CHAPTER

Validation of Landscape-Scale Decision Support Models That Predict Vegetation and Wildlife Dynamics

16

Stephen R. Shifley, Chadwick D. Rittenhouse, and Joshua J. Millspaugh

Models have long been used to assess habitat quality and to predict how wildlife populations would respond to changes or manipulations in habitat (Verner et al. 1986, Morrison et al. 1998). In general, most modern approaches assess the value of habitat by relating a species' needs of food, cover, and other necessities to characteristics of vegetative cover types and other landscape features (e.g., distance to edge, patch size, interspersion of habitat features). Habitat suitability models evaluate the resource attributes considered important to a species' presence, abundance, survival, or reproduction and often result in predictive maps that illustrate suitability (Fig. 16-1), which may then be used to make relative comparisons across management alternatives (Gustafson et al. 2001, Marzluff et al. 2002). Ideally, habitat quality would be summarized in terms of demographic processes (Van Horne 1983, Johnson 2007), but this is often not the case due to insufficient data to measure and model demographic processes.

Models can incorporate the effects of spatial patterns of habitat quality on wildlife population viability by modeling demographic parameters of one or more populations such as carrying capacity, population size, and fecundity as a function of characteristics of the patches or landscape they occupy (Akçakaya 2001, 2002, Larson et al. 2004). However, the locations of habitat patches, populations, or individual home ranges change over time due to processes such as succession, natural disturbance, or anthropogenic alteration (Akçakaya et al. 2004). Therefore, changes in wildlife habitat over time in response to management actions or inaction are important determinants of wildlife population viability. Such broad-scale assessments require the use of dynamic landscape models

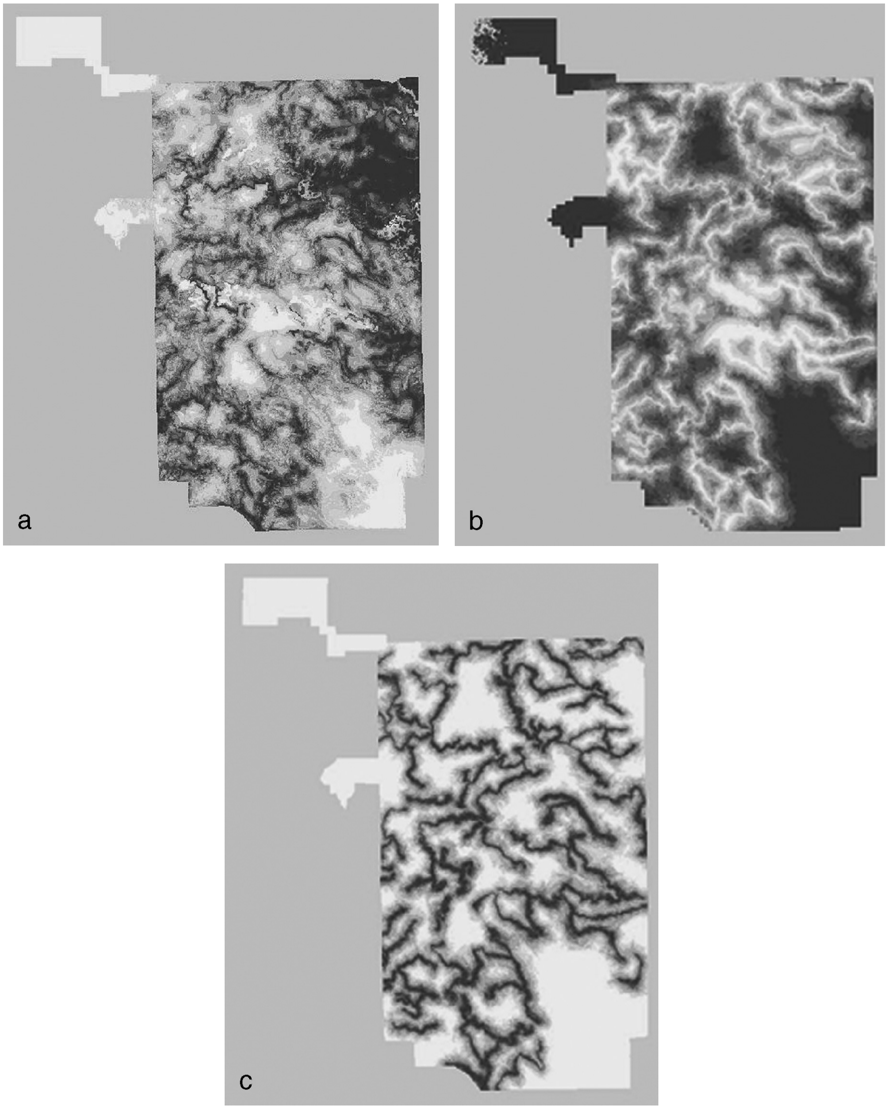

FIG. 16-1

Resource selection functions of (A) elk (*Cervus elaphus*), (B) mule deer (*Odocoileus hemionus*), and (C) white-tailed deer (*Odocoileus virginianus*) in Custer State Park, South Dakota, USA, during summer as determined by radiotelemetry techniques and logistic regression (Woeck 2003).

capable of projecting future habitat conditions (Akçakaya and Brook, this volume; Bekessy et al., this volume; He, this volume).

Landscape models, like all models, are abstractions of reality and should be validated to determine their utility. Under the best circumstances, models will capture the most important features and processes of the real system (Gentil and Blake 1981). But all models are imperfect; they are constrained by imperfect knowledge of the system and limited by resources (e.g., data, computational, human, and financial) for model building or execution (Miller et al. 1976). The challenge is finding or creating a model or group of models that strike a practical balance between usability and capability for one's intended purpose (Millspaugh et al., this volume). Box (1979) stated that "all models are wrong, some models are useful." The essence of landscape model validation and verification is not to determine if a model is wrong—we already know that. Rather we are interested in knowing whether the model is good enough for an intended purpose and whether it is superior to the alternative models that are available (Conroy and Moore 2002). In the case of wildlife habitat models, they represent our best knowledge of animal-habitat relationships. However, the form of the relationships between habitat quality and life history attributes may be unknown. Thus, it may be difficult to validate wildlife habitat suitability models or even to understand what constitutes validation (Van Horne 2002).

Much of our current understanding about simulation model development and validation in ecology is derived from methods originating in engineering and physics. Often model applications in those fields have outcomes that are straightforward and easier to test than those from ecological models (Cartwright 1983). For example, a beam that buckles, a bridge that collapses, or parts that fail with unacceptable frequency give unambiguous feedback on the utility of the models used in their design. The landscape-scale vegetation and wildlife models that are addressed in this book typically lack such clear indicators of outcome. It is inherently difficult to quantify habitat quality for multiple species across a large landscape at a single point in time, particularly when there is no obvious currency (or metric) and methodology for each species. It is even more difficult to quantify how well a model predicts changes in habitat in response to management for those same species. Consequently, with landscape-scale vegetation and wildlife models, we rarely have simple outcomes corresponding to a simple pass or fail. Rather we accumulate strengths and weaknesses or domains where the model appears to be useful and domains where it does not (Starfield and Bleloch 1991).

Evaluation of landscape-scale models is a complex process (Scott et al. 2002). Numerical or statistical prediction accuracy (Mayer and Butler 1993) is only one component of model evaluation (Hamming 1975, Hurley 1986). There are practical and theoretical considerations associated with model evaluation that may eliminate a model from consideration regardless of its statistical accuracy and precision. In this chapter we discuss the important factors that should be considered when evaluating a model for its intended purpose. Model evaluation proceeds with imperfect information (Starfield 1997) and is an iterative process; over time our understanding of model capabilities, data needs, and limitations improves. Models become more useful when applied in an adaptive management framework (Millspaugh et al., this volume). We describe a general process and present specific examples from the literature to help guide evaluation of landscape-scale vegetation and wildlife habitat suitability models. There is not a single recipe suitable for all applications, but the evaluation process will help model developers and model users arrive at a realistic assessment of a model's strengths, limitations, and utility for a stated purpose (Rykiel 1996). Related issues of model transparency, repeatability, and incorporating uncertainty in data are addressed elsewhere in the book (Millspaugh et al., this volume).

GENERAL CONCEPTS AND TERMINOLOGY

Three concepts should be clearly understood in the model validation process. First, evaluating a model in the absence of plainly stated objectives is meaningless. Landscape vegetation and wildlife suitability models can be used to gain insight into the theoretical underpinnings of system processes; to make general predictions of change over large areas and long time frames (Shifley et al. 2006); to make specific predictions for a small area (Marzluff et al. 2002); to address a single species of concern (Akçakaya et al. 2005); to evaluate trade-offs among multiple species (Marzluff et al. 2002, Root et al. 2003, Noon et al., this volume); to make a quick estimate based on readily available data; to establish the critical parameters for a long-term monitoring program; or for countless other purposes. Only when the objectives of model application are known can a model be evaluated and compared with alternatives (Johnson 2001).

Comparison among alternative models is a second core concept in model validation (Conroy and Moore 2002). There are always alternative models. However, alternative models are not always quantitative, and that can make comparisons difficult. For example, "conceptual models" (i.e., mental models) are always an initial point of comparison. Experienced managers may have highly refined conceptual models that work well in specific situations or for some species of wildlife. When conceptual models based on experience are all we have, they invariably are used to make management decisions because they are better than nothing. And "nothing" is the ultimate starting point for model comparisons. With GIS models applied in large landscapes, we sometimes even create a "nothing" or "null" landscape with random patterns and use that as a starting point for model comparison.

Finally, when one is quantitatively evaluating model performance, the currency (i.e., the primary metric) of the model output must be explicitly stated, because it must be measured objectively and compared to observations (Conroy and Moore 2002). When "habitat suitability" is defined in only vague terms, it becomes difficult to validate a model because a variety of competing independent data sources could be used, each offering a different view of model utility. There is an important difference in validating models that predict suitability and those that predict animal abundance. For example, failure of predicted suitability to relate to animal use or abundance does not necessarily mean the model is performing badly if the animal population is well below carrying capacity.

Consequently, additional knowledge about population status is often helpful. Validation with vital rates is sometimes better than with presence-absence data.

The literature relevant to model validation arises from numerous disciplines, and definitions of terms like model validation, verification, or evaluation vary among authors and applications (Morrison and Hall 2002). Also, criteria for successful validation, verification, or evaluation differ among authors (Marcot et al. 1983). Some authors contend that models can never be validated; they can only be falsified as in the context of a statistical hypothesis (e.g., Holling 1978). Others suggest that validation leads to a binomial outcome (true vs. false or good enough vs. not good enough). Others see validation as an ongoing iterative process of constantly refining a model and documenting its strengths and weaknesses (Starfield and Bleloch 1991). Rykiel (1996) provides the most comprehensive and coherent overview of ecological model validation that we have encountered; it is essential reading for anyone engaged in evaluating ecological models. We utilize his terminology whenever possible.

1. *Evaluation*—a nontechnical umbrella term for an assessment of the strengths, weaknesses, and utility of a model for a stated purpose. Evaluation is a broad concept that takes into account whether the model predictions are good enough (Mankin et al. 1979), and whether it is practical to utilize the model with the resources (people, data, money) available for the problem (Johnson 2001). Johnson (2001) argues that "evaluation" is a more appropriate term than "validation" because value is relative and valid is an absolute term. In the context of predicting vegetation and wildlife suitability, a model might not be valid (i.e., it is incorrect), yet still have value for its intended purpose.

2. *Validation*—involves "demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application" (Rykiel 1996). Validation includes a specific context of applicability (e.g., species, geographic region, associated disturbance processes). In wildlife habitat suitability modeling, we sometimes extend this traditional definition to consider how well the model performs under conditions that are different from those used to develop the model (Conroy and Moore 2002).

3. *Verification*—demonstrating (1) that the ecological processes embodied in the model (e.g., vegetation response to simulated disturbance or predicted habitat quality change in response to vegetation change) are correctly and sufficiently embodied in the model's equations or algorithms; and (2) determining that the computer code used to implement the equations and algorithms is correct (*sensu* Rykiel 1996).

4. *Calibration*—improves the agreement between model output and observed data through estimation and adjustment of model parameters. Calibration data are most often distinguished from validation data by partitioning the data in time or space.

THREE ELEMENTS OF MODEL EVALUATION

Model evaluation can be partitioned into three areas: (1) the application environment, (2) conceptual design, and (3) quantitative performance (Buchman and Shifley 1982) (Fig. 16-2). Each of these areas addresses a potentially large number of questions, issues, and concerns associated with a given model. The numerical order of the three elements is relevant only in that it represents increasing levels of investment of resources (time, data, intellectual energy) to evaluate model performance. Therefore, it is usually most efficient to see if a model passes the tests related to the application environment and design before pursuing to the data-intensive or quantitative performance considerations.

Application Environment

Evaluation of the application environment simply determines whether the model is compatible with one's particular problem and available resources. Landscape-scale simulation models, such as those discussed in this book, are notoriously demanding of spatial data and computing power. A number of questions should be asked at this stage. Do you have (or can you acquire) the required data to apply the model to your region of interest? Do you have the necessary computing resources and the technical expertise to apply the model? Do you have the time it takes to acquire, learn, and apply the model? These and the related questions in Fig. 16-2 are practical considerations. Deficiencies in the application environment can be difficult to overcome and can quickly eliminate an otherwise suitable model from consideration. Alternatives to simulation models such as conceptual models or expert opinions may have serious limitations in the other two elements of model evaluation, but they usually offer a favorable application environment (i.e., they are widely available, they are fast, and they work with the available data).

Every model presents a unique application environment, but we have some practical observations relevant to landscape-scale simulation models.

1. These models require a great deal of data. Data for model initialization or for regional recalibration are inevitably limited or partially missing. Although there are ways to estimate missing data, they are usually time-consuming. Our general rule of thumb is that it will take longer to assemble the data and implement the model than anticipated—probably twice as long and maybe more. Despite the availability of some software specifically developed for large-scale modeling activities (Beck and Suring, this volume; Dijak and Rittenhouse, this volume; He, this volume; Larson et al., this volume), compiling the data necessary for landscape modeling is difficult.

2. Model initialization and calibration are the most time-consuming stages of landscape simulation modeling, especially for real landscapes. Our

Application Environment

User support

Is the system thoroughly and clearly documented?
Are user guides and sample projection runs available?
Is training available?
Who is available to answer questions or address problems?
Who maintains the system?
Is the system open-source with options for user modification?
Is there an established user group?
Does the system provide on-line help and user prompts?

Data

What data are required to use the system?
Are the available data compatible with those requirements?
Are there established methods to estimate missing data?
Does the system check for erroneous data values or ranges?
Is model output in a format suitable for the intended use?
Can output easily be reformatted or further processed?
Can user-supplied data be used to test model performance or recalibrate the model?

Computing

What computing capacity is required?
How quickly does the system produce results?
What specialized skills or knowledge are required of a user?
Can the system be reprogrammed if necessary?
Does the system require additional software for implementation or for analysis of results?
Have the program logic and coding been verified?

Design

Flexibility

Can the system address a wide array of issues or problems?
Can it be readily modified for new problems?
Is the system built of modules that can be revised or replaced as necessary?
Can models for other resources readily be added?
Can a wide array of management practices and other disturbance processes be modeled?
Can the system easily be recalibrated to new conditions or new locations?

Bio-logic

Are projection models formulated to incorporate basic biological and ecological theories of change?
Are appropriate feedback mechanisms and other system controls incorporated?
Do system component interact logically (e.g. vegetation change affects wildlife habitat quality and herbivory affects vegetation change)?
Over long projection periods, do predictions approach reasonable limits?
Are there real or hypothetical conditions that cause the model to predict results that are obviously unreasonable?

Performance

Entire system

How accurately and how precisely can the system forecast the quantities of interest?
  Wildlife habitat amount and suitability
  Wildlife population viability
  Vegetation species composition and size structure
  The spatial arrangement of vegetation and other landscape features
  The spatial and temporal distribution of disturbance events
Do forecasts show a systematic bias or loss of precision associated with length of projections, with certain species of wildlife or certain vegetation types, by ecological land type, or by landscape location?

System components

Is it possible to test individual components of the system?
  Wildlife habitat models
  Wildlife population viability models
  Vegetation change models
  Disturbance models
Do individual components show a systematic bias or loss of precision associated with length of projections, with certain species of wildlife or certain vegetation types, by ecological land type or landscape location, or other relevant variables?

FIG. 16-2

Considerations in model evaluation. The Application Environment issues often determine whether or not a model can be applied in a given situation. Design issues address factors that can make a landscape model adaptable to new places and new problems. Performance issues address the quantitative and qualitative evaluations that are typically associated with model validation and verification. This table is reproduced from Buchman and Shifley (1982) with only minor modifications and reordering criteria. Over the past 25 years, capabilities to access and manipulate data and to forecast landscape change and wildlife response have increased exponentially. However, the basic questions associated with model evaluation have changed little.

experience is that roughly 75% of the effort goes into gathering base data and setting up a large problem, 5% goes into running the actual simulation scenarios, and the remaining 20% goes toward summarizing, interpreting, and reporting results.

3. The large investment typically required to initialize a landscape simulation model and apply it on a realistic landscape is a significant barrier to model implementation. However, the models are usually amenable to addressing a wide range of issues. When the first application is completed, the models can be applied repeatedly to address other issues.
4. Purchasing more computing and data storage capacity is cheap relative to the cost of implementing the model. However, with landscape-scale simulation models, it is surprisingly easy to exceed all the computing and data handling capacity available.
5. We advise automating the computing processes to the greatest extent possible. Repeating a few hundred hours of simulation is easier to tolerate when it is fully automated than when frequent operator intervention is required.
6. Mistakes in modeling will occur. We inevitably rerun the simulation analysis to correct errors—usually many times.
7. Expertise in applying GIS, manipulating large amounts of data, and computer programming is a practical necessity for any large-scale application of vegetation and wildlife models (Roloff et al., this volume).

Conceptual Design

This set of criteria determines the extent to which the model appropriately incorporates the underlying theories of ecology, biology, vegetation dynamics, and wildlife population dynamics. Does the model account for the ecosystem processes necessary to address the issues at hand? How do wildlife respond to landscape features and predicted landscape change? This evaluation requires a careful look inside the model to ensure computations are correct (i.e., examination of equations or algorithms in the model); it also requires a detailed look at the model assumptions in the context of the specific application of the model. For some landscape models this step is difficult because the user interface and software documentation do not provide all the information needed. We encourage users to fully investigate the inner-workings of "black box" models to ensure they are comfortable with the assumptions being made. It is tempting to proceed with modeling without fully understanding the modeling components, only to find out later that critical assumptions are not appropriate for the intended application. Users should not be tempted to select software based solely on ease of use.

Theoretical derivation of processes and interactions affecting landscape change is a useful exercise because it provides a structured way to think about the problem of landscape and wildlife change in response to endogenous and exogenous factors and feedback loops. Proper theoretical formulation of the model is essential, but it is not entirely sufficient. Implementation of an elegant conceptual design can suffer from missing data, ill-defined causal pathways, and processes that are difficult to quantify. A model that is well designed from a theoretical standpoint may be impossible to implement and, thus, useless for practical purposes (Hurley 1986). Evaluation of model design must explicitly consider the qualitative trade-offs between design, data requirements, and utility for the intended purpose. All ecological models compromise theoretical detail and complexity for application utility; the important consideration is whether or not the model remains useful for its intended purpose. Well-designed models find a suitable balance among theoretical and practical considerations. Models often continue to be refined to incorporate more theoretical detail as practical issues like data availability or computing limitations are resolved over time.

Quantitative Performance

A difficult part of evaluating landscape-scale models is the quantitative assessment of the model predictions. In many cases it seems nearly intractable to complete a true validation, particularly when considering the availability of appropriate vegetation and wildlife data. When attempting to validate wildlife habitat suitability maps using field data, we often face issues with nonconstant detection probabilities of surveyed wildlife (McKenzie et al. 2006), seasonal movements and migration, variability in wildlife use among seasons and years, difficulties with sampling small populations (Thompson 2004), and technological limitations. Forest vegetation models, because they deal with stationary populations, avoid some of these issues, but they face other constraints such as slow rates of change over time (e.g., with regard to tree species composition). Despite these barriers to model validation, there are a few techniques, although imperfect, that can address the quantitative performance of landscape-scale decision support models.

Data are invariably a limiting factor in quantitative validation of landscape model predictions. For example, if we want to compare model predictions of change for real landscapes to observed change for those landscapes, where do we get the data? Clearly, we do not have a reference map for future time periods. Therefore, validation can be performed only for past or present conditions. If we are interested in a small landscape, roughly 1,000 ha in size, can we find a dozen landscapes where we have observational data on initial conditions, disturbance events, vegetation change over time, and wildlife change over time? Do we have such results for several decades of observation, or for a century, which is a typical planning horizon for forest landscapes? Usually, we do not. In a few cases, it is possible to test individual species over the long term with appropriate data sets. For example, one might use the breeding bird survey as a long-term data source (Newson et al. 2005, Somershoe et al. 2006, Freeman et al. 2007). One can also substitute space for time using widespread vegetation inventories such as Forest Inventory and Analysis (FIA) data (U.S. Forest Service 2008*a*) although limitations exist (Roloff et al., this volume). Testing with other detailed, species-specific options such as expert panels is also possible (e.g., Holthausen et al. 1994).

There are a number of procedures to quantitatively evaluate model performance. Sensitivity analysis is commonly used to evaluate the contribution of individual parameters to model performance (Rykiel 1996), but does not constitute validation (Johnson 2001). By systematically varying model parameters and comparing results across a wide range of landscape conditions, sensitivity analysis helps model users understand how robust the model is to small changes in the modeled relationships. That helps model users understand how much confidence to place in model results. Because of the interactive effects of multiple input parameters in complex models, it is advisable during sensitivity analysis to alter more than one parameter at a time. As Rykiel (1996) suggests, the most important parameters should be estimated well. Sensitivity analysis, although useful, should be complemented by other quantitative evaluation methods. Statistical procedures, including Receiver Operating Characteristic (ROC) curves (Zweig and Campbell 1993, Pearce and Ferrier 2000), confusion matrix-derived measures, such as the Kappa statistic (Cohen 1960, Fielding and Bell 1997), and cross-validation techniques (Boyce et al. 2002), are commonly used in validation tests. The appropriateness of these procedures is dependent on assumptions of the model, the data available, and intended purpose of the model.

We now turn our focus to a review and discussion of the methods used to evaluate models that make quantitative predictions of changes to vegetation over time and habitat suitability on large landscapes.

EVALUATION OF FOREST VEGETATION MODELS

Types of Forest Vegetation Models

Unlike wildlife, forest vegetation is rooted in place. Except as seeds, trees do not move beyond the reach of their roots or crowns in search of suitable habitat. Thus, observing changes over time for individual trees or communities of forest vegetation is relatively easy when compared to similar operations for wildlife. Consequently, models to predict vegetation change for trees and forest stands are well established (e.g., Ek et al. 1988, Vanclay 1994, Dixon 2003, Husch et al. 2003) and are widely used to predict forest change and develop management prescriptions (Miner et al. 1988, Mowrer 1997, Dixon 1998, Twery et al. 2005).

Methods for validating short-term predictions (e.g., up to 50 years) of tree or stand change are documented in numerous sources (e.g., Buchman and Shifley 1982, Reynolds 1984, Miner et al. 1988, Brand and Holdaway 1989, Vanclay 1994, Vanclay and Skovsgaard 1997). These typically compare model predictions with changes observed from remeasured forest inventory plots. The metrics of interest are usually (1) tree size or growth (e.g., diameter, height, and volume), (2) tree mortality, and (3) stand change per hectare (density, volume, diameter distribution). Validation generally consists of measuring the departure of the observed from predicted vegetation change and reporting bias and precision of estimates by species, per hectare, and over time. Occasionally, validation results for forest change also include estimated prediction intervals—expected errors if the model is applied at new locations (Reynolds 1984). In virtually all situations, model predictions and validation procedures for forests are limited to tree species and exclude other vegetation layers that lack abundant inventory data.

Software for implementing tree- and stand-scale tree growth models is comprehensive and widely available (Ek et al. 1988, Dixon 2003). For some landscapes it is now possible to apply tree- or stand-scale models to all stands across the landscape and, thus, forecast change over time for the entire landscape. This approach provides great detail in predicted forest vegetation structure and composition at all locations on the landscape, it is a highly intuitive methodology, and in some cases integrated software is available to visualize landscape vegetation change over time. Validation for the component vegetation change models are typically reported with the documentation of model calibration procedures (e.g., Brand and Holdaway 1989, U.S. Forest Service 2008*b*), so initial estimates of model precision and bias are available in most cases. Also, long-term predicted changes in stand size structure can be evaluated from a theoretical perspective using multidimensional, graphical analysis of modeled changes in stand density, height, mean tree size, volume, and site quality (Leary 1996). Although the tree- and stand-scale vegetation models usually provide a mechanism to model forest regeneration and species succession, forecasting long-term species succession is not their strength (for the same reasons outlined for other categories of landscape-scale models discussed later in this section).

The Forest Vegetation Simulator (FVS) is the most widely utilized system of tree and stand growth prediction models in the United States. It includes automated procedures for comparing predicted tree and stand changes with actual changes observed from remeasured forest inventory plots. It also provides automated procedures for recalibrating the prediction equations based on changes observed from local inventory data (Dixon 2003). In situations in which the landscape extent and the data on initial landscape vegetation conditions are compatible with model requirements, FVS or similar models can be an excellent choice for forecasting forest vegetation with known (or readily determined) levels of precision and bias. Moreover, it is possible to visualize projected forest landscape change over time as three-dimensional renderings using software such as the Landscape Management System (LMS; Millspaugh et al., this volume; Oliver et al., this volume; University of Washington 2008). A typical application would be for a landscape composed of a large, contiguous public or private ownership supported with a current stand-level inventory of forest conditions for each stand or each ecological land type.

Limitations to applying tree- and stand-level models in the specific context of large-scale landscape decision support modeling are the detailed data requirements for initial landscape conditions (e.g., forest inventory data observed or estimated for each site), the burden of carrying for each site highly detailed data that may be unnecessary for evaluating the final objectives (e.g., for estimating wildlife habitat suitability), and a limited capacity to model natural regeneration and species succession over long time periods. Thus, for landscapes larger than about 25,000 ha and for analyses approaching or exceeding the duration of a forest management rotation, the general tendency has been to model vegetation structure and composition using landscape decision support systems that carry far less detail about forest vegetation at each site. These fall into two broad categories: (1) raster-based models that track forest age class and/or tree species presence on sites ranging from 0.01 ha to 1 km^2 mosaiced across the entire landscape (e.g., LANDIS and HARVEST) (Gustafson and Crow 1999; He et al. 1999, 2005; Mladenoff and He 1999; Mladenoff 2004; Gustafson and Rasmussen 2005; Scheller et al. 2007; He et al. this volume), and (2) polygon-based models that track the progression of each landscape polygon (e.g., forest stand) through a finite number of habitat classes defined by the species composition and size structure of the dominant vegetation (e.g., LANDSUM, SIMPPLE, and VDDT/TELSA) (Chew 1995; Keane et al. 1997, 2002; Beukema and Kurz 1998; Barrett 2001; Chew et al. 2007). In both cases the primary emphasis is on predicting long-term (a century or more) changes in species composition and forest size structure in response to succession, harvest, fire, severe weather, insects, and disease.

Validation and Evaluation Considerations

For all landscape vegetation models, quantitative validation of predicted forest change is hampered by (1) a lack of long-term data documenting patterns of tree species succession and (2) a lack of data describing pre- and post-disturbance vegetation for sites affected by harvest, fire, severe weather, insects, or disease. Consequently, quantitative, data-driven validation estimates of predicted changes in species composition over time (with or without disturbance) are problematic. However, there are qualitative procedures for evaluating predicted species dynamics for landscape-scale models of vegetation change. Verification of the predicted rate and pattern of disturbance processes is also required.

Evaluation of landscape-scale forest vegetation models can be subdivided into three parts: (1) validation of forest structure change in the absence of disturbance; (2) validation of the rate of exogenous disturbance; and (3) evaluation of species dynamics as affected by disturbance type (e.g., harvest, fire, wind, none) and by ecological land type. The three elements do not exhaust the range of what could be (and ideally should be) done to evaluate a landscape-scale model of forest vegetation used to support decision making, but they provide an essential first iteration.

Validation of Forest Structure Change.—Validation of forest structure change in the absence of disturbance has strong ties to validation processes developed for traditional forest growth and yield models, and it is the validation component that typically has the best supporting data. In the absence of disturbance, established forest stands (or sites) go through predictable stages of development that can be described in terms of tree size and density, in terms of stand age, or in terms of structural stages (e.g., stand establishment stage, stem exclusion stage, understory reinitiation stage, old-growth stage; Oliver and Larson 1996).

For tree- and stand-based models of forest vegetation change (e.g., FVS) validation estimates typically encompass two to five decades of observed change for disturbed and undisturbed forests. For robust evaluation of model performance, the observed forest changes in the independent validation data set should cover a wide range of forest age, species composition, site quality, and ecological land types; these readily measured attributes are known to influence forest change. Examples and recommendations for validation of forest size and structure change based on inventory data can be found in Reynolds (1984), Brand and Holdaway (1989), Vanclay (1994), and Vanclay and Skovsgaard (1997). However, other types of landscape-scale vegetation change models track less detail about each site and produce outputs that are not directly comparable with remeasured forest inventory data. These models include LANDIS (He et al. 1999, 2005; Mladenoff and He 1999; Mladenoff 2004; Scheller et al. 2007; He et al., this volume), HARVEST (Gustafson and Crow 1999, Gustafson and Rasmussen 2005), LANDSUM (Keane et al. 1997, 2002), SIMPPLE (Chew 1995, Chew et al. 2007) and VDDT/TELSA (Beukema and Kurz 1998). Evaluation of vegetation age and size structure change for these models amounts to verification that the predicted forest age changes with each time step in the simulation and/or that the predicted vegetation structural state changes with the model time step in accordance with the probabilities established when the model was calibrated.

Evaluation of Modeled Disturbances.—Evaluation of modeled exogenous disturbances by harvest, weather, fire, insects, or disease in landscape-scale vegetation models is primarily a verification process. For example, harvest is a disturbance that in reality and within a landscape model is controlled by managers. Model verification should be conducted to determine (1) that the timing, location, and cumulative spatial patterning of modeled harvest operations is consistent with those prescribed in the simulation and (2) species composition and age structure of the modeled landscape are adjusted to reflect the anticipated post-harvest state of the forest site. This is simply verification that modeled harvest events operate as intended. Prescribed fire can be thought of as a similar anthropogenic disturbance that requires the same type of model verification.

Modeled patterns of disturbance by weather, wildfire, insects, or disease are usually developed either by analyzing and quantifying observed patterns based on past disturbances (e.g., records of wildfires or insect damage) or by speculation about future patterns of disturbance. In either case, model verification should be conducted to ensure that the predicted patterns and extent of disturbances implemented in the landscape model are consistent with the observed historical data and/or with speculative scenarios. Landscape models often use stochastic methods to predict the timing and location of disturbances due to weather, wildfire, insects, or disease. Thus, model verification should include examination of variability in the location and timing of disturbance events across multiple runs of a given scenario.

Although opportunities for quantitative validation of predictions of disturbance events in landscape-scale vegetation models are limited, there is a recent example. Yang et al. (2007) used historical data on the location of fire ignitions in the Missouri Ozarks from 1970 to 2002 to develop a probability of map of wildfire ignition risk for a 130,000 ha study area. The wildfires in that region are primarily human-caused, and topography, road locations, land ownership, and proximity to communities were shown to be significant predictors of ignition risk. They then used Monte Carlo methods to repeatedly simulate fire spread across the landscape based on the ignition probability and modeled rates of fire spread in response to vegetation type, topography, prevailing wind direction, and locations of fire breaks. The resulting burn probability map summarized the likelihood that any given point on the landscape would burn. This provided for each site a spatially explicit fire probability that is applicable for landscape change simulations. Quantitative validation of the fire risk model was possible by examining the mapped patterns of actual fires in subsequent years (2003–2004) and using categorical analysis to determine if the sites that really burned occurred disproportionately at locations that had a greater than average predicted probability of burning (Yang et al. 2008).

Evaluation of Changes in Species Composition.—Predicting changes in species composition is the most difficult aspect of modeling landscape-scale, long-term vegetation change and of validating those models. In most forest ecosystems the species composition of regeneration varies greatly from site to site (even within a single stand) and is known to be affected by the species composition of the prior stand; the size and species composition of the advance reproduction; the site productivity; the ecological land type; the type, size, frequency, and intensity of the disturbance event(s) that precipitated the regeneration event; and the type, size, frequency, and intensity of disturbances that follow the regeneration event (e.g., drought or surface fire). After a forest is regenerated, tree species composition can continue to change rapidly as the forest progresses through the stand establishment and stem exclusion stages of development; this may last for three decades or more (Oliver and Larson 1996, Johnson et al. 2002). Due to a lack of long-term, site-specific inventory data describing patterns of forest regeneration, a quantitative comparison of observed and modeled long-term changes in tree (or herbaceous) species composition is rarely possible. Short-term comparisons of observed and modeled changes in species composition (e.g., for two to five decades) are possible but rarely sufficient because they usually do not adequately capture successional shifts in species composition (e.g., the gradual shift from oak [*Quercus* spp.] dominated forests to forests dominated by maples [*Acer* spp.] and other mesic species across much of the eastern United States).

Often the most useful verification of predicted long-term changes in tree species composition is the evaluation of "reasonableness" as judged by local experts. In this process a landscape model of vegetation change is applied to a real landscape or a theoretical test landscape while implementing a wide range of disturbance scenarios. Then the predicted changes in species composition over time are vetted by experts who judge the patterns of species change over time to be reasonable or otherwise. Because of the large number of variables involved in landscape modeling, this procedure is facilitated by including modeled scenarios that implement a single disturbance agent while holding other factors constant (e.g., separate scenarios for no disturbance, low intensity fire, high intensity fire, even-aged management, and uneven-aged management). This is a form of sensitivity analysis. Depending on the number of modeled tree species and the complexity of the modeled ecosystems, experts may need to focus on a few key indicator species or scenarios.

An example is the evaluation of tree species composition modeled using LANDIS in southern Indiana. In that region oaks dominate forest overstories on the majority of acres (Woodall et al. 2005). However, five decades of inventory data show that under the current management regime (characterized by low fire frequency and small harvest openings) maples and other mesic and shade-tolerant tree species are increasing in abundance and gradually displacing oaks via greater regeneration success (Shifley and Woodall 2007). Silvics information (Burns and Honkala 1990) and stand-scale regeneration studies indicate that (1) red oak species (*Quercus* section *Lobatae*) are faster growing and shorter-lived than the white oaks (*Quercus* section *Quercus*); (2) oaks are generally more fire tolerant than their mesic competitors, particularly with repeated fires; (3) white oaks tend to be more tolerant of fire than red oaks; and (4) harvesting provides opportunities for some oaks to successfully regenerate via sprouting (Johnson et al. 2002). Three management scenarios applied to the 81,000 ha Hoosier National Forest in southern Indiana provide a qualitative means to examine and evaluate the reasonableness of predicted long-term changes in species composition (Rittenhouse 2008) (Fig. 16-3).

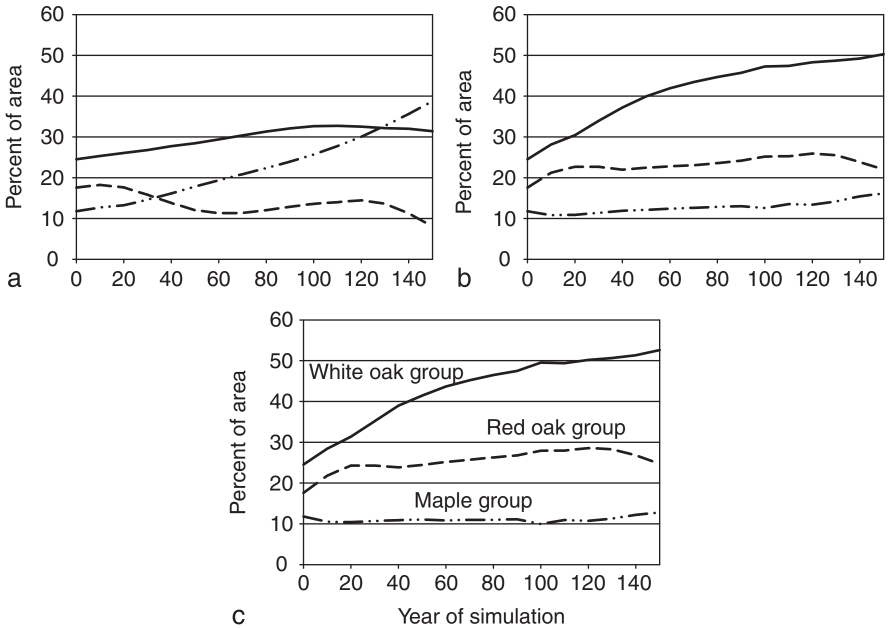

FIG. 16-3

Projected change in dominant tree species composition on each site for three management alternatives on the 81,000 ha Hoosier National Forest. This figure shows projected trends over 150 years for three major species groups: red oaks (*Quercus rubra* L., *Q. Velutina* Lam., *Q. coccinea* Muenchh.), white oaks (*Q. alba*, L., *Q. Prinus* L.), and maple (*Acer saccharum* Marsh., *A. rubrum* L.). With the exception of hickories (*Carya* spp.), the nine remaining species groups were typically dominant on less than 5% of the landscape. Panels ordered in increasing intensity of disturbance: (A) no harvest or prescribed fire, (B) group selection harvest on 3.9% of area per decade and prescribed fire on 3.6% of area per decade, (C) shelterwood or clearcut harvest on 4.3% of area per decade and prescribed fire on 36.7% of area per decade (From Rittenhouse 2008).

EVALUATION OF WILDLIFE HABITAT MODELS

An important consideration in any evaluation of wildlife habitat models is the currency (or metric) used to represent habitat quality because this currency will help decide which type of data to collect or which existing data may be used to validate the model. The most basic measure of habitat suitability is species presence or absence. More complex measures of habitat suitability include species abundance or density (Niemuth et al., this volume), the amount of time spent within different habitat types, vital rates (survival or fecundity), and population viability (Beissinger et al., this volume). Careful consideration of what "suitability" represents is essential in model building, but equally necessary when evaluating model performance. The type of empirical data used for validation should correspond to the objective of the suitability model. For example, if the model is developed to predict breeding habitat suitability, then empirical data used for validation should provide evidence of breeding and ideally some measure of breeding success (e.g., nest success, number of young fledged, season-long or annual fecundity).

Wildlife habitat suitability models are either deductive or inductive (Ottaviani et al. 2004), and the difference between the two is important because some validation approaches are appropriate for only one type. Deductive models include those relying on expert knowledge and published literature (e.g., Dijak and Rittenhouse, this volume). In contrast, inductive models are statistical models developed from wildlife data that are associated with habitat features (e.g., Niemuth et al., this volume). The latter are commonly developed from survey data or radiotelemetry (e.g., resource selection functions; Manly et al. 2002). Validation of deductive models generally involves the use of independent field data, such as animal location, density, or other demographic components (e.g., nest success) to assess the utility of the model. Validation then consists of comparing expected (or predicted) values to observed (or reference) values and quantifying the agreement between the two as a form of accuracy assessment (Rykiel 1996). This comparison is made in different ways depending on several factors, including the type of model and data used in the validation procedure, and whether or not the model output and reference maps are spatially explicit. Validation of inductive models also often uses independent data (e.g., Mladenoff et al. 1999, Luck 2002), but frequently model evaluation includes traditional statistical evaluations, such as resampling (e.g., jackknife and bootstrapping) and data-partitioning (e.g., *k*-fold cross-validation) procedures (Boyce et al. 2002) on the original observations. We view the use of resampling and data-partitioning tools as useful first approximations of model evaluation; however, the use of independent data allows for a richer investigation of model performance and utility issues (e.g., Is the model over-fit? Is performance of the model overrated?).

The most common type of data used to validate wildlife habitat suitability models is presence-absence data. Presence-absence models use data from known occurrences or observations (site is used) and known absences (site is unused). Sites classified as unused must be confirmed via sampling. Otherwise, used locations represent a sample from a distribution of available locations on the landscape. In this sense, presence-only models are a subset of presence-available models because the data contain known occurrences; absence is neither confirmed nor addressed in the model. Several papers have summarized validation procedures for presence-absence models (Fielding and Bell 1997, Boyce et al. 2002, Hirzel et al. 2006) including Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) (Zweig and Campbell 1993, Cumming 2000, Pearce and Ferrier 2000), and confusion matrix-derived measures (Baldessarini et al. 1983) such as the Kappa statistic (Cohen 1960, Fielding and Bell 1997). For presence-available models, Boyce et al. (2002) recommended a *k*-fold cross-validation method and used Spearman-rank correlation to compare observed values (presence only) to categories (bins) of predicted values. The Boyce et al. (2002) approach is also applicable to presence-absence models. All the statistical approaches for validation of presence-absence and presence-available models are applicable to suitability models. When presence and absence data are available, one also has the opportunity to evaluate both omission error (species present, yet model predicts unsuitable habitat) and commission error (species absent, yet model predicts suitable habitat; Ottaviani et al. 2004). We believe that measures such as ROC curves (Fielding and Bell 1997, Pearce and Ferrier 2000) are useful, in part, because they overcome issues with choosing an arbitrary probability threshold when evaluating model performance (although see Termansen et al. 2006).

Roloff and Kernohan (1999) reviewed validation studies for published Habitat Suitability Index (HSI) models and provided a "Checklist for study design and validation considerations for evaluating HSI model performance." They found that the most common deficiencies in HSI validation studies were (1) inadequate consideration of input data variability and how that variability affected interpretation of the final HSI output (Bender et al. 1996); (2) application of the models to inappropriate spatial scales; (3) sampling too narrow a range of habitat conditions; and (4) collection of population data over too short a time frame to reflect variation in population size, density, or demographic rates. Subsequent studies have addressed some of the issues summarized by Roloff and Kernohan (1999). For example, Burgman et al. (2001) used a fuzzy numbers approach for establishing reliability estimates (i.e., confidence intervals) of HSIs. Larson et al. (2004) used a similar approach to create upper and lower limits of ovenbird (*Seiurus aurocapillus*) habitat suitability.

Although many validation procedures exist, relatively few specifically address the use of animal data to assess a suitability model's predictive ability. In the next two sections we highlight general issues and practical approaches for validation of both deductive and inductive wildlife habitat suitability models. Our purpose is to address issues of validation with the explicit goal of quantifying the predictive ability of habitat models using animal data. We focus our discussion around HSI models and not animal counts or viability per se. However, many of the general concepts apply to either objective. Often, we are ultimately interested in the same diagnostics, such as classification rates and the difference between predicted and observed values within the landscape.

General Issues with Validation of Habitat Suitability Models

Here we discuss four specific issues related to validation of habitat suitability models: (1) model uncertainty (also see Millspaugh et al., this volume), (2) types of empirical data used in validation, (3) multiscale validation, and (4) autocorrelation (both spatial and temporal). The increased use of habitat suitability models within large landscapes for policy, conservation, and management decisions (e.g., state and federal land planning) makes a discussion of these issues relevant. Although this is not a complete list of HSI model validation issues (e.g., no discussion of nonconstant detection probabilities [MacKenzie et al. 2006]), it provides an essential starting point that can be expanded with the guidance of other literature that addresses these issues in part or on whole (Scott et al. 2002, Williams et al. 2002, MacKenzie et al. 2006).

Model Uncertainty.—Habitat suitability models include numerous sources of model uncertainty. Regan et al. (2002) defined two main categories of uncertainty: linguistic uncertainty and epistemic uncertainty. Linguistic uncertainty arises from the vagueness, ambiguity, and context dependence of the natural and scientific language used in developing, describing, and applying models. Epistemic uncertainty arises from uncertainty regarding a determinate fact. For example, habitat suitability models often include numerous habitat variables, and each variable is subject to location or assignment error in GIS layers, bias due to study design or data collection methods, and natural variation in biotic or environmental conditions. Rigorous attention to study design and error assessments should be conducted prior to model development.

Model uncertainty also arises from subjective interpretation of wildlife-habitat relationships and the decisions made in developing models (Ray and Burgman 2006). For example, deductive habitat suitability models (e.g., traditional HSI models) use expert knowledge to specify which habitat variables are important to a species, the relative importance of different variables, and how to relate those variables to habitat suitability values (Larson et al. 2003; Dijak and Rittenhouse, this volume). Given that the assumed or known form of the suitability relationship may be logical (presence-absence of feature), linear, or nonlinear (e.g., sigmoid or threshold response), and the HSI equation may use geometric, arithmetic, or logical relationships to calculate the composite HSI value, many different HSI models may be developed from the same set of variables. Ray and Burgman (2006) recommend using “bounded” habitat suitability maps (i.e., maps developed under alternative HSI models) to evaluate subjective uncertainty. This is a variation of sensitivity analysis.

Alternatively, model verification and calibration, conducted prior to validation, may be used to ensure that the model form is correct and all relevant parameters are included. For example, Mitchell et al. (2002) determined the relative contribution of each habitat suitability component (e.g., input data as well as individual suitability indices) to the final HSI model using sensitivity and elasticity analyses (Caswell 1978, Stearns 1992). He quantified how the variability in each component could affect the final HSI values by weighting each individual suitability parameter in the HSI equation. Unequal sensitivity and elasticity values among components suggest variation in model output is due to a subset of model components. In that case, the model may be reduced to a subset of components without substantially affecting model predictions. Beven (1993), Oreskes et al. (1994), Rykiel (1996), Beutel et al. (1999), Morrison et al. (1998), Johnson (2001), and Scott et al. (2002) provide methodological summaries or reviews that address model uncertainty, verification, and calibration.

Data for Model Evaluation.—The second main issue relates to the data type used in model validation. An important consideration in model evaluation is the currency (metric) used to represent habitat quality because this currency will help decide which type of data to collect or which existing data may be used to validate the model. Buskirk and Millspaugh (2006) discussed alternative currencies used in resource selection studies; these are defined as measures of investments by an animal in searching, finding, or using a resource. The currencies include time spent, distance traveled, energy expended, and predation risks incurred. In addition, they defined event sites as those places that animals use intensively or where important life functions occur. Each of these currencies is applicable to habitat model validation using animal data, and they complement more traditional measures of habitat quality.

More complex measures of habitat suitability include species abundance or density, vital rates (survival or fecundity), and population viability (Akçakaya and Brook, this volume; Beissinger et al., this volume; Larson et al., this volume). Van Horne (1983) suggested that habitat quality should be the product of density, survival, and expectation of future offspring (also see Johnson 2007). The data needed to meet this definition of habitat quality are often prohibitive. However, these metrics are the most meaningful and useful, given issues with animal location data (Battin 2004, Rittenhouse 2008) (Fig. 16-4). These metrics must be used when demographics are being modeled. Despite the desirability of demographic data, they can be problematic for validating habitat models and, for this reason, are often not used. For example, other factors, such as territoriality, may limit access to high-quality habitat, resulting in some individuals using low-quality habitat. The existence of demographic phenomena affecting the relationship between populations and habitat (e.g., source-sink) also should be considered to ensure inference about the predictive capability of the model is correct (Conroy and Moore 2002). Thus, it is critical that context be considered. For example, a species below carrying capacity might respond differently to available habitats. In such cases, habitat suitability might be better defined as some demographic parameter.

Spatial Scale and Validation.—The third issue with validation of wildlife suitability models is that validation should address the spatial scale of model development and of the intended application. The level of accuracy or precision should be specified *a priori* because agreement depends in part on the cell size (i.e., landscape resolution; Pontius et al. 2004). In general, maps with coarse resolution have better agreement with observed data than maps with fine resolution because the coarse resolution data are an average or a composite across a relatively large area. This reduces the variation in the observed values. In other words, the focus shifts from spatially explicit prediction at fine resolutions to composition at coarser resolutions. Advancements in computing technology

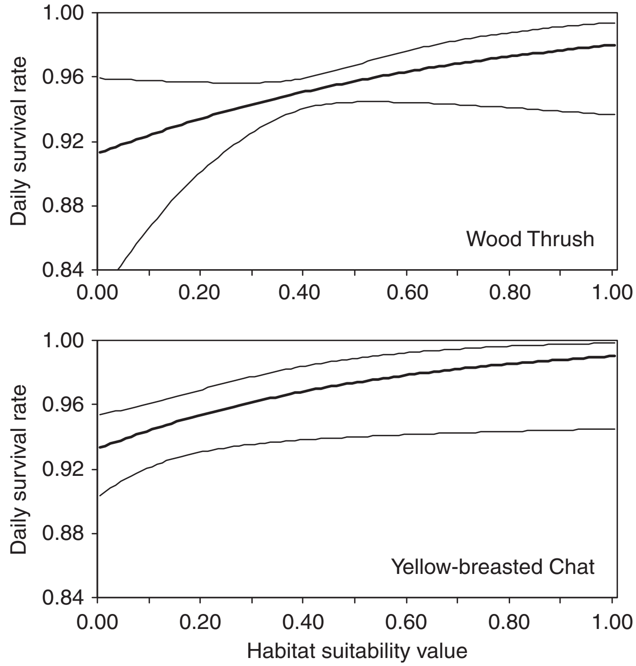

FIG. 16-4

Daily survival rates of wood thrush and yellow-breasted chats in relation to habitat suitability values in Missouri Ozark Forest Ecosystem Project, south-central Missouri, 1991–2002. Thin lines represent Wald 95% confidence limits for the logistic-exposure model (Rittenhouse 2008).

facilitate application of wildlife suitability models to large spatial extents at high resolutions (e.g., small cell sizes), which means that HSI models incorporate multiple spatial scales. Empirical data and the suitability measures developed from them may represent habitat associations at similar or different spatial scales than intended by the model. The most common spatial scale used in wildlife suitability models is the animal's home range size, which corresponds to Johnson's (1980) second-order selection. However, habitat variables at any scale may influence selection and affect the suitability relationship. The correlation of habitat variables within and among scales results in cross-scale correlation; Battin and Lawler (2006) present methods for identifying and incorporating cross-scale correlation in analyses.

Spatial and Temporal Autocorrelation.—The fourth issue with habitat suitability model validation is treatment of spatial and temporal autocorrelation. Spatial autocorrelation results when the distribution of animals or environmental variables in a cell is not independent of surrounding cells (e.g., attraction of con-specifics, dependence in the suitability value in neighboring cells). Spatial autocorrelation may occur in wildlife models with suitability relationships based on patch size (area) functions or the composition of habitat computed within a moving window, because the suitability value is determined in part from a cell's proximity to other cells with similar resource attributes. Spatial autocorrelation may also occur in the animal data used for validation. For example, kernel methods for estimating the distribution and intensity of use by animals inherently are spatially autocorrelated (Marzluff et al. 2004). The concern with spatial autocorrelation is that it may contribute to either stronger agreement or disagreement when comparing the observed map to the predicted map. Without explicitly addressing spatial autocorrelation, wildlife suitability models may chronically under- or over-perform in certain regions of the map. Validation techniques that retain the spatial context of prediction errors (e.g., as maps of prediction errors) rather than providing a global measure of error may facilitate identification of spatial autocorrelation.

Validation of Suitability Models Using Animal Location Data

In this section we focus on approaches that use animal data to validate habitat suitability models, which predict habitat quality and not demographic parameters. These approaches address some of the validation issues discussed in the preceding section, but due to limitations in available animal data, no single approach simultaneously addresses all the issues (Van Horne 1983, Battin 2004). The choice of validation procedure varies by the form of the habitat suitability model and the available animal data, and it should be matched to the intended model use. We primarily discuss procedures that compare the predicted distribution of suitability values to observed locations of individual animals or of distributions (populations) of animals, which may be estimated from radio-telemetry relocations, surveys (e.g., point counts, distance sampling), or territory spot mapping. The features of distributions that are of interest when validating suitability models include the frequency of values with the same classification, the magnitude of the difference between predicted and observed values for a given cell, the cell location, the type of error (agreement or disagreement), and the source of error (e.g., issues of model uncertainty, animal data, spatial scale, autocorrelation).

The first validation approach uses compositional analysis (Aitchinson 1986). Compositional analysis has been used in studies of habitat selection to determine if the proportion of each habitat type within the home range (area used) differs from the proportional occurrence of habitat types at a larger scale (area available—typically the entire study area; Aebischer et al. 1993). A logical extension of compositional analysis for validation is to consider the predicted map as defining "available" habitat and testing whether the proportion of suitability index values within an observed animal's home range (i.e., "used") differ from what is available. This requires categorizing (or binning) suitability values and treating the categorical suitability data as synonymous with a categorical map of habitat types. In addition, the working assumption is that animals should use areas with higher habitat suitability values than is the mean of all available habitat. For example, Ottaviani et al. (2004) validated habitat suitability models for 113 species of terrestrial vertebrates and 82 species of birds using compositional analysis. They compared the mean, covariance structure, and proportion of habitat suitability classes within polygons of used sites to polygons of similar size selected at random from the study area. Differences in modeled habitat suitability class rankings from those for sites that were actually used may reveal which landscape conditions are contributing to agreement or lack thereof. Ottaviani et al. (2004) assumed animal use within a polygon was equal to the raw proportion of habitat suitability classes within the polygon, which means that nonrandom use of habitat suitability classes within polygons was not considered. Millspaugh et al. (2006) suggested weighting the raw proportion of habitat (suitability classes) by the amount of animal use, estimated from empirical data. In this way, both the proportion of each habitat suitability class and the amount of use of that class are considered in the analysis (Millspaugh et al. 2006).

While compositional analysis can indicate that differences exist between observed and predicted values, it does not necessarily provide a spatial context regarding those differences. In other words, do the maps look similar? Pontius et al. (2004) used multiple resolution analysis to answer this question. They partitioned sources of error between observed and predicted values based on cell location and on the quantity of cells with particular suitability values. The Pontius et al. (2004) approach provides a good basis for understanding model performance in terms of the quantity and location of suitability values, and the influence of spatial resolution on model prediction ability. Couto (2003) presented approaches for comparing maps with different classification schemes, which might occur when comparing habitat suitability maps to animal distributions. Cell values in "hard maps" are discrete and mutually exclusive, meaning that a cell has only one value. In contrast, cells in "fuzzy maps" have mixed or uncertain membership, meaning that a cell could have more than one value. Uncertainty in cell membership could arise from model uncertainty, autocorrelation, or differences in cell resolution between predicted and observed maps (i.e., spatial scale). Couto (2003) described measurements based on fuzzy set theory that assessed the frequency, magnitude, source, and nature of errors, for hard maps and fuzzy reference data.

Roloff et al. (2001) validated a spatially explicit habitat effectiveness model for elk (*Cervus elaphus*) in South Dakota using telemetry data. The model scored suitability based on forage quality, quantity, and availability (Fig. 16-5). Using the volume of intersection index (Millspaugh et al. 2004), which compares the surface fit of two utilization distributions (i.e., three-dimensional surfaces), Roloff et al. (2001) compared the suitability model output with the total distribution of elk movements (Fig. 16-6). This global measure of agreement between two utilization distributions offers an overall assessment of fit, but follow-up analyses are required to determine where the model is predicting poorly. We view this approach as appropriate when comparing the global fit of two distributions and when there is interest in determining the relative fit (e.g., does the model perform better or worse during some seasons than others, or does the model perform better for one group of animals versus another group?).

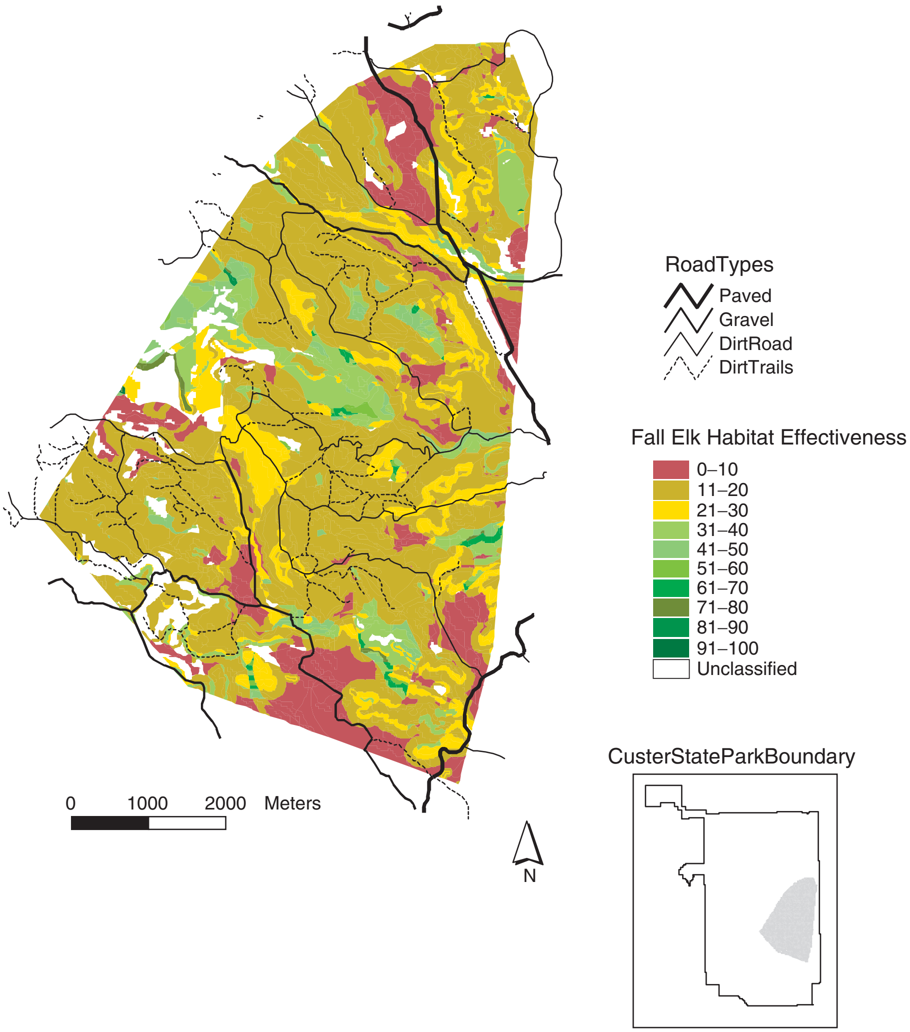

FIG. 16-5

Fall habitat quality scores for an elk subherd in Custer State Park, South Dakota, USA (from Roloff et al. 2001).

As with vegetation models, it is useful to simply summarize basic information (e.g., frequency of correct observations) about model performance using an independent data set. Fielding and Bell (1997) proposed several useful calculations to assess model performance when comparing animal observation data with suitability models. First, they constructed a table that identifies the following: number of sites where the species was predicted to occur and it did; number of sites where the species was predicted to occur and it did not; the number of sites where the

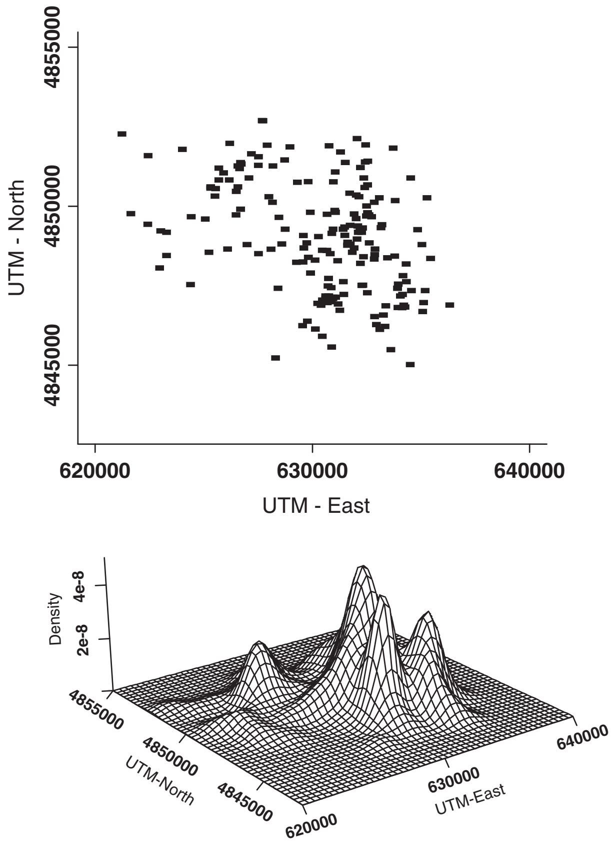

FIG. 16-6

A sample utilization distribution. The upper panel shows the spatial distribution of radiotelemetry observations. The lower panel shows a three-dimensional utilization distribution where the surface height at location x,y represents the intensity of animal use at that point relative to all other locaions (From Millspaugh et al. 2000).

species was predicted to be absent and it was; and finally, the number of sites where the species was predicted to be absent and it was not absent (Table 16-1). From that information, Fielding and Bell (1997) recommended the calculation of several metrics useful for determining model performance (Table 16-2). Simple metrics such as these greatly aid model evaluation and can be used with other measures. For example, an ROC curve (Fielding and Bell 1997, Pearce and Ferrier 2000) is created by plotting sensitivity against 1 – specificity (Table 16-2) across threshold values. The resulting curve gives a measure of model performance (Fielding and Bell 1997, Pearce and Ferrier 2000). We advocate the use of ROC curves; however, it should be noted that the use of ROC curves has recently been criticized in validating species occurrence data (Termansen et al. 2006).

Table 16-1 A Classification Table Used to Summarize the Number of Sites with Observed and Predicted Occurrences of a Species (After Fielding and Bell 1997, Luck 2002)

	Predicted Occurrence	
Observed Occurrence	**Present**	**Absent**
Present	a	c
Absent	b	d

a = number of sites where a species was predicted and observed; b = number of sites where a species was predicted to be present but was not observed; c = number of sites where a species was not predicted to occur yet was observed; and d = the number of sites where the species was predicted to be absent and it was not observed.

Table 16-2 Diagnostic Metrics for Evaluating Model Performance (After Fielding and Bell 1997, Luck 2002) n = the number of cites

Metric	**Calculation**
Correct classification rate[a]	$(a+d)/n$
Kappa[b]	$[a+d-((a+c)(a+b)+(b+d)(c+d))/n]/$ $[n-((a+c)(a+b)+(b+d)(c+d))/n]$
Negative predictive power[c]	$d/(c+d)$
Positive predictive power[d]	$a/(a+b)$
Prevalence[e]	$(a+c)/n$
Sensitivity[f]	$a/(a+c)$
Specificity[g]	$(d/(b+d)$

Variables a, b, c, and d are as defined in Table 16-1.
[a]measure of the number of sites correctly classified.
[b]measure of the improvement to classification over a null (chance) model. Values <0.4 are poor; 0.4–0.75 are good; and >0.75 is excellent (Landis and Koch 1977).
[c]proportion of sites where the species was predicted to be absent yet was present.
[d]proportion of sites where occurrence was predicted to occur and the species did occur.
[e]proportion of prevalence cases.
[f]the true positive rate.
[g]1 – specificity = false positive rate.

FUTURE DIRECTIONS

The wildlife suitability models discussed in this book are intended to be used for planning over large geographic areas. Such broad applications are useful for addressing complex natural resource issues. Landscape-scale decision support models allow us to investigate these large-scale issues in cases where the collection of field data is logistically and economically prohibitive. The large spatial scales addressed by landscape models are beneficial to planning efforts, but it makes the models hard to validate. Collection of independent data at the necessary temporal and spatial scales is costly and difficult. However, when viewed in an adaptive management framework, continued model refinement and evaluation becomes tractable, reduces uncertainty, and facilitates resource management.

We have discussed models for wildlife habitat in the context of forest vegetation management. Over the next century, changes in land use will also greatly affect habitat suitability for many species at many locations. Increases in the area of land devoted to primary homes, second homes, businesses, and transportation can greatly affect habitat suitability as can rural land use shifts into and out of agricultural production. Efforts are underway to link landscape-scale models of vegetation and wildlife to land use change (e.g., Syphard et al. 2007). Effects of global climate change on forest vegetation and avian species are also under investigation (Iverson et al. 2005, Matthews et al. 2007, Prasad et al. 2007). These factors, and probably other macro effects, will be gradually incorporated into future landscape-scale habitat modeling. That will provide new analysis opportunities and compound difficulties associated with model validation.

Management prescriptions at any given time are made using the best available science and the best available data. However, implementation and monitoring of management prescriptions provides new opportunities to learn more about the system, and prescriptions can change over time as better information becomes available (Millspaugh et al., this volume). Models then serve the dual purposes of (1) quantifying what we know (or think we know) about the system and (2) providing a framework for evaluating key uncertainties in our understanding of the system. This role for models is not new (Williams et al. 2002), but is sorely needed in the future. The value of continuous landscape-scale monitoring, within an adaptive management framework of model performance, cannot be overemphasized (Millspaugh et al., this volume).

Experimental approaches offer another solution to validate models and test our understanding of system processes. In the case of wildlife habitat suitability models, experimental manipulation of vegetation conditions offers a useful opportunity to test the strength of habitat suitability relationships. Projects such as the Missouri Forest Ecosystem Project (MOFEP) are unique in their temporal and spatial scale of experimentation (Brookshire and Shifley 1997, Shifley and Brookshire 2000, Shifley and Kabrick 2002). MOFEP is a large-scale, long-term experiment designed to determine the effects of even-aged, uneven-aged, and no-harvest forest management on biotic and abiotic ecosystem attributes at the landscape scale (Brookshire and Shifley 1997). There are nine sites (three replicates of each treatment) ranging from 312 to 514 ha in size. These are large enough to examine landscape-scale effects of vegetation treatments on wildlife species of management concern. The role of experimentation, even at smaller temporal and spatial scales, would be useful in evaluating suitability relationships.

Much is written and discussed in this book about multispecies models (see Flather et al., this volume; Noon et al., this volume). While we agree with the potential value in multispecies models, validation of these models will present new difficulties. Although it might be possible to survey multiple species at the same time (e.g., birds), there are limitations to multispecies surveys. Further, it will be necessary to develop techniques that are suitable for multiple species validation, which are currently lacking in the literature. However, given the increased emphasis on multispecies assessments, we encourage investigation of appropriate validation procedures.

SUMMARY

Models that predict vegetation and wildlife dynamics at large spatial scales, like all models, are abstractions of reality. Under the best circumstances, models capture the most important features and processes of the real system. However, models are imperfect representations of reality, and every model is wrong. Validation is a critical, yet often neglected component of the modeling process. The key to evaluating a model is determining whether a model is useful for its intended purpose. Validation of wildlife models is universally hampered by the type, quantity, and spatial extent of observations of habitat use that can be quantitatively and qualitatively compared with model predictions. Our ability to conceive and design validation procedures far outstrips our ability to implement those procedures based on field data of habitat use and population dynamics. Fortunately, the model validation process helps resolve this deficiency by identifying model shortcomings, data shortcomings, and opportunities for field research or monitoring that can efficiently address those shortcomings. The essence of model validation and verification is understanding and articulating the strengths, weaknesses, and utility of a model for its intended purpose and relative to alternative methodologies. The principal stages of landscape model validation include (1) vetting the explicit and implicit assumptions of the landscape model; (2) verifying the computer code that implements the model, applies the mathematical equations or algorithms, and handles the bookkeeping; and (3) validating model predictions using real (and preferably independent) data that are invariably limited in scope and spatial extent. For models of sufficient utility to be maintained and improved, these three stages are repeated at periodic intervals, typically in an adaptive management framework. In addition, model evaluation must consider pragmatic aspects such as ease of use, data requirements, the user interface, computational demands, and ease of modification or adaptation. The models presented in this book span a range of intended uses, spatial scales and points of focus. There are vegetation models that predict forest vegetation dynamics for trees, for stands, or for large landscapes. Similarly, there are wildlife models that estimate generic habitat quality and those that predict species population dynamics over time. This chapter

addresses universal considerations for evaluation of such models and then presents examples of alternative methodologies from the literature. Despite considerable previous research on model validation, the quantitative validation of landscape model predictions still presents significant practical and technical challenges.

ACKNOWLEDGMENTS

We thank the U.S. Forest Service and the University of Missouri for their support of our research.

LITERATURE CITED

Aebischer, N. J., P. A. Roberston, and R. E. Kenward. 1993. Compositional analysis of habitat use from animal radio-tracking data. *Ecology* 74:1313–1325.

Aitchinson, J. 1986. *The statistical analysis of compositional data*. Chapman and Hall, London, UK.

Akçakaya, H. R. 2001. Linking population-level risk assessment with landscape and habitat models. *Science of the Total Environment* 274:283–291.

Akçakaya, H. R. 2002. *RAMAS GIS: Linking landscape data with population viability analysis (version 4.0t)*. Applied Biomathematics, Setauket, New York, New York, USA.

Akçakaya, H. R., J. Franklin, A. D. Syphard, and J. R. Stephenson. 2005. Viability of Bell's sage sparrow (*Amphispiza belli* ssp. *belli*): Altered fire regimes. *Ecological Applications* 15:521–531.

Akçakaya, H. R., V. C. Radeloff, D. J. Mladenoff, and H. S. He. 2004. Integrating landscape and metapopulation modeling approaches: Viability of sharp-tailed grouse in a dynamic landscape. *Conservation Biology* 19:526–537.

Baldessarini, R. J., S. Finklestein, and G. W. Arana. 1983. The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry* 40:569–573.

Barrett, T. M. 2001. *Models of vegetation change for landscape planning: A comparison of FETM, LANDSUM, SIMPPLLE, and VDDT.* U.S. Forest Service, Rocky Mountain Forest and Range Experiment Station, General Technical Report RMRS-GTR-76-WWW, Ogden, Utah, USA.

Battin, J. 2004. When good animals love bad habitats: Ecological traps and the conservation of animal populations. *Conservation Biology* 18:1482–1491.

Battin, J., and J. J. Lawler. 2006. Cross-scale correlations and the design and analysis of avian habitat selection studies. *Condor* 108:59–70.

Bender, L. C., G. J. Roloff, and J. B. Haufler. 1996. Evaluating confidence intervals for habitat suitability models. *Wildlife Society Bulletin* 24:347–352.

Beukema, S. J., and W. A. Kurz. 1998. *Vegetation dynamics development tool: User's guide version 3.0*. ESSA Technologies Ltd., Vancouver, British Columbia, Canada.

Beutel, T. S., R. J. S. Beeton, and G. S. Baxter. 1999. Building better wildlife-habitat models. *Ecography* 22:219–223.

Beven, K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* 16:41–51.

Box, G. E. P. 1979. Robustness in scientific model building. Pages 201–236 *in* R. L. Launer, G. N. Wilkinson, editors. *Robustness in statistics.* Academic Press, New York, New York, USA.

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological Modelling* 157:281–300.

Brand, G. J., and M. R. Holdaway. 1989. Assessing the accuracy of TWIGS and STEMS85 volume predictions: A new approach. *Northern Journal of Applied Forestry* 6:109–114.

Brookshire, B. L., and S. R. Shifley, editors. 1997. *Proceedings of the Missouri Ozark Forest Ecosystem Project symposium: An experimental approach to landscape research*. U.S. Forest Service, North Central Research Station, General Technical Report NC-193, St. Paul, Minnesota, USA.

Buchman, R. G., and S. R. Shifley. 1982. Guide to evaluating forest growth projection systems. *Journal of Forestry* 81:232–234.

Burgman, M. A., D. R. Breininger, B. W. Duncan, and S. Ferson. 2001. Setting reliability bounds on suitability indices. *Ecological Applications* 11:70–78.

Burns, R. M., and B. H. Honkala. 1990. *Silvics of North America*. Volume 1 and 2. U.S. Forest Service, Agriculture Handbook 654, Washington, D.C., USA.

Buskirk, S. W., and J. J. Millspaugh. 2006. Metrics of use and availability in studies of resource selection. *Journal of Wildlife Management* 70:358–366.

Cartwright, N. 1983. *How the laws of physics lie*. Clarendon Press, Oxford, United Kingdom.

Caswell, H. 1978. A general formula for the sensitivity of population growth rate to changes in life history parameters. *Theoretical Population Biology* 14:215–230.

Chew, J. D. 1995. *Development of a system for simulating vegetative patterns and processes at landscape scales*. Dissertation, University of Montana, Missoula, USA.

Chew, J. D., C. Stalling, K. Moeller, E. Bella, R. Schreiner, R. Ahl, and T. Jones. 2007. User guide for SIMPPLLE, Version 2.5 (Draft). U.S. Forest Service, General Technical Report RMRS- xxx. Published on line. http://www.fs.fed.us/rm/missoula/4151/SIMPPLLE/files/publications/SIMPPLLE-MANUAL-2-5.pdf

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37–46.

Conroy, M. J., and C. T. Moore. 2002. Wildlife habitat modeling in an adaptive framework: The role of alternative models. Pages 205–218 *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. *Predicting species occurrences: Issues of accuracy and scale.* Island Press, Washington, D.C., USA.

Couto, P. 2003. Assessing the accuracy of spatial simulation models. *Ecological Modelling* 167:181–198.

Cumming, G. S. 2000. Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography* 27:441–455.

Dixon, G. 1998. *Evaluating stand density management alternatives using the Forest Vegetation Simulator*. U.S. Forest Service, Forest Management Service Center, DRAFT paper, Fort Collins, Colorado, USA.

Dixon, G. E. compiler. 2003. *Essential FVS: A user's guide to the Forest Vegetation Simulator*. U.S. Forest Service, Forest Management Service Center, Internal Report, Fort Collins, Colorado, USA.

Ek, Alan R., S. R. Shifley, T. E. Burk, editors. 1998. *Forest growth modeling and prediction* (Volumes 1 & 2). U.S. Forest Service, North Central Forest Experimental Station, Genral Technical Report NC-120, St.Paul, Minnesota, USA.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:31–46.

Freeman, S. N., D. G. Noble, S. E. Newson, and S. R. Baillie. 2007. Modelling population changes using data from different surveys: The common birds census and the breeding bird survey. *Bird Study* 54:61–72.

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

Gustafson, E. J., and T. R. Crow. 1999. HARVEST: Linking timber harvesting strategies to landscape patterns. Pages 309–322 *in* D. J. Mladenoff and W. L. Baker, editors. *Spatial modeling of forest landscape change: Approaches and applications.* Cambridge University Press, Cambridge, United Kingdom.

Gustafson, E. J., N. L. Murphy, and T. R. Crow. 2001. Using a GIS model to assess terrestrial salamander response to alternative forest management plans. *Journal of Environmental Management* 63:281–292.

Gustafson, E. J., and L. V. Rasmussen. 2005. *HARVEST for Windows user's guide: Version 6.1*. U.S. Forest Service, North Central Research Station, St. Paul, Minnesota, USA. <http://www.ncrs.fs.fed.us/4153/Harvest/v61/download/Harv61%20User.pdf>.

Hamming, R. W. 1975. How do you know the simulation is relevant? *Simulation* 25:163–167.

He, H. S., W. Li, B. R. Sturtevant, J. Yang, Z. B. Shang, E. J. Gustafson, and D. J. Mladenoff. 2005. *LANDIS 4.0 user's guide. LANDIS: A spatially explicit model of forest landscape disturbance, management, and succession*. U.S. Forest Service, North Central Research Station, General Technical Report NC-263, St. Paul, Minnesota, USA.

He, H. S., D. J. Mladenoff, and J. Boeder. 1999. An object-oriented forest landscape model and its representation of tree species. *Ecological Modelling* 119:1–19.

Hirzel, A. H., G. Le-Lay, V. Helfer, C. Randin, and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199:142–152.

Holling, C. S., editor. 1978. *Adaptive environmental assessment and management.* John Wiley and Sons, New York, New York, USA.

Holthausen, R. S., M. J. Wisdom, J. Pierce, D. K. Edwards, and M. M. Rowland. 1994. *Using expert opinion to evaluate a habitat effectiveness model for elk in western Oregon and Washington*. U.S. Forest Service, Pacific Northwest Research Station, Research paper PNW-RP-479, Portland, Oregon, USA.

Hurley, J. F. 1986. Summary: Development, testing, and application of wildlife-habitat models—The manager's viewpoint. Pages 151–153 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. *Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates.* University of Wisconsin Press, Madison, USA.

Husch, B., T. W. Beers, and J. A. Kershaw. 2003. *Forest mensuration*. John Wiley and Sons, New York, New York, USA.

Iverson, L. R., A. M. Prasad, and M. W. Schwartz. 2005. Predicting potential changes in suitable habitat and distribution by 2100 for tree species of the eastern United States. *Journal of Agricultural Meteorology* 61:29–37.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. *Ecology* 61:65–71.

Johnson, D. H. 2001. Validating and evaluating models. Pages 105–122 *in* T. M. Shenk, A. B. Franklin, editors. *Modeling in natural resource management: Development, interpretation, and application.* Island Press, Washington, D.C., USA.

Johnson, M. D. 2007. Measuring habitat quality: A review. *Condor* 109:489–504.

Johnson, P. S., S. R. Shifley, and R. Rogers. 2002. *The ecology and silviculture of oaks*. CABI Publishing, New York, New York, USA.

Keane, R. E., D. Long, D. Basford, and B. A. Levesque. 1997. Simulating vegetation dynamics across multiple scales to assess alternative management strategies. Pages 310–315 *in Conference Proceedings—GIS 97, 11th Annual Symposium on Geographic Information Systems. Integrating Spatial Information Technologies for Tomorrow.* GIS World, Vancouver, British Columbia, Canada.

Keane, R. E., R. Parsons, and P. Hessburg. 2002. Estimating historical range and variation of landscape patch dynamics: Limitations of the simulation approach. *Ecological Modelling* 151:29–49.

Larson, M. A., W. D. Dijak, F. R. Thompson, III, and J. J. Millspaugh. 2003. *Landscape-level habitat suitability models for twelve species in southern Missouri*. U.S. Forest Service, North Central Research Station, General Technical Report NC–233, St. Paul, Minnesota, USA.

Larson, M. A., F. R. Thompson, III, J. J. Millspaugh, W. D. Dijak, and S. R. Shifley. 2004. Linking population viability, habitat suitability, and landscape simulation models for conservation planning. *Ecological Modelling* 180:103–118.

Leary, R. A. 1996. Testing models of unthinned red pine plantation dynamics using a modified Bakuzis matrix of stand properties. *Ecological Modelling* 98:35–46.

Luck, G. W. 2002. The habitat requirements of the rufous treecreeper (*Climacteris rufa*). 2. Validating predictive habitat models. *Biological Conservation* 105:395–403.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier, San Diego, California, USA.

Mankin, J. B., R. V. O'Neill, H. H. Shugart, and B. W. Rust. 1979. The importance of validation in ecosystem analysis. Pages 309–317 *in* H. H, Shugart, R. V. O'Neill, editors. *Systems ecology.* Dowden, Hutchinson and Ross, Stroudsburg, Pennsylvania, USA.

Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. McDonald, and W. P. Erickson. 2002. *Resource selection by animals: Statistical design and analysis for field studies*. Second edition. Kluwer Academic, Dordrecht, The Netherlands.

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

Marzluff, J. M., J. J. Millspaugh, K. R. Ceder, C. D. Oliver, J. Withey, J. B. McCarter, C. L. Mason, and J. Comnick. 2002. Modeling changes in wildlife habitat and timber revenues in response to timber management. *Forest Science* 48:191–202.

Marzluff, J. M., J. J. Millspaugh, P. Hurvitz, and M. S. Handcock. 2004. Resource utilization of an avian nest predator: Relating resources to a probabilistic measure of animal use. *Ecology* 85:1411–1427.

Matthews, S. N., L. R. Iverson, A. M. Prasad, and M. P. Peters. 2007. *Climate change bird atlas: A spatial database of 147 bird species of the eastern United States*. U.S. Forest Service, Northern Research Station, NRS-4151, Delaware, Ohio, USA. <http://www.nrs.fs.fed.us/atlas/bird>.

Mayer, D. G., and D. G. Butler. 1993. Statistical validation. *Ecological Modelling* 68:21–32.

Miller, D. R., G. Butler, and L. Bramall. 1976. Validation of ecological system models. *Journal of Environmental Management* 4:383–401.

Millspaugh, J. J., K. J. Raedeke, G. C. Brundige, and R. A. Gitzen. 2000. Elk and hunter space use sharing in the Southern Black Hills, South Dakota. *Journal of Wildlife Management* 64:994–1003.

Millspaugh, J. J., R. A. Gitzen, B. J. Kernohan, M. Larson, and C. Clay. 2004. Comparability of three analytical techniques to assess joint space use. *Wildlife Society Bulletin* 32:148–157.

Millspaugh, J. J., R. M. Nielson, L. McDonald, J. M. Marzluff, R. A. Gitzen, C. D. Rittenhouse, M. W. Hubbard, and S. L. Sheriff. 2006. Analysis of resource selection using utilization distributions. *Journal of Wildlife Management* 70:384–395.

Miner, C. L., N. R. Walters, and M. L. Belli. 1988. *A guide to the TWIGS program for the North Central United States*. U.S. Forest Service, North Central Forest Experiment Station, General Technical Report NC-125, St. Paul, Minnesota, USA.

Mitchell, M. S., J. W. Zimmerman, and R. A. Powell. 2002. Test of a habitat suitability index for black bears in the southern Appalachians. *Wildlife Society Bulletin* 30:794–808.

Mladenoff, D. J. 2004. LANDIS and forest landscape models. *Ecological Modelling* 180:7–19.

Mladenoff, D. J., T. A. Sickley, and A. P. Wydeven. 1999. Predicting gray wolf landscape recolonization: Logistic regression models vs. new field data. *Ecological Applications* 9:37–44.

Mladenoff, D. J., and H. S. He. 1999. Design and behavior of LANDIS, an object-oriented model of forest landscape disturbance and succession. Pages 163–185 *in* D. J. Mladenoff, and W. L. Baker, editors. *Advances in spatial modeling of forest landscape change: Approaches and applications.* Cambridge University Press, Cambridge, United Kingdom.

Morrison, M. L., and L. S. Hall. 2002. Standard terminology: Toward a common language to advance ecological understanding and application. Pages 43–52 *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. *Predicting species occurrences: Issues of accuracy and scale.* Island Press, Washington, D.C., USA.

Morrison, M. L., B. G. Marcot, and R. W. Mannan. 1998. *Wildlife-habitat relationships: Concepts and applications*. Second edition. University of Wisconsin Press, Madison, USA.

Mowrer, H. T., K. Barber, J. Campbell, N. Crookston, C. Dahms, J. Day, J. Laacke, J. Merzenich, S. Mighton, M. Rauscher, R. Sojda, J. Thompson, P. Trenchi, and M. Twery. 1997. *Decision support systems for ecosystem management: An evaluation of existing systems*. U.S. Forest Service, Rocky Mountain Forest and Range Experiment Station, General Technical Report RM-GTR-296, Fort Collins, Colorado, USA.

Newson, S. E., R. J. W. Woodburn, D. G. Noble, S. R. Baillie, and R. D. Gregory. 2005. Evaluating the breeding bird survey for producing national population size and density estimates. *Bird Study* 52:42–54.

Oliver, C. D., and B. C. Larson. 1996. *Forest stand dynamics*. Update edition. John Wiley and Sons, New York, New York, USA.

Oreskes, N., K. Shrader-Frechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641–646.

Ottaviani, D., G. J. Lasinio, and L. Boitani. 2004. Two statistical methods to validate habitat suitability models using presence-only data. *Ecological Modelling* 179:417–443.

Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133:225–245.

Pontius, R. G. Jr., D. Huffaker, and K. Derman. 2004. Useful techniques of validation for spacially explicit land-change models. *Ecological Modelling* 179:445–461.

Prasad, A. M., L. R. Iverson, S. Matthews, and M. Peters. 2007. *Climate change tree atlas: A spatial database of 134 forest tree species of the eastern United States*. U.S. Forest Service, Northern Research Station, Delaware, Ohio, USA. <http://www.nrs.fs.fed.us/atlas/tree>.

Ray, N., and M. A. Burgman. 2006. Subjective uncertainties in habitat suitability maps. *Ecological Modelling* 195:172–186.

Regan, H. M., M. Colyvan, and M. A. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12:618–628.

Reynolds, M. R. 1984. Estimating the error in model predictions. *Forest Science* 30:454–469.

Rittenhouse, C. D. 2008. Wildlife response to spatial and temporal changes in forest habitat. Dissertation, University of Missouri, Columbia, USA.

Roloff, G. J., and B. J. Kernohan. 1999. Evaluating the reliability of habitat suitability index models. *Wildlife Society Bulletin* 27:973–985.

Roloff, G. J., J. J. Millspaugh, R. A. Gitzen, and G. C. Brundige. 2001. Verification of a spatially explicit habitat model for Rocky Mountain elk. *Journal of Wildlife Management* 65:899–914.

Root, K. V., H. R. Akçakaya, and L. Ginzburg. 2003. A multispecies approach to ecological valuation and conservation. *Conservation Biology* 17:196–206.

Rykiel, E. J., Jr. 1996. Testing ecological models: The meaning of validation. *Ecological Modelling* 90:229–244.

Scheller, R. M., J. B. Domingo, B. R. Sturtevant, J. S. Williams, A. Rudy, E. J. Gustafson, and D. J. Mladenoff. 2007. Design, development, and application of LANDIS-II, a spatial landscape simulation model with flexible temporal and spatial resolution. *Ecological Modelling* 201:409–419.

Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. 2002. *Predicting species occurrences: Issues of accuracy and scale.* Island Press, Washington, D.C., USA.

Shifley, S. R., and B. L. Brookshire, editors. 2000. *Missouri Ozark Forest Ecosystem Project: Site history, soils, landforms, woody and herbaceous vegetation, down wood, and inventory methods for the landscape experiment*. U.S. Forest Service, North Central Research Station, General Technical Report NC-208, St. Paul, Minnesota, USA.

Shifley, S. R., and J. M. Kabrick, editors. 2002. *Proceedings of the 2nd Missouri Ozark Forest Ecosystem Project symposium: Post treatment results of the landscape experiment*. U.S. Forest Service, North Central Research Station, General Technical Report NC-227, St. Paul, Minnesota, USA.

Shifley, S. R., F. R. Thompson, III, W. D. Dijak, M. A. Larson, and J. J. Millspaugh. 2006. Simulated effects of forest management alternatives on landscape structure and habitat suitability in the Midwestern United States. *Forest Ecology and Management* 229:361–377.

Shifley, S. R., and C. W. Woodall. 2007. The past, present, and future of Indiana's oak forests. Pages 295–304 *in* D. S. Buckley and W. K. Clatterbuck, editors. *Proceedings of the 15th central hardwood forest conference*. U.S. Forest Service, Southern Research Station, e-General Technical Report SRS-101, Asheville, North Carolina, USA.

Somershoe, S. G., D. J. Twedt, and B. Reid. 2006. Combining breeding bird survey and distance sampling to estimate density of migrant and breeding birds. *Condor* 108:691–699.

Starfield, A. M. 1997. A pragmatic approach to modeling for wildlife management. *Journal of Wildlife Management* 61:261–270.

Starfield, A. M., and A. L. Bleloch. 1991. *Building models for conservation and wildlife management*. Second edition. Burgess International Group, Edina, Minnesota, USA.

Stearns, S. C. 1992. *The evolution of life histories*. Oxford University, Oxford, United Kingdom.

Syphard, A. D., K. C. Clarke, and J. Franklin. 2007. Simulating fire frequency and urban growth in southern California coastal shrublands, USA. *Landscape Ecology* 22:431–445.

Termansen, M., C. J. McClean, and C. D. Preston. 2006. The use of genetic algorithms and Bayesian classification to model species distributions. *Ecological Modelling* 192:410–424.

Thompson, W. L. editor. 2004. *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters.* Island Press, Washington, D.C., USA.

Twery, M. J., P. D. Knopp, S. A. Thomasma, H. M. Rauscher, D. E. Nute, W. D. Potter, F. Maier, J. Wang, M. Dass, H. Uchiyarna, A. Glende, and R. E. Hoffman. 2005. NED-2: A decision support system for integrated forest ecosystem management. *Computers and Electronics in Agriculture* 39:24–43.

U.S. Forest Service. 2008*a*. Forest inventory and analysis national program. USFS webpage. <http://www.fia.fs.fed.us/>. Accessed 3 January 2008.

U.S. Forest Service. 2008*b*. Forest vegetation system. USFS webpage. <http://www.fs.fed.us/fmsc/fvs/>. Accessed 12 January 2008.

University of Washington. 2008. Landscape management system. UW webpage. <http://lms.cfr.washington.edu/index.php>. Accessed 12 January 2008.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. *Journal of Wildlife Management* 47:893–901.

Van Horne, B. 2002. Approaches to habitat modeling: The tensions between pattern and process and between specificity and generality. Pages 63–72 *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. *Predicting species occurrences: Issues of accuracy and scale.* Island Press, Washington, D.C., USA.

Vanclay, J. K. 1994. *Modelling forest growth and yield: Applications to mixed tropical forests*. CAB International, Wallingford, United Kingdom.

Vanclay, J. K., and J. P. Skovsgaard. 1997. Evaluating forest growth models. *Ecological Modelling* 98:1–12.

Verner, J., M. L. Morrison, and C. J. Ralph, editors. 1986. *Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates.* University of Wisconsin Press, Madison, USA.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and management of animal populations: Modeling, estimation, and decision making*. Academic Press, San Diego, California, USA.

Woeck, B. N. 2003. *Demographics and space use of mule deer and white-tailed deer in Custer State Park, South Dakota*. Thesis, University of Missouri, Columbia, USA.

Woodall, C., D. Johnson, J. Gallion, C. Perry, B. Butler, R. Piva, E. Jepsen, D. Nowak, and P. Marshall. 2005. *Indiana's forests 1999-2003 (Part A)*. U.S. Forest Service, North Central Research Station, Resource Bulletin NC-253A, St. Paul, Minnesota, USA.

Yang, J., He, H., Shifley, S. R., and Gustafson, E. J. 2007. Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands. *Forest Science* 53:1–15.

Yang, J., H. He, and S. R. Shifley. 2008. Spatial controls of occurrence and spread of wildfires in the central hardwood forests of the Missouri Ozark Highlands. *Ecological Applications*, *In press*.

Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39:561–577.