United States
Department of
Agriculture

Forest Service

Pacific Northwest
Research Station

General Technical
Report
PNW-GTR-232
MAY 1989

Alchemy and Uncertainty: What Good Are Models?

Fred L. Bunnell

Foreword

Resource managers in the United States and Canada must face increasing demands for both timber and wildlife. Demands for these resources are not necessarily incompatible with each other. Management objectives can be brought together for both resources to provide a balanced supply of timber and wildlife. Until recently, managers have been hampered by lack of technique for integrating management of these two resources. The goal of the Habitat Futures Series is to contribute toward a body of technical methods for integrated forestry in British Columbia in Canada and Oregon and Washington in the United States. The series also applies to parts of Alberta in Canada and Alaska, California, Idaho, and Montana in the United States.

Some publications in the Habitat Futures Series provide tools and methods that have been developed sufficiently for trial use in integrated management. Other publications describe techniques not yet well developed. All series publications, however, provide sufficient detail for discussion and refinement. Because, like most integrated management techniques, these models and methods have usually yet to be well tested, before application they should be evaluated, calibrated (based on local conditions), and validated. The degree of testing needed before application depends on local conditions and the innovation being used. You are encouraged to review, discuss, debate, and-above all-use the information presented in this publication and other publications in the Habitat Futures Series.

The Habitat Futures Series has its foundations in the Habitat Futures workshop that was conducted to further the practical use and development of new management techniques for integrating timber and wildlife management and to develop a United States and British Columbia management and research communication network. The workshop—jointly sponsored by the USDA Forest Service and the British Columbia Ministry of Forests and Lands, Canada—was held on October 20-24, 1986, at the Cowichan Lake Research Station on Vancouver Island in British Columbia, Canada.

One key to successful forest management is providing the right information for decisionmaking. Management must know what questions need to be asked, and researchers must pursue their work with the focus required to generate the best solutions for management. Research, development, and applicatipn of integrated forestry will be more effective and productive if forums, such as the Habitat Futures Workshop, are used to bring researchers and managers together for discussing the experiences, successes, and failures of new management tools to integrate timber and wildlife.

British Columbia Ministry of Forests and Lands — R.M. Ellis

U.S. Department of Agriculture, Forest Service — Richard S. Holthausen

Author

FRED L. BUNNELL is a professor of forest wildlife management, Faculty of Forestry, University of British Columbia, Vancouver, British Columbia V6T 1W5 .

Abstract

Bunnell, Fred L. 1989. Alchemy and uncertainty: what good are models? Gen. Tech. Rep. PNW-GTR-232. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 27 p.

Wildlife-habitat models are increasing in abundance, diversity, and use, but symptoms of failure are evident in their application, including misuse, disuse, failure to test, and litigation. Reasons for failure often relate to the different purposes managers and researchers have for using the models to predict and to aid understanding. This paper examines these two purposes and the nature of problems or failures in model application. A five-step approach toward solution is presented: (1) recognize the problem, (2) nurture modeling teams, (3) match purpose with test, (4) confront basic beliefs, and (5) evade known errors. Nurturing of modeling teams requires recognition of different risks, rewards, and timeframes of researchers and managers. Most wildlife-habitat models have mixed purposes (prediction and understanding). Each purpose should and can be evaluated separately. Some basic beliefs surrounding how we test statements are based more on faith than reason. These can be obstacles to using and testing models. Models of wildlife habitat are efforts to apply research to management; a better approach to using and testing models is a better approach to applying research on wildlife-habitat relations.

Keywords: Applied research, evaluation, habitat, models, wildlife.

Contents

1 **Introduction**
1 **The Problem**
2 Our Failures
4 Some Definitions
4 Historical Roots
5 Psychological Roots
6 Statistical Roots
7 The Root Mat
7 An Approach
8 **Nurturing Teams**
10 **Matching Purpose and Test**
12 Evaluating Prediction
13 Evaluating Understanding
13 Evaluating Wildlife-Habitat Models
14 **Doing Better**
19 **Summary**
22 **Acknowledgments**
22 **References**

Introduction

If only numbers and diversity measured success, wildlife-habitat models would be termed successful. Symptoms of failure, however, are evident. Waste, impeded growth of knowledge, and mismanagement are consequences of four symptoms of failure. Some models are being used for the wrong purposes. Some models, built at large cost, are not being used. Other models are never seriously evaluated. A few models have become "exhibit A" in litigation. These symptoms of failure are each related to how we (researchers and managers) choose the best use for and best test of models. We need more confidence in the uses and tests of models. This is the issue I examine here—the "validation problem" or "what good are models?"

I do not discuss all the reasons why models should be evaluated—Marcot and others (1983) present a compelling list. Nor do I list all means of validation; the Wildlife 2000 Symposium (Verner and others 1986) provides a large array. I focus on some broad reasons why wildlife-habitat models fail rather than problems with model structure, sampling, and statistics. I believe the broad reasons are associated with the different viewpoints of researcher and manager. As Walters (1986:viii) noted, "...management is done *by* people, as well as for people." I focus on people.

To some extent, the problems presented in this paper are more potential than real, but because they cause waste and other detrimental effects, the symptoms of failure suggest a need for concern. A few problems I discuss are just appearing. This paper is partly an early alert. My experience tells me models can be useful. Furthermore, we are approaching holy grails of modeling teams—the prediction of wildlife diversity or the abundance of specific species. Discussion of failures may seem negative; I discuss the negative to pursue the positive. I want to examine failures and learn from them, foresee failures and avoid them, and understand failures and eliminate them.

Psychological and statistical problems arise when asking, What good are models? Researchers and managers have different psychologies or worldviews as a result of their training and roles. These views influence their notions of purpose, risk, reward, and uncertainty and, thus, statistics and the appropriate ways of testing. I have attempted to capture this web of intertwined notions with the phrase "alchemy and uncertainty."

I view wildlife-habitat models as statements of how research findings apply to the management of wildlife habitats. Any approach to improving the use and testing of such models is an approach to applying research more effectively. I begin with the structure of the problem: failures and their historical, psychological, and statistical roots; then I examine an approach toward solution. The approach starts with the structure and nurture of modeling teams. Nurture of these teams requires us to recognize the utility of alchemy (not knowing why things work) and different kinds of uncertainty. The approach continues by noting that models can be evaluated only relative to their purpose (matching purpose and test). I recognize two broad purposes: to predict and to further understanding. The paper concludes with a list of 10 ways of accomplishing these purposes better and a summary.

The Problem

At the simplest level, the numbers and kinds of wildlife-habitat models are growing rapidly but so are instances where the models are perceived as wasteful. In the introduction, I noted four ways—symptoms of failure—by which modeling can waste time, talent, and funding.

Our Failures

Misuse—Models are being used for the wrong purpose. Such misuse often results from a misunderstanding of management and research goals. In 1973, 10 researchers gathered to collect their experience and data into a model of barren-ground caribou *(Rangifer tarandus;* Bunnell and others 1975). The kinds of data were diverse and had not been related before; combined they made some unexpected predictions about the influence of forest fires on caribou numbers. Output was soon used to justify approaches to northern fire management. The purpose of modeling was to examine interactions in the data collected, expose gaps in the data, and refine the research program. Confusion about the model's relative utility to research and management generated argument that is still healthy 15 years later. A more subtle case involved a simple model that related size of mammalian home ranges to body weight, food habits, and habitat productivity. The authors (Harestad and Bunnell 1979) provided, r^2 values and analyses of covariance. The model was not for prediction; the statistical tests were to document departures from previous published relations and differences between herbivores and carnivores. Some researchers misread the model's purpose and equated the highly significant r^2 values with predictability for specific instances rather than for general relations.

Disuse—Models may fall into disuse. This is good if a more useful tool is available. Disuse is wasteful, however, if building the model was expensive and the model was never used. Disuse is troubling if the model represents the best statement available. By definition most of these instances are invisible, but we all know of some. Four broad, related reasons exist why wasteful disuse occurs: (1) the model does not address the question or purpose of the potential user; (2) a potential user never existed, just someone who might be convinced; (3) the model asks the right question but is too complex to be used or requires too much data; and (4) overall output was not sufficiently accurate. In most instances, wasteful disuse occurs because of a poor understanding of the user's goal. It occurs most often when researchers build models for managers.

The fourth reason for disuse is often a poor one. Because most testing of models uses overall output, which is often the weakest part, models are sometimes abandoned that have some good features. One example is a very simple model of bear harvest *(Ursus* spp.; Bunnell and Tait 1980) that was developed for bear management and used by agencies in several States, Provinces, and Territories. It was too simple to incorporate all known, significant biological relations for bears, so it grew (Taylor and others 1987a). The demand came from researchers within Government agencies. The new model is likely the best current expression of bear population dynamics but requires at least 10 years of data to run. Researchers may keep tinkering with the model so it will serve one goal—aiding understanding. But researchers do not set harvest policy. The current complexity and data needs of the model almost ensure that it will never be used widely for its original, management purpose. Small pieces of the model are now being broken out and published as research articles (for example, Taylor and others 1987b).

Lack of evaluation—The third symptom of failure I note is not evaluating the model. Fortunately, this tends to discredit the model, but we need to ask if discredit was appropriate. At least four excuses exist for not evaluating models. First, it may take too long. Evaluations of the accuracy of overall output of certain large forest resource models (for example, FORPLAN [Hoekstra and others 1987] or TASS [Mitchell and Cameron 1985]) would take decades. We avoid this by evaluating pieces. Second, political will to enact the model's best policy and then monitor the outcome is lacking. Instead, we choose some compromise policy and never evaluate the model. I have addressed this problem elsewhere (Bunnell 1974, 1976b); it is not a trivial obstacle (Walters 1986:7). Third, users do not identify closely enough with the model's purpose to want to evaluate it. Wildlife-habitat models examining forest resources may require large-scale, long-term manipulation for evalution. If we have built the model well, the individuals charged with manipulating wildlife and habitat are among the users. If these users are not committed to the model's purpose, they will likely not be committed to long-term evaluation. Fourth, we do not know how to evaluate the model. I believe that credibility or confidence is important enough to the use of resource management models that potential approaches to evaluation must be considered before or during building of the model.

Legal challenges—The fourth symptom of failure in modeling is legal challenges to the validity of models and their use in making management decisions. The challenges have begun and will likely increase. These challenges include projected rates of sedimentation, elk *(Gervus elaphus)* population responses, and the sizes of viable populations from red cockaded woodpeckers *(Dendrocopus borealis)* to grizzly bears *(Ursus arctos).* Two characteristics are common among legal challenges to models. First, the model is held to be developed for a purpose other than that for which it is being used; it is ill-adapted to its use. Second, the assumptions or structure of the model are invalid; it has not been well tested and is not highly credible.

I discussed symptoms of failure in modeling in some detail because they are important challenges and contrary to other observations. The growing numbers and diversity of models and their potential uses are clearly evident in Verner and others (1986) and were evident at the 1986 Habitat Futures workshop. Such growing numbers and diversity unquestionably demonstrate that modeling is successful. The symptoms of failure, nonetheless, are equally real but just beginning to appear. Because modeling has so many advantages to contribute, failures must be examined as well as accomplishments. Two features are common among the problems previously listed—poor understanding of purpose and failure to evaluate the models usefully or convincingly. These common features are linked; we cannot evaluate effectively without knowing the purpose of the model. These features also have common, intertwined roots—historical, psychological, and statistical. If we understand the roots, the problem becomes soluble. Understanding requires some simple definitions.

Some Definitions

My taxonomy of models and model builders is simple; there are two kinds of each. Models have two tasks: models for prediction and models for understanding. For this elegance I am indebted to Caswell (1976). The distinction is similar to what Innis (1973) termed as "output utility" and "conceptual utility." Predictive models are designed to provide accurate, quantitative predictions of the response of one (or more) variable(s) (for instance, deer or cavity-nesting birds) to changes in another, often larger, group of variables (for instance, those describing a large collection of trees). Models for understanding represent combinations of best guesses or hypotheses in a theoretical statement about how the system operates; they are created to increase insight into the system's connections and behavior. The set of demands or task environment for these two kinds of models is different. In short, they must be evaluated for their "goodness" by different criteria and, probably, by different means.

The two kinds of modeling team members are researchers and managers (fig. 1). Walters (1986) would add policy analysts and decision makers to this team. Researchers create new concepts or expose new facts; managers make decisions about big yellow machines or other forces that change the landscape. It is no great step to equate the manager with the predictive model and the researcher with the model for understanding.

Historical Roots

Models explicitly relating wildlife to components of forested habitat appeared more than two decades ago (Bunnell 1974), and some early ones were relatively complex (Walters and Bunnell 1971). These models generally were the creation of teams of researchers and managers or were produced by researchers for managers. Many models were developed to guide management decisionmaking, and managers were best qualified to specify kinds of management actions and appropriate solutions. Researchers., however, usually had a piece of the puzzle and were best qualified to create appropriate relations between wildlife and its habitat. Teams arose, blending different skills and philosophies.

Three dramatic changes have occurred since the beginning of the development of wildlife-habitat models. First, computers allow us to misuse models at superhuman speed and produce impressive stacks of invalid output. Second, the number and variety of models has grown dramatically, challenging the diversity of species to be managed. Third, resource managers have begun using the models. Now more than ever we must test models effectively. Differences in psychological roots confuse our approach to testing.

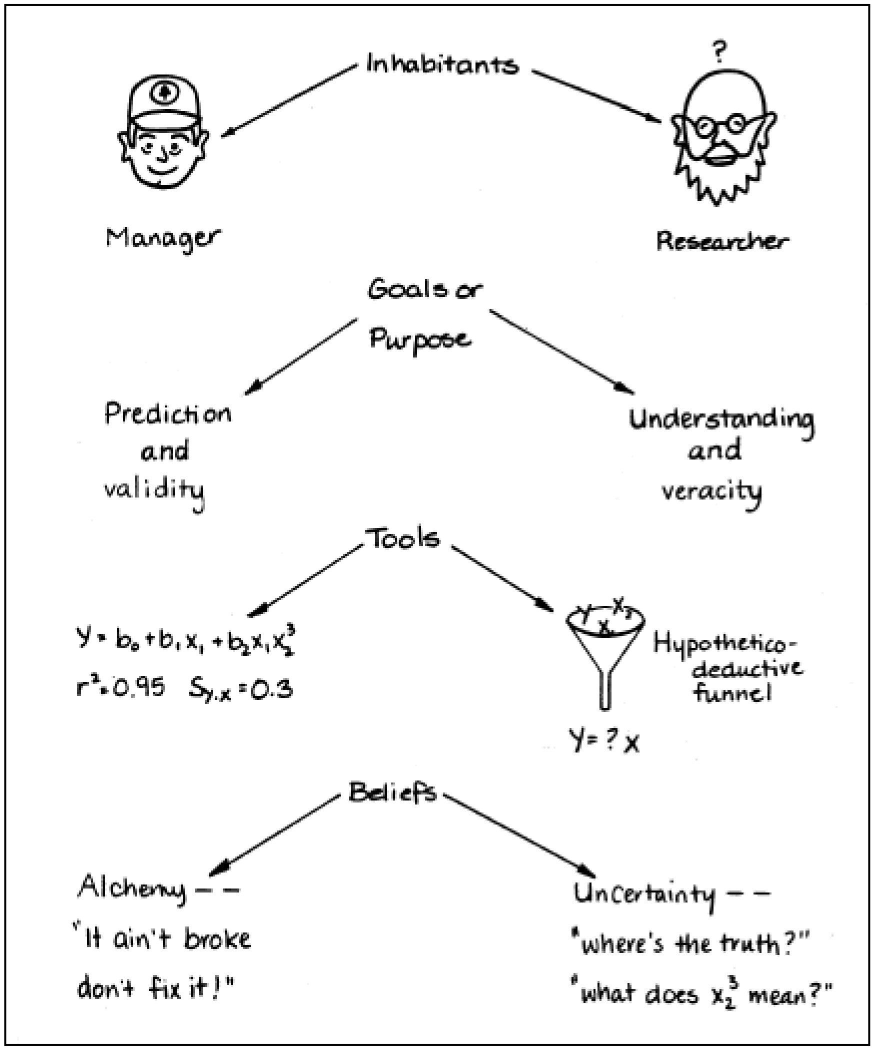

Figure 1-A simplified view of the world of wildlife-habitat modeling.

Psychological Roots

Most models are developed by teams or for managers by researchers. The manager seeks a timely, economical, easily applied tool to facilitate decision making in a complex system. Models are one such tool. Researchers see how their creations can help build that tool and thus aid decision-making. The model is a common goal. The problem is not the goal but the psychological roots brought to pursue and evaluate the goal. Ellis (1986) reviews the disparate needs, risks, and constraints among team members. He discusses Government foresters, industrial foresters, Government wildlife biologists, and academic researchers. I examine two team members: researchers and managers.

In my view, managers are big on alchemy and researchers are big on uncertainty (fig. 1). Alchemy ignores understanding; components are combined until they work. For example, if a multiple regression model could relate deer numbers to areas of different seral stage (for example, Bunnell 1979), many managers would not be concerned with understanding the relative roles of edge effects, forage versus cover, interspersion, and juxtaposition. These effects would be completely entangled in regression coefficients with peculiar units. If asked to wait until researchers untangle the coefficients, the manager's answer is likely to be, "If it ain't broke, don't fix it" (fig. 1). A test of the predictive powers or validity of such a model (for example, high r^2, low SE) would be sufficient for the manager. Many researchers would view such tests as faulty experimental design or poor technique and be very uncertain about the contributions to knowledge of such tests. Part of the researcher's training taught that the best way of separating knowledge from speculation is to deal with uncertainty by the hypothetico-deductive method (for example, Romesburg 1981). No clean hypothesis exists in the multiple regression model-it combines too many real processes.

Statistical Roots

A basic question is, "Can we validate the model?" To answer that question, the purpose or task of the model must be known.

Any model is an incomplete picture of reality (Bunnell 1973) and necessarily artificial (Simon 1969). Models of systems are systems themselves, comprised of mathematical variables and expressions (Caswell 1976, Walters 1971). The building of artificial systems is a design problem, and the process of design is essentially a search for agreement between features of the artificial system and a set of demands placed on it by the designer (Alexander 1964, Simon 1969). One cannot evaluate the success or failure of a design attempt without specifying the demands—what task is the model to perform?

I have reduced the tasks to two: prediction and understanding (fig. 1). So, can a model be validated? The short answer is yes and no, depending on the dictionary you use. If your dictionary defines "validate" as "sufficiently supported by actual fact" or "well grounded," you can probably agree on "sufficiently" or "well" and arrive at "yes." You could, however, have a dictionary that sends you from "validate" to "verify." Flipping to "verify," you encounter trouble because the dictionary may define it as "to establish the truth, accuracy or reality of." Accuracy can have remarkably little to do with truth or reality. Ellis (1986:619) recognized this point by terming the utility of research as "veracity" and that of development as "validity."

The basic form of a purely predictive model is probably multiple regression (fig. 1). If potential independent variables are evaluated stepwise, they can be included or rejected on the basis of statistics that indicate their contribution to the accuracy of the model. Statistics describing the model will show how well predictions can be made over the range of variables measured. This ability to predict can be termed "validity." It can be a powerful management tool. But notice-nowhere is there any hint of testing the truth or reality of the model. This empirical approach focuses on statistical validity of model output; the conceptual level is hidden from direct testing (Jacoby and Kowalik 1980, Marcot and others 1983). For this reason, I have termed its utility alchemical, a kind of magic. We do not know why it works.

Models for understanding are different and seek to increase our ability to explain how nature works. Recent articles (for example, Gill 1985, McNab 1983, Romesburg 1981) indicate that wildlife researchers are attempting to adhere more closely to the hypothetico-deductive method and falsifiable hypothesis of Popper (1959). That approach forces us to conclude that no method exists to show a theory (model) to be true or to have complete veracity.

So, can a model be validated? If the model is for prediction with accuracy and precision the important criteria, the answer is, Yes, it is relatively easy. The notion of truth does not plague us, and criteria can be established. If the model is for understanding and obtaining truth, the answer is no (Bunnell 1973, Popper 1959). The model however, can be corroborated (see the section, Matching Purpose and Test).

The Root Mat

The three types of psychological roots are entangled. If wildlife-habitat modeling efforts are to be pointed in a useful direction and ultimately used, both researchers and managers must be involved at the outset (for example, Bunnell 1974, Ellis 1986, Thomas 1979, Walters 1986). The different skills of researcher and manager often have been brought together and blended successfully. Blended contributions also have created part of our problem. The different psychologies of team members assign different purposes to the goal of modeling (for example, Ellis 1986), and evaluations must acknowledge the purpose (Caswell 1976).

Even if pursuit of the goal or model was harmonious, evaluating the success of that pursuit may not be. What initially seems to be successful integration may inadvertently encourage conflicts in beliefs. When the potentially embarrassing question is addressed, "What good is this model?" team members may follow different priesthoods: one with a strong bent toward alchemy and the other toward the pursuit of truth (fig. 1). The type of team member characterized, as "researcher" will be repelled by the alchemical but delighted if the model increases understanding by some unmeasurable amount. The type characterized as "manager" will want to know how much confidence can be placed in the model. Depending on the model's purpose and appropriate means of evaluation, the best evaluation may be unfair, unhelpful, and probably irritating to one or another type.

An Approach

The four kinds of failure of wildlife-habitat models I noted at the outset have common elements of unclear purpose or inadequate evaluation. Each failing can be avoided if the purpose is clear and an appropriate test of the model can be devised. Different, entangled roots underlie the general problem. These roots are to some extent inseparable, but an approach can be taken to distinguish them. First, recognize the need and possibility for better evaluating models. Next, separate the psychological and statistical problems as best as possible; we must recognize competing purposes and psychologies embedded within models. We then need to nurture modeling teams, in part by matching purpose with test, while confronting basic beliefs. Wildlife-habitat models have mixed purposes, but each can be evaluated separately. Finally, evade errors to which tests of wildlife-habitat models are particularly susceptible. Wildlife-habitat models are statements of applied research. The approach outlined for implementing these models is one of applying research more effectively.

Nurturing Teams

The need for teams, rather than individuals, to address applied research problems has been noted by many (for example, Bunnell 1974, 1976a, 1976b, 1985; Callaham 1984; Ellis 1986; Thomas 1979; Walters 1986). Teams are necessary in part because nature is not divided up as university curricula are or as Government management agencies are (Bunnell and Dumont 1973) and in part because single ownership of a problem encourages competition and conflict. "It is, therefore, usually necessary to impose some credible structure [team] with a fixed life span that is charged with solving the problem" (Bunnell 1985:183); in this case, applying research. Teams of both researchers and managers are not only necessary, but they also can reduce changes of failure and, more importantly, can be fun, exciting, and great learning opportunities. Consider first reducing failures.

The structure and nurture of modeling teams can reduce the chance of each of the four failings noted. Misuse of the model suggests a misread purpose. Disuse of the model suggests no interest in the model's purpose or design. Lack of evaluation and legal challenge suggest misread or abused purpose and difficulties in evaluation, which can be entangled in purpose.

Failure to use the model or lack of interest is perhaps the easiest to correct. The cure is simple and most simply stated by Peters and Waterman (1982:14): get "close to the customer." "Close" means more than within reach of the customer's wallet. It means designing the model to meet the customer's needs. We do that best by being close enough to understand both the needs and their reasons. Validity and simplicity are not synonyms of veracity. Failure often occurs when researchers dominate model building, and the search for understanding produces a tool with little validity or that is too complex to use easily. Managers' needs for validity and simplicity must be shared with research team members; conversely, the researchers' attention to detail should help to ensure credibility.

Failure through misread use or misuse of a model may arise either because the purpose of the model was never clear or different team members perceived different purposes. Many, perhaps most, models have incorporated both the search for accurate prediction and understanding into their objectives. The researcher usually has not held a hidden agenda, suppressing pursuit of truth while overtly developing a management tool. More often, both manager and researcher have recognized that the predictive device had to incorporate some best guesses or theoretical constructs.

When it is time to evaluate a model, researchers naturally want to examine how well their best guesses withstand scrutiny. Managers want simplicity and a reliable measure of confidence. The problem remains that evaluation of understanding or of prediction may use different means. If the manager chooses to evaluate for predictability when the researcher wants to evaluate for understanding, the most likely outcome is trouble.

The broad cause of all these potential failures is the same: different psychologies and purposes of team members. Researchers tend to value cerebral anarchy and function best when unrestrained. They may bemoan the notion that "demand pull" sees tools used more quickly and more often than "supply push," but the notion appears true (for example, Callaham 1984). So, focus on the fun and opportunities. Managers would usually rather be right than consistent. Researchers like to see their results applied. The opportunities are there; nurture them.

First, develop the model so it belongs to everybody (Bunnell 1974). It may not be neutral, but it is impersonal, and it can't fight back. Criticize it, test it appropriately; you learn from that. Second, negotiate limits to the cerebral anarchy. Understanding responses of managed systems is not necessarily increased by more basic research or more detailed analysis (see Walters 1986 for compelling examples). Also, researchers can provide some quick best guesses while intensively probing other relevant questions (Bunnell 1985). The astute manager will not use the pull of demand to completely restrain the researcher's quest. Doing so, the manager could lose the most creative researchers and miss some important biological truths (veracity) necessary for decisionmaking. Third, exploit adaptive management to test the model. Walters (1986) provides the best summary of adaptive management. Briefly, the potentials of natural populations are learned through experience with management itself rather than through basic research or through development of general ecological theory. The latter two should be used to guide the management experiment. The manager's manipulation can exploit the evolution of the researcher's best guesses. Both groups benefit and a growing edge is maintained on research and management (Bunnell 1976b, 1985). Fourth, recognize the different senses of risk, time, and reward that team members have (Bunnell 1985, Callaham 1984, Ellis 1986).

Multiple purposes of the model are more troublesome than incomplete communication. Only partial solutions exist. A major step is to recognize that different people may perceive different goals for the model. The goals themselves are equally valid and valuable but often must be evaluated differently. Both kinds of goals can be evaluated in the same model or in different portions of the model. Similarly, several significant, detailed relations can be condensed into simpler ones that may increase the model's validity and management utility, but this will decrease veracity (not all the biological pieces will be represented).

One useful approach is to use different forms of the same model for research and management purposes (fig. 2). For example, one detailed model of the operative temperature of black-tailed deer (Parker and Gillingham 1986) uses about 250 lines of computer code. It is a useful tool for asking how different abiotic variables influence the deer's cost of thermoregulation and represents the inner research track of figure 2. For management purposes this detail can be reduced to one multiple regression equation that explains 95 percent of the variation accommodated by the original model (Hovey and Bunnell, in press). The detailed model helps to define areas where managers should be concerned about thermal cover; the simpler model can explore broad consequences if that concern modifies forestry practice.

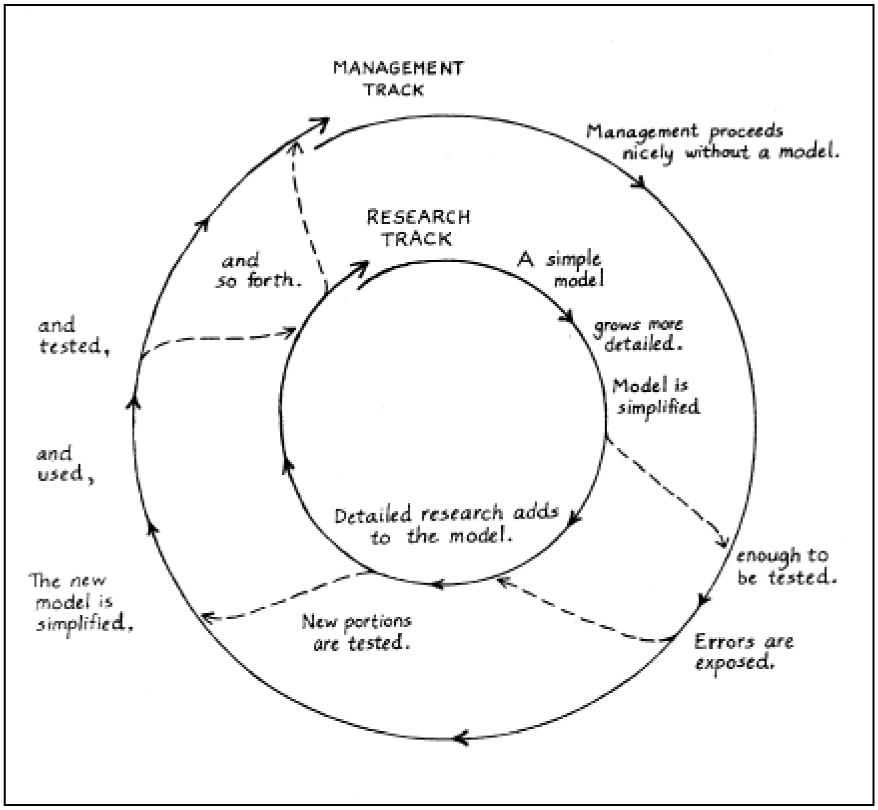

Figure 2—Relations between detailed research models and simpler management models. Testing the models through adaptive management benefits both researcher and manager.

There are five important points to having separate research and management tracks (fig. 2) in modeling. First, the simpler model often cannot be built without the detailed one. Second, development, testing, and use of both models can proceed at their own pace. Their development is spread through time. Third, the manager benefits from the researcher's refined tool (credibility), and the researcher benefits from the manager's use of the simpler tool (testing). Fourth, the two major purposes are still mixed in each model, but one purpose dominates each. Fifth, the mixed purposes may still require separate tests (for example, fig. 3). The separation of tracks (fig. 2) recognizes different purposes, evaluates them differently, but combines their benefits; the team approach is thereby nurtured.

Matching Purpose and Test

Many criteria exist for evaluating or testing models; Marcot and others (1983) summarized 23. Criteria must fit the purpose of the model. Success of managers depends on their ability to make accurate predictions. The appropriate criterion for testing a management model is thus the ability to make empirically correct predictions in all instances. Mankin and others (1975) and Gass (1977) called this criterion "validity." Success of researchers depends on their ability to describe new but true relations. The appropriate criterion is "veracity," but no one knows how to measure that. Other of the criteria summarized by Marcot and others (1983) will be appropriate; for example, "usefulness" (at least some model predictions are empirically correct; Shrank and Holt 1967) will be important to managers.

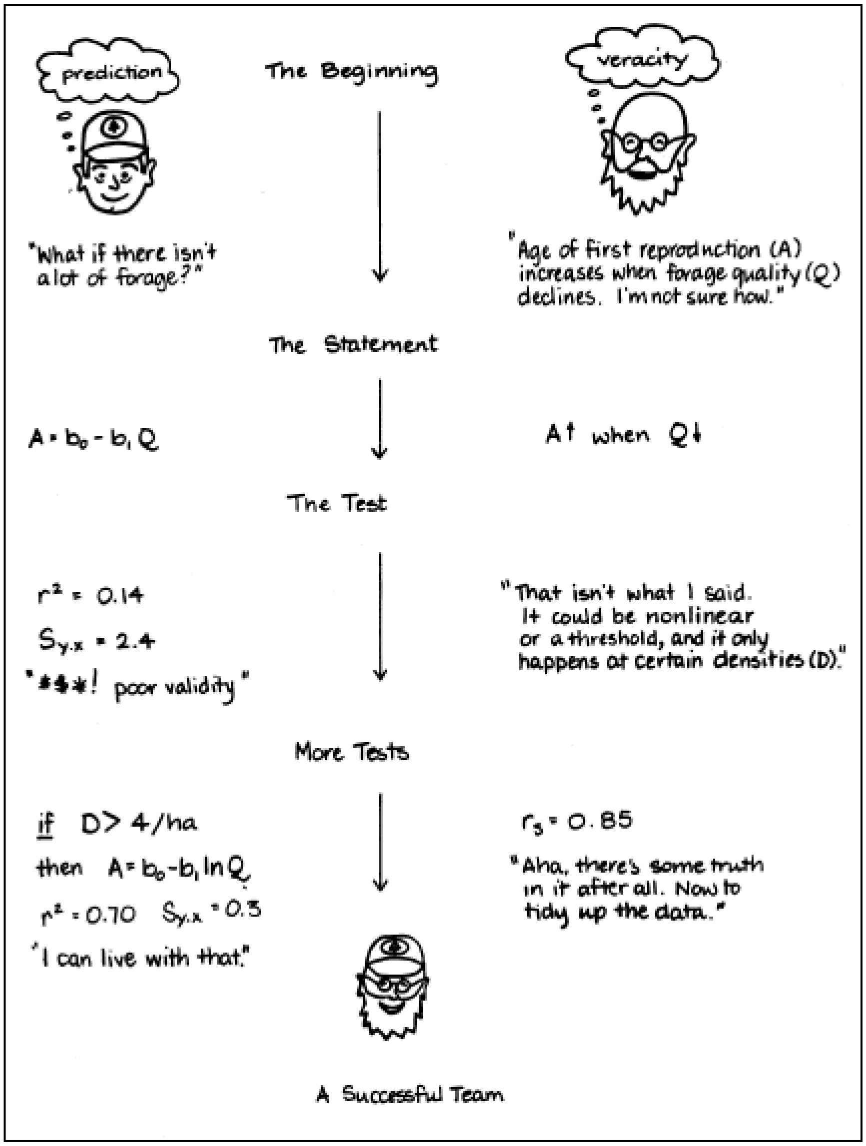

Figure 3— Testing a wildlife-habitat model with two purposes: prediction and understanding. For maintaining a successful team, the purposes should be tested differently.

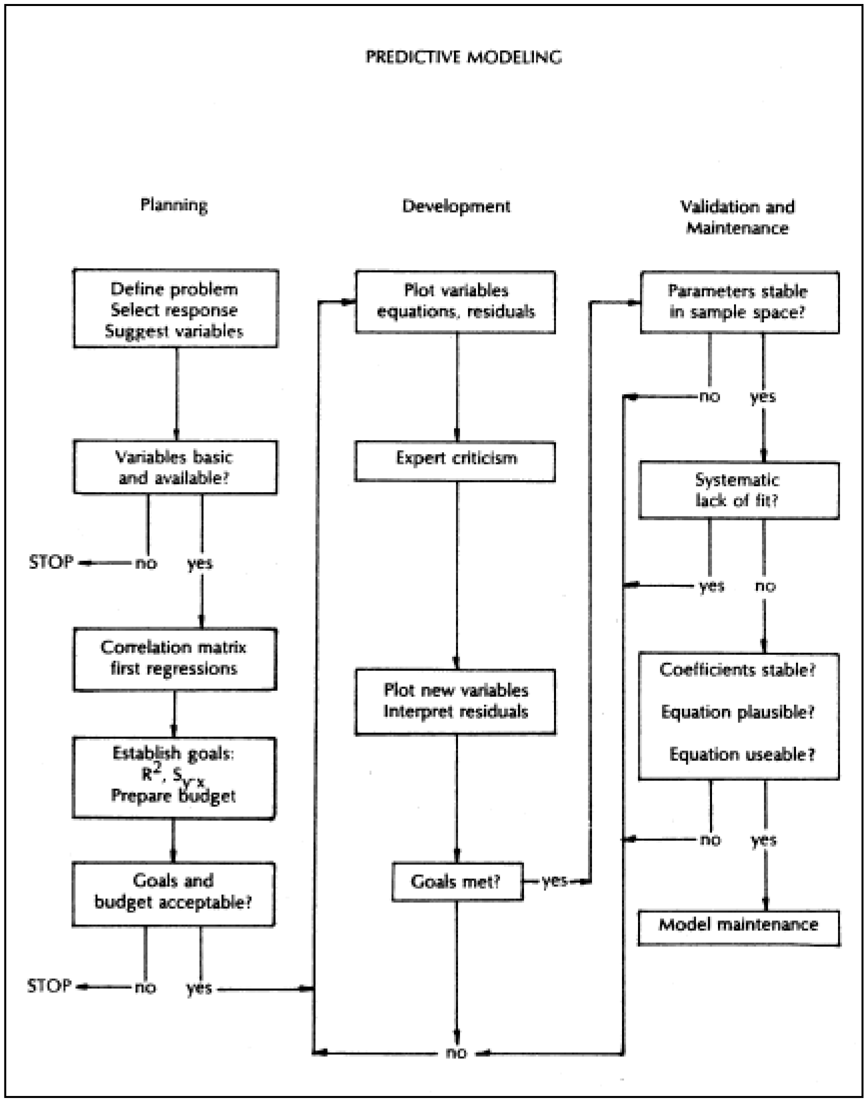

Figure 4—Flow chart showing steps in construction of a predictive model by multiple regression analysis (adapted from Draper and Smith 1966).

Evaluating Prediction

Simple predictive models may be as simple as volume or yield tables (the guts of FORPLAN) and are usually empirically based. Their validity can be tested simply by asking how closely does the model fit the data; later tests compare the model to another body of data. Appropriate statistics are those testing goodness of fit and include r^2, R^2, i^2, and the standard error of the regression $s_{y \cdot x}$. Figure 4 illustrates the approach as it could be applied to a multiple regression model.

With sufficient data, a similar approach can be used with simulation models having dynamic interactions. One compares time series of output with observed data, or compares the output under different known and measured conditions. Goals for accuracy or precision can be assigned in terms of (1) the costs of improving the model versus costs of a given level of error, R^2, $s_{y \cdot x}$, or (2) some measure of goodness of fit (fig. 3). We compute the appropriate statistic and evaluation is easy, sometimes.

Actually, simulation models can be very hard to test for their predictive powers by standard statistics. One way suggested by several (for example, Farmer and others 1982, Marcot and others 1983, McKenny 1967, Van Horn 1969) is the "Turing test" (Turing 1950). Professionals try to distinguish between unlabeled summaries of model predictions and empirical observations. This is a kind of "soft testing" (as defined by Marcot and others 1983) in which standard statistics are not used. As they note, such testing involves "...the deceptively difficult task of articulating [assumptions and theories] in clear and precise statements, and then assessing their limitations and pertinence to the modeling objectives" (Marcot and others 1983:319). Such tests may help little in litigation.

Evaluating Understanding

Models pursuing understanding or truth cannot be validated with. criteria like those of figure 4. We can, however, corroborate the model. My dictionary defines "corroborate" as "to strengthen or give additional strength to; to make more certain; to add assurance to." Note, the difference is one of philosophy, not of statistics. Many criteria within the list of Marcot and others (1983) apply equally to validation and corroboration. The difference is we cannot establish quantifiable measures of veracity. No one is prepared to state that if the model predicts correctly 90 percent of the time, it represents truth; if it is accurate only 75 percent of the time, it is untrue. Truth eludes the operational definitions of validity. How much difference between model and data is required for rejection is unknown; no matter how many tests the model passes, it may fail the next test.

Philosophers of science have recognized the problem (for example, Popper 1963:33) as have modelers (for example, Bunnell 1973:170, Caswell 1976:320-21, Schamberger and O'Neil 1986:8). "As long as your model output is wrong you continue to learn something. Once it fits the data you do not know what to believe. It is then predictive but it will provide no further insight [understanding] into the dynamics of the system. ...The model is of most use [to seeking understanding] when it is clearly wrong" {Bunnell 1973:170-171). Stated simply, we have no way of recognizing truth when we find it or of measuring how close we are to it.

Evaluating Wildlife-Habitat Models

Most wildlife-habitat models have mixed purposes. That is a problem only if we ignore it. Most tests of these models have used overall output and general ability to predict; assumptions, particularly about how key relations are combined, were embedded unchallenged within the-evaluation. The test of predictive ability (validity) on overall output ignores or obscures potential tests of understanding at finer levels. But we can validate or corroborate a model at any level, including tests of assumptions, variables, components, and overall output (for example, Schamberger and O'Neil 1986). The tests differ at different levels, become less robust at higher levels of aggregation, and are always less rigorous in efforts to corroborate than in efforts to validate.

Efforts to evaluate current and future wildlife-habitat models should consider six points. First, even if mixing the two purposes (prediction and understanding) in models is bad form, it is usually unavoidable. Second, to assign one dominant purpose to a mixed-purpose model and then proceed with testing is poor form because you lose information and alienate team members. Third, we benefit from testing at all levels of model structure and from recognizing the distinct purposes combined in the models. Fourth, we should recognize the utility of models or model segments that have been refuted (untrue) but validated (predict with a specified degree of accuracy), or have been corroborated (respond in the right direction) but invalidated (inaccurate). We should keep the useful parts. Fifth, a researcher can quantify the accuracy or precision in a predictive model for a manager but cannot quantify the degree to which the model has been corroborated or strengthened. We should recognize the distinction; it avoids useless discussion. Sixth, because most wildlife-habitat models contain assumptions about the nature of interactions, the model is a theoretical statement. It can be treated, and sometimes tested, as a predictive model with overall output; but such tests ignore the theory involved in the model and may contribute little to understanding. We should apply more than one kind of test at different levels and specific to each purpose.

In the example illustrated (fig. 3), the manager has used tests that quantify validity but with help from the researcher has constrained these to the context in which the relation holds. The researcher's quest has encompassed a larger range. He doesn't know the form of the relation and has used weaker statistics appropriate to ordinal data. He is confident that his general understanding is correct but cannot predict over this range. The different tests have matched the different purposes (fig. 1) and produced a successful team (fig. 3).

Doing Better

We need both a strategy and tactics to test our models. The major strategic step is to recognize different purposes in the model. The tactics must incorporate that recognition.

Were I Moses, I would propose 10 commandments for evaluating models; I'm not and the following are suggestions:

1. Recognize the psychological problems. Because the model has theory or guess embedded in it, Popper's use of corroboration is appropriate:

> So long as a theory withstands detailed and severe tests and is not superseded by another theory in the course of scientific progress, we may say that it is corroborated (Popper 1963 :33).

We never find truth, but we can corroborate. Alchemy has an important role; it permits management.

2. Recognize the practical problems. Clear criteria for prediction or validity (r^2, standard error of regression, and cost of improving the model versus the cost of a given error) can be developed by team members early in the modeling process (for example, fig. 3). Clear criteria for testing veracity do not exist. Statistical criteria for rejection are usually not available for overall output, and without such criteria it is difficult to decide how much disagreement between model and data is required for rejection (Bunnell 1973, Caswell 1976). Strictly, the process of testing for understanding must continue indefinitely; no matter how many tests the model has passed, it may fail the next one. Some of the 23 criteria for validation presented by Marcot and others (1983), however, will be applicable. Do not use just one kind of test but do match purpose and test.

3. Complexity potentially can increase model realism (for example, Best and Stauffer 1986, Diehl 1986, Rotenberry 1986), but it will make testing and application more difficult (Bunnell 1973, Walters 1986).

   Primarily because we know so little about the interaction of key habitat and wildlife components, an increase in complexity often increases the number of guesses or assumptions (fig. Sa). Attempts at corroboration, particularly overall output, cannot assign errors to a specific model component or relation. The result is a decrease in understanding with increasing complexity (fig. 5b). I first presented figure 5b in 1973 based on experience and a superficial survey of models. Costanza and Sklar (1983) evaluated the notion rigorously for 87 models. They concluded that model effectiveness really is a dome-shaped function of model complexity: we can predict a few things well or a lot of things very poorly; between these extremes is some balance where reasonable accuracy is maintained without losing too much biological realism (see Walters 1986: 185-190). Note that the peak in the curve of figure 5b will shift to the right as initial understanding increases. Recognition of that can exploit the two-track system of figure 2.

   I refer here to adding "big pieces" to the model—another habitat component. Adding finer levels of detail (for example, physiology versus whole animal) may not be helpful for different reasons. In most current models, the larger, slowly moving pieces generally reflect the dynamics of the smaller, faster moving pieces sufficiently well for useful prediction Thus the simplest models can be helpful (fig. 5a). By treating only big pieces, however, understanding is limited to questions that can be asked of these pieces (for example, shrubs, deer, trees) and any "why" question (for example, Why is thermoregulation important here?) is ignored.

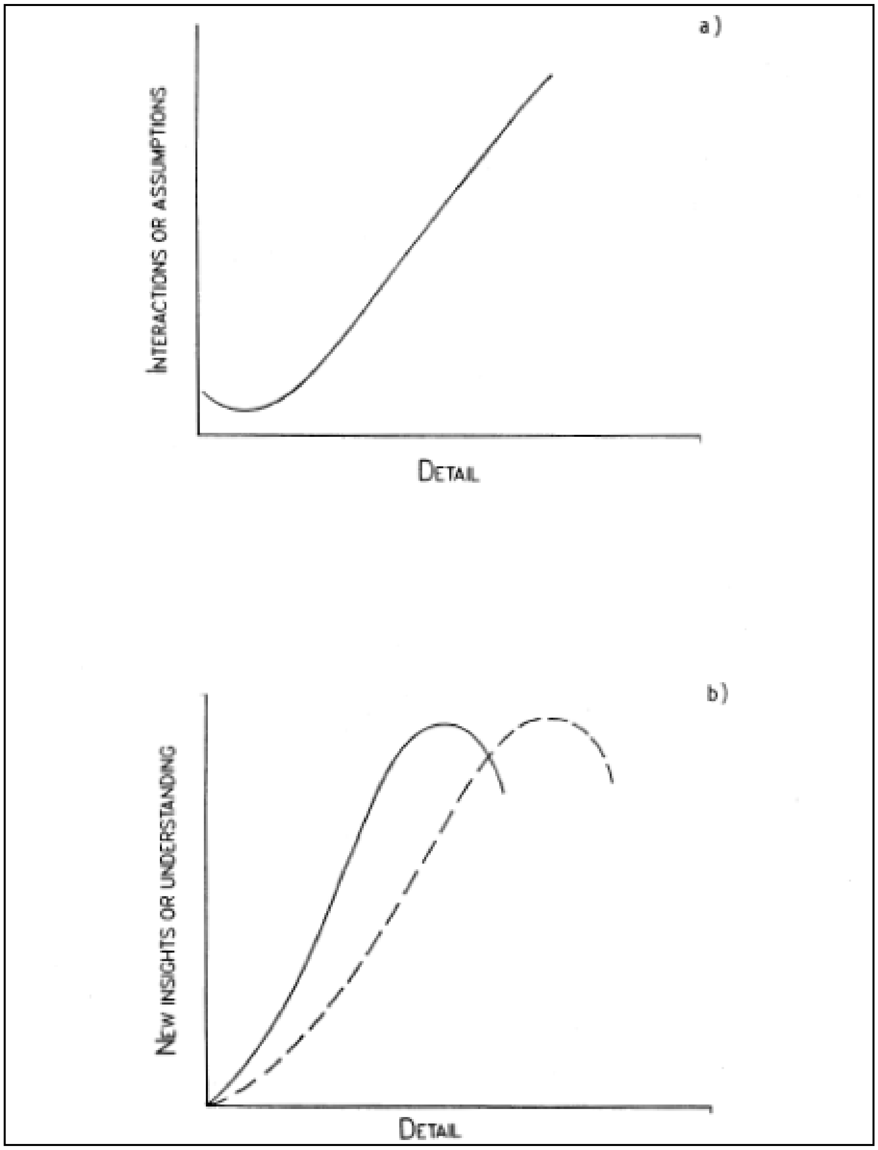

Figure 5–Relations between detail, interactions, and understanding in models (adapted from Bunnell 1973).
a) As complexity or detail in the model increases, the number of interactions and assumptions increases exponentially. The lowest level of detail can introduce new assumptions.
b) As detail increases, the effectiveness of the model in prediction and contributing to understanding initially increases but then declines as a function of initial understanding of the system modeled. The solid line indicates little initial understanding of the system model; the dashed line indicates greater initial understanding.

4. Test the pieces.

   Tests of entire models are generally appropriate only for prediction and then may lack appropriate statistics. Tests for understanding will do better with smaller pieces (for example, fig. 5) that include any interaction constraints. This approach is akin to the suggestion of Schamberger and O'Neil (1986) that comprehensive testing should proceed from assumptions to overall output. Such an approach to evaluation is time-consuming and costly (Caswell 1976; Overton 1977) but may provide the most productive results in terms of model improvement or corroboration (Schamberger and O'Neil 1986). If understanding and prediction are to progress, it is important to assign error to a piece of the model, not necessarily the entire model (for example, fig. 3). But note, tests of management utility still will be based on prediction of the larger process, not its parts. Tests of pieces primarily address understanding of the parts when separated. Almost all failures of predictability cited for models deal with overall output or system response. That failing in validity does not mean there is no truth or veracity within different levels of the model. Test apparent truth at each level.

5. Test alternatives.

   Popper (1959) noted that theories should not be tested in a vacuum but compared to alternative theories. The approach has been called "multiple hypotheses" (Chamberlin 1897) and "strong inference" (Platt 1964). Platt outlines the approach as:

   i) devise alternative hypotheses,

   ii) devise a crucial experiment, the outcome of which will exclude at least one of the hypotheses,

   iii) carry out the experiment, and

   iv) return to i), refining the problem further and further.

   I prefer to term these alternative hypotheses explicitly as competing hypotheses. For example, we might view an apparent "edge-effect" for deer as resulting from either (1) some desired component (for example, forage) is better near the boundary of two types or (2) each type is providing a desirable resource and deer are sufficiently lazy that their distribution away from the edge declines quickly. Explicit statements of competing hypotheses generally assist the design of crucial experiments or sampling schemes. It is often difficult to decide whether model behavior is sufficiently aberrant to merit rejection; it is much easier to decide which of two models is better. Unfortunately, alternative models are remarkably rare in ecology (Caswell 1976; see also Verner and others 1986).

6. Use designed experiments.
   Even if these experiments can be applied only to portions of models, they often provide the strongest tests. Strong inference with competing hypotheses can be a powerful aid to experimental design. Remember, management is an experiment; it can be cautious or probing. If things really are complex out there, we will not extend understanding until we test complex hypotheses. Methods are available. Factorial experiments can isolate causes and measure their interactions (Cochran and Cox 1957). Response-surface methods (Box and others 1978) and logistic regression (Neter and Wasserman 1974) are applicable. Provided one has alternative models, comparison of simulations can be used (for example, Gilbert and others 1976), although the statistics may be less rigorous (but see also Walters 1986).

7. Use sensitivity analyses sparingly.
   Sensitivity analysis could be employed to rank the parameters to which a misbehaving variable is most sensitive. These are then the prime candidates for adjustment. The concept makes sense only for predictive models. Such adjustments represent *ad hoc* modifications of model statements and assume only a small subset (sensitive parameters) of the statements is open to correction (Caswell 1976). The approach is unhelpful if the model is to aid understanding.

8. Begin early and proceed forever.
   A prevailing impression seems to be that validation begins when the model is finished. As Platt (1964) and Caswell (1976) observed, generating alternative models, devising tests, and revising models should be carried out component by component, interaction by interaction while the model is being constructed. This notion is implicit in the observations of Bunnell (1974) and Schamberger and O'Neil (1986). Uncertainty diminishes slowly, and we will live for years with models that do not inspire strong confidence. The ultimate test will still be that of predicting responses to management actions. Moderate confidence is better than little confidence. While testing and extending confidence, the modeling team must continue to "service what it sells." Improvements and useful simplifications should be incorporated.

9. Remember the mathematics and the context.
   The researcher may have stated variable y (for example, deer) varies in the same direction as variable x (for example, red huckleberry plus salal biomass). The model might express this as $y = a + bx$, which is a much more explicit and quite different statement (for example, fig. 3). A test of this latter predictive statement might invalidate it; but a test of the initial statement itself might be corroborated. The test must be appropriate to the statement, not some subsequent mathematical transformation. The wrong test can alienate contributors and users.

   It is equally important to test different contexts. The limited "yes" or "no" approach of simple hypotheses does not test the context in which the model predicts well or poorly. We want to know whether it does better on some sites than others, for some species than others, or ion some seasons than others.

10. Don't throw it all away.
    I have often stated that models should be disposable, but it is not enough to reject a model and throw it away. Walters (1986:44) called such rejection the "all or nothing" myth. The myth is especially dangerous for wildlife-habitat models. Southwood (1977:359) warned, "It is important that we do not visualize habitat as a rigid causal templet (or template in the engineering sense)." Add management actions to the biological variability, and perfec_ prediction is impossible. We must be able to assign fault, at least tentatively, to some portion of the model. It is highly unlikely it is all wrong. Unfortunately, the majority of our past tests have employed the entire model (Schamberger and O'Neil 1986, Verner and others 1986); exceptions do exist (for example, Blenden and others 1986, Brush and Stiles 1986). We can improve our ability to progress incrementally by paying particular attention to points 4, 5, 6, and 9 above.

Summary

Were I about to begin a career building wildlife-habitat models, the preceding comments might dissuade me. Happily, however, I am two decades into it; there's no turning back. Also happily, there are enough maverick individuals to squash the potential problems and retain the utility of modeling. I also know many examples where modeling has benefited management and research. Were models not being used, the potential problems discussed would not exist. We still have a lot to learn about statistical approaches to testing models, but that is not my concern here. We can do better at using what we already have, and we can begin that now. Most of the problems, potential and real, can be overcome. They will not, however, be overcome by ignoring them. Steps to take in overcoming them are steps in applying our research better to wildlife and forest management. I list the following steps as a kind of line up—ducks in a row. They do not represent targets for management or research marksmen, but hurdles we should surmount together. Here is the lineup.

1. Relax, we've just begun.
   About 20 years ago, Charles Elton (1966:62) forcibly argued that "Definition of habitats, or rather lack of it, is one of the chief blind spots in zoology..." Physicists have been playing physics for about 3,000 years; we have just begun with habitat. You have gotten in early. Enjoy the excitement of new ventures.

2. Learn to nurture teams.
   We need teams because both managers and researchers are part of the solution. In wildlife-habitat management they are each other's clients. Their different skills can be most effectively exploited if their different senses of risk, time, purpose, and rewards are recognized. In its rightful place (management), nothing is wrong with alchemy, but it has little place in the quest for understanding. The team, however, needs both. Respect those needs.

3. Treat models as tools not goals.
   The goal is to apply research findings usefully to predict management effects. Moving toward that goal, the manager wants confidence in his or her prediction, the researcher wants to understand why the prediction worked or failed. The model is but one tool to reach the goal. The researcher should realize results can often be applied without understanding them. Art precedes understanding. The manager should recognize that tests of validity may contribute little to understanding. Confusion about how best to handle the tool should not be transferred to the common goal. The model is simply one feature of adaptive management. Explicit models of dynamic behavior spell out assumptions and predictions clearly enough that errors can be detected and used as a basis for further learning.

4. Match the purpose and test.
   Most wildlife-habitat models will have two purposes—prediction and understanding. Any test is purpose-specific. A model cannot be evaluated outside its task environment. The distinction between validity and veracity must be clear. The researcher often feels that inability to predict (validity) undermines veracity; actually they are separate features. Models often reach well beyond relations about which researchers are confident. Likewise there is some truth in the notion that our efforts to simplify can hide real complexity and create surprises. Select tests appropriate to each purpose (for example, fig. 3). A test of only one purpose can alienate or mislead team members. If a model, is rejected for predicting poorly, it is not necessarily a poor contribution to understanding (see "commandments"' 4, 5, 6,9, and 10). Similarly, the successful manager's bias for action will move the manager faster than the researcher creeping up on understanding. It is useful to recognize this explicitly in modeling procedures (for example, fig. 2). The manager may, however, extract an excellent best guess from the researcher providing it is a guessing game and not an assault on truth; the latter makes the researcher more hesitant.

5. Confront theologies.
   When we consider testing models, some of our basic beliefs are founded more on metaphysics than observation. They constitute "theologies." Among them are (1) the "religion of. statistics" (for example, Salsburg 1985) with its adherence to strict hypotheses that tend to ignore context and (2) a kind of "uncertainty principle" among some ecologists that concludes system responses cannot be predicted. These "theologies" can impede applied research and management and are obstacles to effective testing of models. Although they probably contain some element of reason, if not truth, it is good to confront them. Any worthwhile theology can profit from doubt. Besides we will not effectively develop and test our models without confronting them. Fortunately, we can exploit the strengths of these theologies by directing them to appropriate tests. Strict adherents of the "religion of statistics" can test the model's simple assumptions or individual relations; those awed by the ecologists' "uncertainty principle" can develop the complex tests of interactions.

6. Test the models.

Techniques useful for evaluating predictive capability have been applied to wildlife habitat models: multiple regression (Blenden and others 1986, Hammil and Moran 1986), logistic regression (Smith and Connors 1986), discriminant functions (Block and others 1986, James and Lockerd 1986, Mosher and others 1986). The techniques are most correctly applied when evaluating model assumptions (for example, Blenden and others 1986) or model components (Hammil and Moran 1986). These techniques and others noted under the 10 "commandments" are also appropriate for testing understanding. We have no good excuse for not evaluating models better.

7. Watch out for ambush.

Even if we apply our statistics well, we can be ambushed during model tests. Nature is sufficiently full of surprises and uncertainties that we will never evade all of them. At least three sources of ambush are lurking for wildlife-habitat models: regression towards the mean when we select extreme habitats to test, misapplying interval or ratio-scale tests to ordinal scales of measurement, and the use of physical-condition indices as a wildlife-response variable (Bunnell 1987; Hovey and Bunnell, in preparation). Undoubtedly more exist (see Walters 1986), but if we evade these, we will have done well.

8. Recognize management as experiment.

To do otherwise ignores reality. It also encourages unhealthy distinctions between researcher and manager. Walters (1986:7) summarized it best:

> The fact that adaptive learning through management 'experiments' may proceed much more quickly than through conservative management and basic research has been noticed by some practicing managers for many years and has helped fuel an unhealthy split and mutual contempt between managers and researchers in many agencies. This split makes the valuable basic advances that do occur much more difficult to put into practice, and isolates researchers from the wealth of experimental opportunities afforded by whole-system manipulations by managers.

The split deprives both managers and researchers of an important "growing edge" (Bunnell 1985). We keep confronting questions that only hard experience can answer; no amount of testing component processes will suffice. Predictive models fail. We need to ask "…whether to use management policies that will deliberately enhance [hard] experience" (Walters 1986:vii). Management by experiment, with models as one tool, reduces both time and cost of testing models (for example, fig. 2). Calls for such action are not new (Bunnell 1974, 1976b; Ellis 1986; McNab 1983). We have the techniques (some noted above) and descriptions of the general process (Holling 1978, Walters 1986). We should begin in earnest. Monitoring will be necessary.

Acknowledgments

This paper derives from experience provided by research and resource management agencies in 10 countries. Those in Canada have been particularly generous in providing opportunities to learn: the Natural Sciences and Engineering Council of Canada (NSERC), the Canadian Wildlife Service, the Science Council of British Columbia, and the British Columbia Ministries of Forests and Lands and Environment and Parks. The Council of Forest Industries in British Columbia and NSERC provided a Senior Industrial Fellowship that allowed me to visit the world of managers. Many individuals have challenged my biases and encouraged rethinking, but I am particularly indebted to Rick Ellis and Carl Walters. Like many doing applied research at the University of British Columbia, I owe a special debt to Ralf Yorque.

References

Alexander, Christopher. 1964. Notes on the synthesis of form. Cambridge, MA: Harvard University Press. 216 p.

Best, Louis B.; Stauffer, Dean F. 1986. Factors confounding evaluation of bird-habitat relationships. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat" relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 209-216.

Blenden, Michael D.; Armbruster, Michael J.; Baksett, Thomas S.; Farmer, Adrian H. 1986. Evaluation of model assumptions: the relationships between plant biomass and arthropod abundance. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 11-14.

Block, William M.; Brennan, Leonard A.; Gutierrez, R.J. 1986. The use of guilds and guild-indicator species for assessing habitat suitability. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 109-113.

Box, George E.P.; Hunter, William G.; Hunter, J. Stuart. 1978. Statistics for experimenters. New York: John Wiley and Sons. 653 p.

Brush, Timothy; Stiles, Edmund W. 1986. Using food abundance to predict habitat use by birds. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 57-64.

Bunnell, F.; Dauphine, D.C.; Hilborn, R. [and others]. 1975. Preliminary report on computer simulation of barren-ground caribou management. In: Luick, J.R.; Lent, P.C.; Klein, D.R.; White, R.G., eds. Proceedings, The first international reindeer and caribou symposium; 1982; Fairbanks, AK. Special Rep. 1. Fairbanks, AK: University of Alaska: 189-193.

Bunnell, F.L. 1974. Computer simulation of forest-wildlife relations. In: Black, H.C., ed. Wildlife and forest management in the Pacific Northwest. Corvallis, OR: School of Forestry, Oregon State University: 39-50.

Bunnell, F.L. 1976a. Forestry-wildlife: whither the future. Forestry Chronicle. 52(3): 147-149.

Bunnell, F.L. 1976b. The myth of the omniscient forester. Forestry Chronicle. 52(3): 150-152.

Bunnell, F.L. 1985. Forestry and black-tailed deer: conflicts, crises, or cooperation. Forestry Chronicle. 61(2): 180-184.

Bunnell, F.L. 1987. Ambushed by desire: potential errors in testing wildlife-habitat models. Technical Report. Vancouver, BC: Faculty of Forestry, University of British Columbia. 16 p.

Bunnell, F.L.; Tait, D.E.N. 1980. Bears in models and reality—implications to management. In: Martinka, Clifford J.; McArthur, Katherine L., eds. Bears—their biology and management. Proceedings, 4th international conference on bear research and management; 1977 February; Kalispell, MT: The Bear Biology Association: 15-23.

Bunnell, Fred. 1973. Theological ecology or models and the real world. Forestry Chronicle. 49(4): 167-171.

Bunnell, Fred L. 1979. Deer-forest relationships on northern Vancouver Island. In: Wallmo, O.C.; Schoen, J.W., eds. Sitka black-tailed deer: Proceedings of a conference; 1978 February 22-24; Juneau, AK. Series R10-48. Juneau, AK: U.S. Department of Agriculture, Forest Service, Alaska Region: 86-101. In cooperation with State of Alaska, Department of Fish and Game.

Bunnell, Fred; Dumont, William E. 1973. Foresters, forestry and wildlife. Forestry Chronicle. 49(2): 66-70.

Callaham, Robert Z. 1984. Managing for applications, not just for research and development. Journal of Forestry. 82(4): 224-227.

Caswell, H. 1976. The validation problem. In: Patten, B.C., ed. Systems analysis and simulation in ecology. New York: Academic Press: 313-325.

Chamberlin, T.C. 1897. Studies for students. The method of multiple working hypotheses. Journal of Geology. 5: 837-848.

Cochran, William G.; Cox, Gertrude M. 1957. Experimental designs. 2d ed. New York: John Wiley and Sons, Inc. 611 p.

Constanza, R.; Sklar, F.H. 1983. Mathematical models of freshwater wetland and shallow water ecosystems: an articulated review. Unpublished paper prepared for SCOPE international conference on freshwater wetlands and shallow water bodies; 1983 August, Tallinn, U.S.S.R.

Diehl, Barbara. 1986. Factors confounding predictions of bird abundance from habitat data. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 229-334.

Draper, N.R.; Smith, H. 1966. Applied regression analysis. New York: John Wiley and Sons. 407 p.

Ellis, R.M. 1986. The role of wildlife habitat research in forest management in Canada. In: The 18th IUFRO World Congress, Division 1, Volume II, Forest Environment and Silviculture; 1986 September 7-21; Ljubljana, Yugoslavia. Ljubljana, Yugoslavia: Yugoslav IUFRO World Congress Organizing Committee: 611-624.

Elton, Charles. 1966. The pattern of animal communities. London: Methuen & Co. Ltd. 432 p.

Farmer, Adrian H.; Armbruster, Michael J.; Terrell, James W.; Schroeder, Richard L. 1982. Habitat models for land use planning: assumptions and strategies for development. In: Sabol, Kenneth, ed. Transactions of the 47th North American wildlife and natural resources conference; 1982 March 26-31; Portland, OR. Washington, DC: Wildlife Management Institute: 47-56.

Gass, Saul I. 1977. Evaluation of complex models. Computers and Operations Research. 4(1): 27-35.

Gilbert, N.; Gutiérrez, A.P.; Fraser, B.D.; Jones, R.E. 1976. Ecological relationships. Redding, CA: W.H. Freeman. 156 p.

Gill, R. Bruce. 1985. Wildlife research—an endangered species. Wildlife Society Bulletin. 13(4): 580-587.

Hammill, James H.; Moran, Richard J. 1986. A habitat model for ruffed grouse in Michigan. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates, Madison, WI: University of Wisconsin Press: 15-18.

Harestad, A.S.; Bunnell, F.L. 1979. Home range and body weight—a re-evaluation. Ecology. 60(2): 389-402.

Hoekstra, Thomas W.; Dyer, A.A.; LeMaster, Dennis C. tech. eds. 1987. FOR PLAN: an evaluation of a forest planning tool. Gen. Tech. Rep. RM-140. Fort Collins, CO: U.S. Department of Agriculture, Forest Service. 164 p.

Holling, C.S., ed. 1978. Adaptive environmental assessment and management. New York: John Wiley and Sons. 377 p.

Hovey, Fred W.; Bunnell, Fred L. [In preparation]. Potential limitations of physical condition indices as measures of habitat quality.

Hovey, F.W.; Bunnell, F.L. [In press]. A simulation model of effects of silvicultural practices on black-tailed deer energetics during winter. IWIFR. Victoria, BC: B.C. Ministry of Environment and Parks and B.C. Ministry of Forests and Lands. 21 p.

Innis, George. 1973. Simulation of ill-defined systems: some problems and progress. Simulation Today. 2(1): 33-36.

Jacoby, Samuel L.S.; Kowalik, Janusz S. 1980. Mathematical modeling with computers. Engelwood Cliffs, NJ: Prentice-Hall. 292 p.

James, Douglas A.; Lockerd, M. Joseph. 1986. Refinement of the Shugart-Patten-Dueser model for analyzing ecological niche patterns. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 51-56.

McKenny, James L. 1967. Critique of: "Verification of computer simulation models." Management Science. 14(2): B102-B108.

McNab, John. 1983. Wildlife management as scientific experimentation. Wildlife Society Bulletin. 11: 397-401.

Mankin, J.G.; O'Neill, R.S.; Shugart, H.H.; Rust, B.W. 1975. The importance of validation in ecosystem analysis. In: Innis, George S., ed. New directions in the analysis of ecological systems. Part 1 : Simulation Councils Proceed. Series 5(1): 63-71.

Marcot, Bruce G.; Raphael, Martin G.; Berry, Kristen H. 1983. Monitoring wildlife habitat and validation of wildlife-habitat relationships models. In: Sabol, Kenneth, ed. Transactions, 48th North American Wildlife and Natural Resources Conference; 1983 March 19-24; Kansas City, MO. Washington, DC: Wildlife Management Institute: 315-329.

Mitchell, K.J.; Cameron, I.R.; 1985. Managed stand yield tables for coastal Douglas-fir: initial density and precommercial thinning. Land Management Rep. 31. Victoria, BC: B.C. Ministry of Forests. 69 p.

Mosher, James A; Titus, Kimberly; Fuller, Mark L. 1986. Developing a practical model to predict nesting habitat of woodland hawks. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 31-36.

Neter, John; Wasserman, William. 1974. Applied linear statistical models: regression, analysis of variance, and experimental design. Homewood, IL: Richard D. Irwin, Inc. 842 p.

Ovenon, W. Scott. 1977. A strategy of model construction. In: Hall, Charles A.S.; Day, John W., Jr., eds. Ecosystem modeling in theory and practice: an introduction with case histories. New York: John Wiley and Sons: 49-74.

Parker, K.L.; Gillingham, M.P. 1986. Estimating thermal environments for mule deer *(Odocoileus hemionus hemionus).* WHR-24. Victoria, BC: B.C. Ministry of Environment and Parks. 63 p.

Peters, Thomas J.; Waterman, Roben H. Jr. 1982. In search of excellence: lessons from America's best run companies. New York: Harper and Row. 360 p.

Platt, John R. 1964. Strong inference. Science. 146(3642): 347-353.

Popper, Karl R. 1959. The logic of scientific discovery. London, England: Hutchinson and Co., Ltd. 480 p.

Popper, Karl R. 1963. Conjectures and refutations: the growth of scientific knowledge. London: Routledge and Kegan Paul. 412 p.

Romesburg, H. Charles. 1981. Wildlife science: gaining-reliable knowledge. Journal of Wildlife Management. 45{2): 293-313.

Rotenberry, John T. 1986. Habitat relationships of shrubsteppe birds: even "good" models cannot predict the future. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 217-222.

Salsburg, David S. 1985. The religion of statistics as practiced in medical journals. The American Statistician. 39: 220-223.

Schamberger, Melvin L.; O'Neil, L. Jean. 1986. Concepts and constraints of habitat model testing. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 5-10.

Shrank, William E.; Holt, Charles C. 1967. Critique of: "Verification of computer simulation models." Management Science. 14{2): B104-B106.

Simon, Herbert A. 1969. The sciences of the artificial. Cambridge, MA: M.I.T. Press. 123 p.

Smith, Kimberly G.; Connors, Peter G. 1986. Building predictive models of species occurrence from total-count transect data and habitat measurements. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 45-50.

Southwood, T.R.E. 1977. Habitat, the templet for ecological strategies. Journal of Animal Ecology. 46{2): 337-365.

Taylor, M.K.; Bunnell, F.L.; DeMaster, D.P. [and others]. 1987a. ANURSUS: a population analysis system for polar bears. International Conference Bear Research and Management. 7: 117 -125.

Taylor, M.K.; Carley, J.S.; Bunnell, F.L. 1987b. Recruitment rates of age-structured populations with multi-annual reproductive cycles. Marine Mammal Science. 3(2): 171-179.

Thomas, Jack Ward. 1979. Introduction. In: Thomas, Jack Ward, tech. ed. Wildlife habitats in managed forests—the Blue Mountains of Oregon and Washington. Agric. Handb. 553. Washington, DC: U.S. Department of Agriculture, Forest Service: 10-21.

Thomas, Jack Ward. 1986. Wildlife-habitat modeling-cheers, fears, and introspection. In: Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press. xix-xxv.

Turing, A.M. 1950. I. Computing machinery and intelligence. Mind. 59: 433-460.

Van Horn, Richard. 1969. Validation. In: Naylor, Thomas H., ed. The design of computer simulation experiments. Durham, NC: Duke University Press: 232-251.

Verner, Jared; Morrison, Michael L.; Ralph, C. John, eds. 1986. Wildlife 2000: modeling habitat relationships of terestrial vertebrates. Madison, WI: University of Wisconsin Press. 470 p.

Walters, Carl J. 1971. Systems ecology: the systems approach and mathematical models in ecology. In: Odum, E.P., ed. Fundamentals of ecology. Toronto: W.B. Saunders Company: 276-292.

Walters, Carl J. 1986. Adaptive management of renewable resources. New York: MacMillan Publishing Company. 374 p.

Walters, Carl J.; Bunnell, Fred. 1971. A computer management game of land use in British Columbia. Journal of Wildlife Management. 35(4): 644-652.

★ U.S. GOVERNMENT PRINTING OFFICE: 1989—693-563

Bunnell, Fred L.1989. Alchemy and uncertainty: what good are models? Gen. Tech. Rep. PNW-GTR-232. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 27 p.

Wildlife-habitat models are increasing in abundance, diversity, and use, but symptoms of failure are evident in their application, including misuse, disuse, failure to test, and litigation. Reasons for failure often relate to the different purposes managers and researchers have for using the models to predict and to aid understanding. This paper examines these two purposes and the nature of problems or failures in model application. A five-step approach toward solution is presented: (1) recognize the problem, (2) nurture modeling teams, (3) match purpose with test, (4) confront basic beliefs, and (5) evade known errors. Nurturing of modeling teams requires recognition of different risks, rewards, and timeframes of researchers and managers. Most wildlife-habitat models have mixed purposes (prediction and understanding). Each purpose should and can be evaluated separately. Some basic beliefs surrounding how we test statements are based more on faith than reason. These can be obstacles to using and testing models. Models of wildlife habitat are efforts to apply research to management; a better approach to using and testing models is a better approach to applying research on wildlife-habitat relations.

Keywords: Applied research, evaluation, habitat, models, wildlife.

Pacific Northwest Research Station
319 S.W. Pine St.
P.O. Box 3890
Portland, Oregon 97208

U.S. Department of Agriculture
Pacific Northwest Research Station
319 S.W. Pine Street
P.O. Box 3890
Portland, Oregon 97208

Official Business
Penalty for Private Use, $300

BULK RATE
POSTAGE +
FEES PAID
USDA-FS
PERMIT No. G-40

Do NOT detach Label