defines a temporal period and the spatial domains of interest for the analyses; (2) in each domain, the user defines a minimum average abundance required for "persistence" and a minimum population trend required for "stable or increasing population size"; (3) a plausible but worst-case set of assumptions about model structure and parameter values is chosen; and (4) the model is run numerous times and the result "success or failure" is recorded for each domain on each run. The results of step 4 provide the model output of primary interest.

The Recovery Plan proposed establishing a network of "Designated Conservation Areas" to be distributed throughout the owl's range and managed for Spotted Owls. Harvest of old-growth forests was envisaged outside these areas, and owl populations were expected to decline as a result. Owls were predicted to persist throughout the current range of the owl, however, and to stabilize sometime after the harvest of old-growth forests ceased in about 50 yr. The word "stable" in the phrase "remain stable or increase" thus means remain stable after the period of general population decline. Owls are long-lived so stability might not occur for a few decades after harvest of old-growth ceased, and long-term fluctuations in population size may occur. These facts suggest that the period 70–170 yr from the present would be an appropriate period for analyzing population size and trend.

Description of the model

As noted above, MOSAIC is spatially explicit and individual-based. Habitat is classified using a two-category system (suitable vs. unsuitable). Suitable habitat is defined in several different ways:

1) Areas classified as "suitable" by agency biologists;

2) areas that have not previously been harvested and that occur at elevations lower than a user-specified value;

3) regenerating areas that were suitable prior to harvest and are older than a user-specified age.

The smallest spatial unit defined by the model is a 1200 × 1200 m "cell." Simulations begin as early as 1950 (for certain evaluations of the model) and extend to 2160 (i.e., 170 yr from 1990). The model requires the amount of suitable habitat in each cell during each year of the simulation. These amounts were determined using GIS methods and extrapolation backwards in time using harvest records and forwards in time using the management proposals in the Recovery Plan. The "amount" of suitable habitat in a given cell and time is approximately the proportion of the cell covered by suitable habitat. More specifically, each cell is divided into nine smaller cells and each small cell is coded "0" if less than half of it is suitable habitat and "1" otherwise. The habitat score for the cell is the sum of these nine numbers and thus ranges from 0 to 9. Cells of different size are relatively easy to produce using these GIS methods. A complete description of the model would provide additional details about the methods used to project amounts of suitable habitat backwards and forwards; these methods were complex, but some of the details are important for interpretation of the model results.

Each 1200 × 1200 m cell is given a score that depends on the amount of suitable habitat, the number of owls using the cell, and the amount of suitable habitat in surrounding cells. The rules for defining these variables and combining them into a single cell score are flexible. These scores are updated constantly as birds settle and die and habitat change occurs. Territories and home ranges are also given scores that depend on the benefits and the costs of the territory. The "benefit" is the sum of the cells' scores. The "cost" is an increasing exponential function of size, so that eventually the cost of adding a cell is greater than the benefit, even if the cell is of high quality, so the net value of the territory declines. The user controls the parameter in the exponential function, and this provides a convenient method for determining territory size and owl density.

Birth and death rates are entered by the user. They may be sex-specific, habitat-specific, and age-specific up to age 30 yr. Birth is defined as the production of a young owl that survives until soon after nest departure. Users have the choice of entering rates individually or entering the parameters in the equation

$$r = \alpha - \beta(\tau - \text{age})^2 \quad (1)$$

where α, β, and τ are parameters, and age is in years at the start of the breeding season.

The program first distributes owls across the landscape, allowing them to delineate territories, mate, or remain floaters using the routine SEARCH described below.

The year is divided into four seasons: breeding, fall dispersal, wintering, and spring dispersal. Annual survival is partitioned into four season-specific rates whose product equals the user-specified annual rate. Various options are available for calculating the seasonal rates. At the start of the breeding season, the program determines whether each bird lives or dies using the age- and sex-specific survival rates. If either member of a pair dies, that pair does not produce young. If both members survive, then the number and sex of offspring is determined using the birth rates applicable to the pair. Territorial birds have a small (user-specified) probability of dispersing that depends on whether their mate survives, on the bird's sex, and on the quality of its territory. Adults that will disperse, and all young, are placed in a special array (MOVERS$).

At the end of the breeding season, floaters are given an opportunity to fill vacancies in the territorial population. All floaters within a user-specified distance of the vacancy are identified and one is chosen. Older birds, or birds closer to the vacancy, may be given a competitive advantage in filling these territories.

During the fall dispersal period, the program randomly selects birds from MOVERS$ and calls a DISPERSE routine. DISPERSE is an iterative routine in which a bearing and distance are chosen from user-specified distributions and used to determine the location in which the dispersing bird settles. Each time a new bearing is chosen, the deviation from the past bearing is categorized as "left," "straight ahead," and "right" with a user-specified definition of straight-ahead (e.g., $\pm 5°$). A tendency to alternate left and right bearings can be included; it tends to keep the birds moving in a straight line. The program also assesses habitat ahead of the dispersing bird and includes an option for giving the bird more probability of settling or turning sharply if the habitat ahead is substantially lower in quality than the habitat the bird is currently in.

Each time the bird lands, another routine, SEARCH, is called in which the bird defines the best home range it can find, decides whether that home range is suitable, and then either settles or jumps to a new location. SEARCH is a complex routine and will be described only briefly here. A tentative home range is defined following the minimum convex polygon approach (e.g., White and Garrott 1990). The program then determines whether the net value of the territory would be increased by (a) subtracting the best corner cell, (b) adding the best cell adjacent to the tentative territory, or (c) doing both (a) and (b) above. If the answer to any question is "Yes," then the change is made and the program begins these three questions again. This part of the routine ends when the answers to all three questions are "No" or when a user-specified number of alterations has been made. The program then passes the territory score back to DISPERSE, which determines whether the bird settles or not. The entire process is first carried out using rules designed to simulate a bird searching for a territory; if the bird chooses not to settle then the process is repeated using rules designed to simulate a bird searching for a home range (i.e., attempting to settle as a floater). The main difference between these analyses is that birds searching for a territory are influenced strongly by the presence of an unmated territorial bird of the opposite sex and by the number of other territorial birds, whereas these factors are less important to birds attempting to settle as floaters.

During the winter period, birds are again selected randomly and their death or survival is determined using the applicable rates. At the end of the season, vacancies in the territorial population are filled using the same routine as used at the end of the breeding season. Following this process, various summary statistics describing the population (e.g., age and sex distribution, number of territorial birds and floaters, age distribution of first-time breeders, amount of suitable habitat, distribution of territory scores) are recorded for each domain. Habitats are then updated (optionally) and the next year of the simulation begins.

At the end of the simulation, the program can display a large number of graphs and other descriptive statistics or it can repeat the entire process and store the summary data for later examination.

Three assumptions, or sets of assumptions, in the model have a particularly large effect on the projections it makes. First, the population trend is particularly sensitive to the adult survival rate. Population change from one year to the next can be written as "starting population size plus births minus deaths":

$$N_1 + N_1bs_0 - N_1(1 - s) = N_2,$$

where N_1 and N_2 are the population sizes in years 1 and 2, and b, s_0, and s are the per capita birth, first-year survival, and subsequent survival rates. Thus,

$$bs_0 + s = N_2/N_1.$$

An absolute change in adult survival rates causes an equal change in population trend (i.e., if s is replaced by $s + \alpha$, then N_2/N_1 is replaced by $[N_2/N_1] + \alpha$). This is not true for the birth and first-year survival rates because they multiply each other (i.e., if b is replaced by $b + \alpha$, then N_2/N_1 is replaced by $[N_2/N_1] + s_0\alpha$). Since b and s_0 are both in the range 0.2 to 0.4, this considerably dilutes the effect on population trend of changing either one of them. Thus, obtaining accurate estimates of adult survival rates and of how these rates vary in different habitats is of particular importance in using the model.

A second critical assumption, or set of assumptions, concerns the ability of dispersing juveniles or floaters to find vacant but suitable territories. The Recovery Plan provides for the maintenance of suitable habitat, of sufficient size to support several thousand owls, distributed throughout the range. No one is proposing plans that would maintain only a few hundred owls. Thus, if owls always find suitable habitat, then little threat exists due to genetic factors or demographic stochasticity. On the other hand, the Designated Conservation Areas (DCAs) are separated by substantial distances in which the habitat is potentially hostile to owls, and the number of owls in these DCAs, particularly at present since many of them have been heavily harvested, is small enough that these local populations could easily disappear due to chance events. If immigration rates are too low, this process would eventually lead to a severe reduction in the owl's range or even to extinction. The extent of movements by dispersing birds and floaters thus has a major impact on the model's predictions.

A third important issue, pointed out by K. S. McKelvey and B. R. Noon in their owl-modelling work, is the ability of owls to distinguish marginal from superior habitat. In their simulations, population trend varied from negative to positive according to whether owls could distinguish territories on which fitness was 0.98 from territories on which fitness was 1.02. This issue needs additional investigation, but McKelvey and Noon's work suggests that the assumptions made about such discrimination abilities may have significant impacts on model predictions.

Other issues that warrant close examination include the system used to classify habitat quality, birth rates and first-year survival rates and how they are affected by habitat quality, and tendencies of territorial birds to abandon their territories, especially after the death of a mate.

Analyses of model reliability

Structure of the model.—The model is sufficiently flexible that most factors of possible importance to owls can be accommodated by proper choice of parameter values (see below). In a few respects, however, the model's structure does force the user to make certain assumptions. Perhaps the most important is that the model classifies habitat using only two categories: suitable and unsuitable. No distinction is made between nesting and foraging habitat or between summer and winter habitat. The basis for this decision was that the model was intended for application in areas where the habitat was largely either old growth, and suitable for all activities, or harvested since 1960 and not suitable for any activities. This approach seems reasonable for the Olympic Peninsula, the Cascade Mountains west of the Cascade Crest, and most of the federal land in the Klamath Province. In other areas, selective harvest has been extensive or substantial areas are covered by medium-age stands, and the two-category system may be less appropriate (Thomas et al. 1990).

Parameter values.—One important class of assumptions includes those related to whether habitat is suitable or unsuitable. Most of the data on present suitability of habitat comes from the agencies and has been reviewed numerous times (Thomas et al. 1990, Bart et al. 1993). The biggest concern is probably over whether the biologists used a "certainly suitable" or a "might possibly be suitable" definition. This issue was important on the Oregon Coast range where the latter definition was apparently used. As noted above, however, the model does not apply particularly well to this area anyway because large areas are covered by 60–80 yr old stands and it is unclear how well a two-category habitat classification system applies to such environments. In most other areas, the current environment is either clearly suitable or not suitable at all. One possible exception to this generalization is that some decision was needed on the upper elevational limit for suitable habitat. This issue is more important in Washington because more high-elevation areas occur there, but even there the total area involved is relatively small. Related to the question of whether suitable habitat has been correctly identified is the issue of when habitat becomes suitable in the future. Little good information exists on this issue, though Bart and Forsman (1992) showed that extensive stands of 50–80 yr old forest were essentially uninhabited by owls. On the

TABLE 1. Summary statistics describing dispersal behavior of Spotted Owls.

Period	Dispersal distance (km)		Dispersal period duration (d)		Av. distance dispersed per day		Index of dispersal path width*
	Mean	SD	Mean	SD	Mean	SD	
Fall	21	20	13	11	4.6	5.3	5.6
Spring	39	12	43	34	2.1	2.3	20.6

* SD(l_{ij}) Where l_{ij} is the perpendicular distance from location j of bird i to the straight line connecting birds i's locations at the start and end of dispersal.

other hand, telemetry data shows that foraging owls begin to make extensive use of stands as young as 60 yr old (Thomas et al. 1990). This issue would have important ramifications for population performance during a transition period when young stands are growing up in Designated Conservation Areas, and for the predicted suitability of areas with small amounts of old growth (which could be used for nesting and roosting) surrounded by extensive 50–80 yr old stands. The model would classify such stands as unsuitable when in fact they might support viable owl populations. The model makes the conservative assumption that such areas would be unsuitable.

The parameters affecting dispersal behavior were estimated using data from Miller (1989) who followed 31 radio-transmittered birds from their natal site for up to 17 mo. I characterized their movements during fall and spring dispersal by determining the mean and standard deviation of total straight-line distance moved, duration of the dispersal period, average distance moved per day (calculated from instances when the birds were relocated on consecutive days), and the "index to dispersal path width" (Table 1), a measure of how much dispersing birds deviated from their general dispersal bearing. This analysis provides the summary statistics that the owl model should conform to. Fall dispersal involved shorter distances and durations, longer distances moved per day, and a narrower path width than spring dispersal, so separate sets of parameters should be used for fall and spring dispersal.

Change in habitat is known to have some influence on dispersing juveniles, at least when the change is drastic. Thus, dispersing, radio-tagged owls on the Olympic Peninsula, which is surrounded by water on three sides, have always turned back into the Peninsula when reaching water; the same is true of birds reaching the heavily harvested lands lying to the south of the Peninsula (E. D. Forsman, *personal communication*; D. W. Hays, *personal communication*). On the other hand, Spotted Owl juveniles have frequently been observed crossing heavily fragmented areas. We investigated this issue by classifying habitat into three categories (good, fair, poor) and recording cases in which dispersing juveniles encountered borders between better and worse habitat. We recorded whether the birds kept going, turned to stay in good habitat, or settled within a few kilometres of the border. In 40% of 27 cases, dispersing juveniles approaching a transition from better to worse habitat settled within a few kilometres or turned to stay in better habitat. In the rest of the cases, they continued into the poorer habitat. Slight evidence existed suggesting the birds were more likely to settle if they had travelled a longer, rather than shorter, distance from the nest, which would certainly not be surprising. These results suggest that the options described above should be set so that birds have a 30–50% chance of turning or settling when they encounter heavily fragmented habitat.

During periods of residence in winter and summer, birds were normally recorded within a circle of just 1 or 2 km diameter, but they occasionally ventured well outside this area for short periods (usually only one location). These excursions were recorded 2 times during birds' first winter, 4 times during each bird's first summer (i.e., when they were 12 mo old), and 3 times during birds' second winter. No differences in the distances moved during these periods were detectable. The distances had an average value of 13 km and a maximum value of 25 km. Such excursions are probably typical of most birds, particularly those who do not yet have a territory. The model simulates the information obtained by birds on these excursions by assuming that floaters learn of any vacancies in the territorial population within a user-specified distance of the floater's location. The data and rationale above suggest that a reasonable value for this distance for Spotted Owls is in the 15–30 km range.

Empirical data on the relationship between birth and survival rates and amount of habitat were provided by Bart and Forsman (1992), Thomas et al. (1990), and Bart et al. (1993). Significant relationships were found between birth rates and amount of habitat and between survival rates and amount of habitat. The estimates of adult survival rates are probably less reliable, because adults in areas with little suitable habitat may have emigrated and would usually have been counted as dead. This would have accentuated the apparent relationship between survival and habitat. On the other hand, a positive relationship between birth and survival rates and amount of suitable habitat is consistent with many other aspects of Spotted Owl biology (e.g., tendency to forage in old-growth and avoid young stands; declines in density in areas with little old growth), and the conservative assumption is certainly that both birth and survival rates decline as the amount of suitable habitat declines.

Limits on the birth and survival rates can also be established by recognizing that if the rates are too low then the floater population must decline (assuming that floaters fill vacant territories), but at present most populations appear still to have a substantial floater population. Consideration of these issues shows (Bart 1995) that if birth rates have been at the levels reported from field studies, and the first-year survival rate has been 0.40, for the past few decades, then the adult survival rate must have been at least 0.91 or 0.90. If it had been any lower, floaters would have disappeared by now.

Predictions of the model.—A detailed description of the realism of the model's secondary and primary predictions would occupy more space than is available here. The secondary predictions of interest include density, territory sizes and shapes, dispersal distances and patterns, and age ratio of first-time breeders. Parameter values in the model must be set so that all of these outputs conform to empirical evidence. The most important primary prediction to examine is the distribution and abundance of present-day owl populations when the model is started with conditions as they were in 1960 and habitat is removed in a realistic way.

Synthesis

Assumptions and parameter values for the model should be chosen to satisfy three general criteria:

1) Birth and death rates are within the ranges observed in field studies;

2) dispersal paths, territory sizes and shapes, and densities of territorial birds are similar to results from field studies;

3) projections using these values that start in 1960 yield present-day distributions and abundances of territorial owls and age ratios of first-time breeders that match results from field studies.

The criteria above should be used to identify a plausible parameter space. This space should then be sampled in a systematic manner to yield "best-case" and "worst-case" parameter sets, and these should be used to generate the predictions of future owl population performance.

The process above may show that the predictions of the model are sufficiently robust to existing uncertainties about the organism's behavior and demography that high confidence can be placed in the model's predictions. Alternatively, the evaluation may show that the primary predictions vary widely depending on which set of plausible assumptions is used as input values for the exercise. This might have the effect of reducing reliance on models, but it may also lead to greater acceptance of the modelling results because the rationale for the predictions will be clearer (Mankin et al. 1975). Furthermore, the process described above seems likely to have the salutary effect of reinforcing the view that models are "assumption analyzers" (a phrase suggested by Richard Holthausen) rather than black boxes that generate single predictions of unknowable reliability. As Botkin (1977:217) wrote, in a discussion of the use of computers in modelling, "By operating the model the computer faithfully and faultlessly demonstrates the implications of our assumptions and information. It forces us to see the implications, true or false, wise or foolish, of the assumptions we have made. It is not so much that we want to believe everything that the computer tells us, but that we want a tool to confront us with the implications of what we think we know." This view of modelling may be the most valid one of all, and the procedures above should help ensure that the assumptions, whose consequences we explore with the model, are clearly identified.

Acknowledgments

I thank S. Earnst, R. Holthausen, and P. Kareiva for comments on earlier drafts of the manuscript and R. Stehn, R. Mauck, and M. Hofschen for assistance in developing **MOSAIC**. The U.S. Fish and Wildlife Service provided financial support.

Literature Cited

Bart, J. 1995. Evaluation of population trend estimates calculated using capture–recapture and population projection methods. Ecological Applications, *in press.*

Bart, J., R. G. Anthony, M. Berg, J. H. Beuter, W. Elmore, J. Fay, R. J. Gutierrez, H. T. Heintz, Jr., R. S. Holthausen, K. Lathrop, K. Mays, R. Nafziger, M. Pagel, C. Sproul, E. E. Starkey, and J. C. Tappeiner. 1993. Recovery plan for the Northern Spotted Owl—final draft. United States Government Printing Office, Washington, D.C., USA.

Bart, J., and E. D. Forsman. 1992. Dependence of Northern Spotted Owls on old-growth forests. Conservation Biology **62**:95–100.

Berry, K. H. 1986. Introduction: development, testing, and application of wildlife habitat models. Pages 3–4 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

Botkin, D. B. 1977. Life and death in a forest: the computer as an aid to understanding. Pages 213–234 *in* C. A. S. Hall and J. W. Day, Jr., editors. Ecosystem modelling in theory and practice: an introduction with case studies. John Wiley & Sons, New York, New York, USA.

Caswell, H. 1976. The validation problem. Pages 313–325 *in* B. C. Patten, editor. Systems analysis and simulation in ecology. Volume 4. Academic Press, New York, New York, USA.

Cole, C. A., and R. L. Smith. 1983. Habitat suitability indices for monitoring wildlife populations—an evaluation. Transactions of the North American Wildlife and Natural Resources Conference **48**:367–375.

Conroy, M. J., Y. Cohen, F. C. James, Y. G. Matsinos, and B. A. Maurer. 1995. Parameter estimation, reliability, and model improvement for spatially explicit models of animal populations. Ecological Applications **5**:17–19.

Costanza, R. 1985. Articulation, accuracy and effectiveness of mathematical models: a review of freshwater wetland applications. Ecological Modelling **27**:45–68.

DeAngelis, D. L., and L. J. Gross, editors. 1992. Individual-based models and approaches in ecology. Chapman & Hall, New York, New York, USA.

Dedon, M. F., S. A. Laymon, and R.H. Barrett. 1986. Evaluating models of wildlife–habitat relationships of birds in black oak and mixed-conifer habitats. Pages 115–119 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

DeWitt, C. T., and J. Goudrian. 1974. Simulation of ecological process. Centre for Agricultural Publishing and Documentation, Wageningen, Netherlands.

Dunning, J. B., Jr., D. J. Stewart, B. J. Danielson, B. R. Noon, T. Root, R. H. Lamberson, and E. E. Stevens. 1995. Spatially explicit population models: current forms and future uses. Ecological Applications **5**:3–11.

Farmer, A. H., M. J. Armbruster, J. W. Terrell, and R. L. Schroeder. 1982. Habitat models for land use planning:

assumptions and strategies for development. Transactions of the North American Wildlife and Natural Resources Conference **47**:47–56.

Gardner, R. H., R. V. O'Neill, J. B. Mankin, and J. H. Carney. 1981. A comparison of sensitivity analysis and error analysis based on a stream ecosystem model. Ecological Modelling **12**:173–190.

Garrett, M. 1975. Statistical techniques for validating computer simulation models. U.S. International Biome Program, Grassland Biome Tehnical Report 286. Colorado State University, Fort Collins, Colorado, USA.

Gentil, S., and G. Blake. 1981. Validation of complex ecosystem models. Ecological Modelling **14**:21–38.

Grant, W. E. 1986. Systems analysis and simulation in wildlife and fisheries science. John Wiley & Sons, New York, New York, USA.

Hall, C. A. S., and J. W. Day, Jr. 1977. Ecosystem modelling in theory and practice: an introduction with case studies. John Wiley & Sons, New York, New York, USA.

Holt, R. D., S. W. Pacala, T. W. Smith, and J. Liu. 1995. Linking contemporary vegetation models with spatially explicit animal population models. Ecological Applications **5**:20–27.

James, L. D., and S. J. Burges. 1981. Selection, calibration and testing of hydrologic models. Monograph. American Society of Agricultural Engineers, St. Joseph, Michigan, USA.

Krebs, J. 1980. Ornithologists as unconscious theorists. Auk **97**:411.

Laymon, S. A., and R. H. Barrett. 1986. Developing and testing habitat-capability models: pitfalls and recommendations. Pages 87–92 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

Mankin, J. B., R. V. O'Neil, H. H. Shugart, and B. W. Rust. 1975. The importance of validation in ecosystem analysis. Pages 63–71 *in* G. S. Innis, editor. New directions in the analysis of ecological systems, Part 1. Simulation Councils Proceedings Series. Volume 5, number 1. Simulation Councils, Lajolla, California, USA.

Marcot, B. G., M. G. Raphael, and K. H. Berry. 1983. Monitoring wildlife habitats and validation of wildlife–habitat relationships models. Transactions of the North American Wildlife and Natural Resources Conference **48**:315–329.

McKelvey, K., B. R. Noon, and R. H. Lamberson. 1992. Conservation planning for species occupying fragmented landscapes: the case of the Northern Spotted Owl. Pages 424–450 *in* P. M. Kareiva, J. G. Kingsolver, and R. B. Huey, editors. Biotic interactions and global change. Sinauer, Sunderland, Massachusetts, USA.

Meyer, C. F. 1971. Using experimental models to guide data gathering. Journal of the Hydraulic Division, American Society Civil Engineers **10**:1681–1697.

Miller, D. R. 1979. Model validation through sensitivity analysis. Pages 292–295 *in* H. H. Shugart and R. V. O'Neil, editors. Systems ecology. Benchmark Papers in Ecology. Volume 9. Dowden, Hutchinson, & Ross, Stroudsburg, Pennsylvania, USA.

Miller, D. R., D. E. Weidhaas, and R. C. Hall. 1973. Parameter sensitivity in insect population modelling. Journal of Theoretical Biology **42**:263–274.

Miller, G. S. 1989. Dispersal of juvenile northern spotted owls in western Oregon. Thesis. Oregon State University, Corvallis, Oregon, USA.

Naylor, T. H., and J. M. Finger. 1971. Validation. Pages 153–164 *in* T. H. Naylor, editor. Computer simulation experiments with models of economic systems. John Wiley & Sons, New York, New York, USA.

Orlob, G. T. 1975. Present problems and future prospects of ecological modeling. Pages 283–312 *in* C. S. Russell, editor. Ecological modeling in a resource management framework. Resources for the Future Inc., Washington, D.C., USA.

Overton, W. S. 1977. A strategy of model construction. Pages 49–74 *in* C. A. S. Hall and J. W. Day, Jr., editors. Ecosystem modeling in theory and practice. John Wiley & Sons, New York, New York, USA.

Raphael, M. G., and B. G. Marcot. 1986. Validation of a wildlife–habitat relationships model: vertebrates in a Douglas-fir sere. Pages 129–138 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

Reckhow, K. H., and S. C. Chapra. 1983. Confirmation of water quality models. Ecological Modelling **20**:113–133.

Rotenberry, J. 1986. Habitat relationships of shrubsteppe birds: even "good" models cannot predict the future. Pages 217–222 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

Russell, C. S., editor. 1975. Ecological modeling in a resource management framework. Resources for the Future Inc., Washington, D.C., USA.

Schamberger, M. L., and L. J. O'Neil. 1986. Concepts and constraints of habitat-model testing. Pages 5–10 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

Starfield, A. M., and A. L. Bleloch. 1986. Building models for conservation and wildlife management. Macmillan, New York, New York, USA.

Straskraba, M., and A. H. Gnauck. 1985. Freshwater ecosystems modeling and simulation. Elsevier, New York, New York, USA.

Swartzman, G. L., and S. P. Kaluzny. 1987. Ecological simulation primer. Macmillan, New York, New York, USA.

Thomas, J. W. 1986. Wildlife-habitat modeling—cheers, fears, and introspection. Pages xix-xxv *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

Thomas, J. W., E. D. Forsman, J. B. Lint, E. C. Meslow, B. R. Noon, and J. Verner. 1990. A conservation strategy for the Northern Spotted Owl. United States Government Printing Office, Washington, D.C., USA.

Tomovic, R. 1963. Sensitivity analysis of dynamic systems. McGraw-Hill, New York, New York, USA.

Tomovic, R., and M. Vukobratovic. 1972. General sensitivity theory. Elsevier, New York, New York, USA.

Turner, M. G., G. J. Arthaud, R. T. Engstrom, S. J. Hejl, J. Liu, S. Loeb, and K. McKelvey. 1995. Usefulness of spatially explicit population models in land management. Ecological Applications **5**:12–16.

Verner, J., K. S. McKelvey, B. R. Noon, R. J. Gutierrez, G. L. Gould, Jr., and T. W. Beck. 1992. The California spotted owl: a technical assessment of its current status. General Technical Report PSW-GTR-133. Pacific Southwest Research Station, USDA Forest Service, Albany, California, USA.

Verner, J., M. L. Morrison, and C. J. Ralph, editors. 1986. Wildlife 2000. University of Wisconsin Press, Madison, Wisconsin, USA.

White, G. C., and R. A. Garrott. 1990. Analysis of wildlife radiotelemetry data. Academic Press, New York, New York, USA.