# FORUM
# FORUM
# FORUM

# Strategies for ecological extrapolation

*Debra P. C. Peters and Jeffrey E. Herrick, USDA ARS, Jornada Experimental Range, Box 30003, MSC 3JER, NMSU, Las Cruces, NM 88003-0003, USA (debpeter@nmsu.edu). – Dean L. Urban, Nicholas School of the Environment and Earth Sciences, Duke Univ., Durham, NC 27708-0328, USA. – Robert H. Gardner, Appalachian Laboratory, Univ. of Maryland Center for Environmental Science, Frostburg, MD 21532, USA. – David D. Breshears, Los Alamos National Laboratory Atmosphere, Climate and Environmental Dynamics Group (EES-2) Mail Stop J495, Los Alamos, NM 87545, USA.*

Errors associated with the spatial extrapolation of information are an important issue in ecology. Although improvements in instrumentation and software have increased our ability to represent complex behaviors, a measurable reduction in prediction errors is not always achieved with increased complexity. This difficulty in generating accurate predictions is due, in part, to trade-offs in error propagation that exist between simple and complex approaches: simple models often exclude key processes resulting in unknown prediction biases while complex approaches reduce prediction bias at the expense of increased estimation and measurement error. Our inability to anticipate and measure these trade-offs impedes our ability to define and control prediction errors.

We develop a framework to evaluate and compare three classes of approaches to the problem of spatial extrapolation. Trade-offs in realism and potential errors for these classes are illustrated using a case study of the northern spotted owl. A variety of spatial problems are then considered and alternatives to dealing with spatial information are explored. Comparisons of model alternatives show that compelling reasons must exist before complex, spatially explicit models are required for prediction. Given the complexity of these issues, our framework assists ecologists in clarifying these problems and in selecting the most appropriate approach for a given objective.

Ecologists are increasingly faced with the need to use short-term, local measurements to assess patterns of change at landscape, regional and even global scales (Wessman 1992, Dale et al. 2001). Because broad-scale assessments rely on data obtained at much finer scales, statistical or simulation methods are often used to extrapolate data in space (i.e. "scaling"; here we use the term "scaling" in a restricted sense to refer to the use of measurements made at one spatial scale to provide information at a different, and typically broader, scale) and time (i.e. "forecasting"). Extrapolation of information across scales is fraught with difficulties (Levin 1992, Clark et al. 2001), many of which directly affect the precision and accuracy of results (Heuvelink 1998). These difficulties are further exacerbated when spatial interactions produce large responses that overwhelm local effects. For example, fine-grained patterns of plant community development depend on competitive interactions among plants (Tilman 1994), but local competitive effects are less important when contagious processes such as fire determine vegetation patterns (Miller and Urban 1999). Under these conditions, accurate predictions depend upon our ability to identify the relevant spatial processes and to determine their effects on system dynamics. Although others have classified predictive methods based on the degree of spatial information included (Baker 1989), or provided quantitative methods to extrapolate information across scales (King 1991), guidelines to determine when the inclusion of spatial information increases the accuracy of ecological predictions have yet to be developed.

Our goal is to provide a framework for extrapolation relevant to a broad range of ecologists from different disciplines, with different perspectives, and with different levels of mathematical expertise or familiarity with quantitative methods. Models of extrapolation range from simple to complex; our approach provides guidance in the selection of the appropriate model and the likely consequences of that selection to errors in prediction. Although scaling (spatial) and forecasting (temporal) predictions can be distinguished, the conceptual and issues of both are similar. Thus, we focus on spatial extrapolation and recognize the relevance to temporal applications (terms and definitions provided in Table 1).

The importance of spatial information to the accuracy of predictions affects a diversity of ecological problems, including climate change (Eastman et al. 2001), wildlife conservation (Hansen et al. 1993), biodiversity (Dale et al. 1994), species invasions (Pitelka et al. 1997), meta-population dynamics (Bascompte and Sole 1996), spread of disturbance such as fire (Heyerdahl et al. 2001), soil erosion (Pimentel and Skidmore 1999), and carbon sequestration (Solomon et al. 1993, Bachelet et al. 2001). Spatial information is not always required, particularly when spatial effects are inconsequential or the scales for prediction are sufficiently large. However, the relative value of including spatial effects is rarely addressed, in part because there is no clearly defined set of guidelines for this evaluation.

The need to include spatial information is only justified when the errors associated with predictions are substantially reduced. The literature on prediction errors is extensive (O'Neill 1973, Gardner et al. 1981, Parysow et al. 2000), and will not be reviewed here. However, two "error" issues are noteworthy: (1) large quantities of detailed information at local scales do not ensure accurate predictions at broad spatial scales; (2) comparisons of prediction errors for alternative model formulations (Gardner et al. 1982, Reynolds and Acock 1985) show that simple approaches may be sufficient when processes are linear and parameters are uncertain (Deutschman et al. 1999).

We present a framework to determine when the inclusion of spatial information increases the accuracy of ecological predictions. This framework includes a simple classification system based on the degree of spatial interactions. Sources of error due to "problem misclassification" are identified followed by an example of trade-offs in model selection using a case study of the northern spotted owl. Finally, a discussion of the characteristics of ecological systems that require spatial approaches is provided along with a statistical evaluation of errors.

## Classes of models

A framework for relating data or information on one variable from one scale to another scale can be developed from the general expression for a regionalized variable used in geostatistics. If a variable z at location i includes a combination of a broad-scale forcing (trend) component plus a local spatial component, then:

$$z_i = f(\underline{x_i}) + s_i + \varepsilon \tag{1}$$

where the vector $\underline{x}$ is a set of explanatory variables at location i, s is a function of values measured in the neighborhood of location i, and $\varepsilon$ is an appropriate error term. Equation 1 provides the basis for a classification of three classes of approaches: class 1 is a non-spatial approach that makes no explicit reference to location, and simplifies to $z = f(\underline{x}) + \varepsilon$; class 2 is a spatially implicit approach that references location, $\underline{x_i}$, but not the neighborhood, $s_i$, thus estimating z is equal to $f(\underline{x_i}) + \varepsilon$; and class 3 is a spatially explicit approach that corresponds to the full geostatistical model, and includes an explicit reference to the neighborhood of the location of interest (Eq. 1). We next describe these three approaches in detail.

## Class 1. Non-spatial models

The simplest prediction uses a model where the independent variables, $\underline{x}$ are measured without regard to spatial location:

$$z = f(\underline{x}) + \varepsilon \tag{2}$$

Although the spatial distribution of the x's may be non-random, this distribution is unknown, insignificant or prohibitively expensive to measure. As an example, one may wish to predict an ecological property, such as biomass, as a function of a set $(\underline{x})$ of explanatory variables associated with a plant community type.

Table 1. Definitions of terms.

| | |
|---|---|
| Predict[1]: | to declare or indicate in advance, to foretell on the basis of observation, experience or scientific reason |
| Extrapolate (interpolate): | to extend the spatial extent or refine the resolution of measured data to arrive at a broader scale or finer grained estimate |
| Synoptic estimators: | any estimator that provides a prediction of the state of the system simultaneously over a broad area, either as steady-state (average) conditions or at a specified point in time |
| Scenario: | definition of the boundary conditions for a model extrapolation or prediction |
| Project: | a conjecture about the future based on observations, measurements or experience, subject to a specified model or scenario. Estimates are not necessarily associated with explicit incorporation of errors or uncertainties |
| Forecast: | an explicit projection, including error estimates, of an event or condition at a specified future time, based on a specific model, scenario, and initial conditions |
| Sampled area: | the location and scale (both spatial and temporal) of measurements which will serve as the basis of predictions |
| Target area: | the area (defined spatially and temporally) to which predictions will be applied |

[1] "Predict" is used here as "to foretell in advance based on observation, experience or scientific reason." This definition is more general than often used in earth sciences (Pielke 2002) and ecology (Clark et al. 2001). We believe that a general definition is preferred when common usage makes a stricter definition unnecessary and confusing.

Independent samples provide point estimates for each community type, and prediction is effected by assigning a single value of z to each community. Thus, the z's vary from location to location as a function of changes in the explanatory variables, but identical values of z will always be predicted for the same set of variables.

A familiar non-spatial example is the prediction of net primary production by community type based on plot-scale estimates that are averaged and weighted by the area to extrapolate to each community (Lieth and Whittaker 1975). Modeling approaches based on "look-up tables" are similarly non-spatial. For example, one can simulate effects of temperature, precipitation, and soil type on vegetation growth, and save predictions in a "look-up table". One applies this look-up table by matching the descriptors for each study plot with the corresponding tabulated values.

If maps of variables describing each location are available, then maps of predicted values can be generated from look-up tables. These apparently spatial predictions are, in reality, non-spatial models because the neighborhood variables, $s_i$, are unknown or unused. In statistical terms, spatial trends within the landscape are small, autocorrelations among explanatory variables cannot be detected (or are ignored), landscape context or neighborhood effects are unimportant, and the absence of these relationships is assumed to have marginal effects on the error term.

## Class 2. Spatially implicit models

The spatially implicit approach is distinguished from non-spatial predictions by the inclusion of significant spatial relationships among explanatory variable(s):

$$z_i = f(\underline{x_i}) + \varepsilon \qquad (3)$$

Because the physical environment is spatially structured and ecological processes are often correlated with spatial variation in environmental factors, significant spatial structure in explanatory variables is the rule rather than the exception, thus justifying this approach for many objectives (Legendre et al. 1989, Borcard et al. 1992). The use of correlated and spatially structured driving variables generates spatially-structured output, even if the model does not reference location. Consider the previous example of a model driven by a vector of input variables. If the input vectors are sampled at particular locations – such as with geospatial data – then the model output will have spatial structure corresponding to the input. Simulations of net primary production, nitrogen mineralization, and decomposition rates in the Central Great Plains using statistical models or biogeo-chemical simulation models illustrate this approach (Burke et al. 1991, 1997). Most gap models that simulate grassland or forest successional dynamics (Shugart 1984,

Botkin 1993, Peters 2002) and most biogeographic models for regional to global scales (Melillo et al. 1995, Neilson and Running 1996) provide further examples.

The spatially implicit case is essentially a non-spatial model driven by spatially structured data. This spans a spectrum of cases. At its simplest, this approach has very few driving variables and little structure that is not appreciably different from a non-spatial model. As the number of variables increases and the degree of auto-correlation among variables increases, the likelihood that spatial structure will be generated also increases. This trend can be amplified by including driving variables that reference spatial context, such as distance to stream or upslope contributing area. Although these variables are calculated at a point, after they are computed they are spatially implicit because location is not tracked explicitly. In the limit, of course, if one measured enough spatially structured variables then every point would be unique and this approach would be essentially spatially explicit.

The spatially implicit approach does not include spatial processes, nor does it explicitly reference the neighborhood or landscape context of the study plot. Yet these models, when evaluated over an empirical joint distribution of input variables, result in output maps that are spatially structured.

## Class 3. Spatially explicit models

A spatially explicit approach is distinguished by the inclusion of neighborhood effects ($s_i$) for the estimation of z that results in Eq. 1. Neighborhood effects are important when three conditions exist: (a) transfers among sites of materials, organisms, energy, etc. are large; (b) rates of transfer are determined by local conditions; and (c) variables governing the local-scale transfers are spatially heterogeneous. A variety of statistical, empirical or physical-based approaches may be used to estimate $S_i$ by accounting for effects of spatially related cells. "Structure functions" such as variograms used in geostatistics (kriging) explicitly address local spatial structure (Rossi et al. 1992). Similarly, in spatial regressions (e.g. conditional or simultaneous autoregression), a term is fit to account for the effect at location i of variables measured at neighboring locations (Keitt et al. 2002).

Familiar examples of spatially explicit models include cellular automata (Hogeweg 1988), models that compute seed dispersal likelihood based on the locations of target sites relative to source sites (Coffin and Lauenroth 1994, Clark et al. 1998), and models of contagious distur-bances such as fire and disease that simulate interacting cells (Turner et al. 1989). Likewise, hydrology models with 3-dimensional structure that simulate lateral fluxes

(e.g. RHESSYS, Band et al. 2001) are spatially explicit, whereas hydrology models driven by topographic indices (e.g. TOPMODEL, Beven and Kirkby 1979) are spatially implicit. Spatially explicit coupled atmospheric-ecological process models include Eastman et al. (2001) and Lu et al. (2001) where atmospheric winds and turbulence provide horizontal transfers of material among locations.

The decision to simulate space explicitly has serious implications for sampling and model-fitting. Explicit inclusion of local effects in $S_i$ requires a much larger set of data for model calibration and testing than for the other two approaches. Data needed to estimate such models cannot be collected using independent (i.e. not autocorrelated) samples; it is the lack of such independence that makes these models interesting.

## Sources of error and consequences of class selection

The distinction between the three approaches is determined by two factors: (1) the presence or absence of spatial dependencies in the explanatory variables (absent in non-spatial models, present in spatially implicit and explicit models); and (2) the presence or absence of a neighborhood function (present only in spatially explicit models). Distinctions between approaches are important because the error structure associated with predictions differs by approach. Simple models are often optimal when information is imprecise (O'Neill 1979, Reynolds and Acock 1985). However, more complicated models may be better when dynamics are complex and extensive data are available; yet these data may be expensive to collect and contain a number of small errors that accumulate to produce disproportionately large uncertainties in predictions (Gardner et al. 1980). The relative trade-offs between omitting key parameters ("errors of omission" in simple models) and errors associated with including many parameters of unknown importance ("errors of commission" in complex models) (Fig. 1) are best determined by explicit consideration of the sources of error in each approach.

Four sources of error are most relevant to uncertainty in predictions from spatial models (Table 2): Measurement error is an attribute of the data, and may be due to imprecision in measurements, or because the phenomenon of interest cannot be directly measured. For example, there may be uncertainties in estimates of soil nutrients due to uncontrolled sampling error or to variability in lab results. Model error stems from representation of a process of interest, including deciding which drivers to include. All models are abstractions of reality and subject to model error. These errors become serious when non-linear processes are represented as linear functions or the size of the neighborhood and magnitude of local interactions are incorrectly specified.
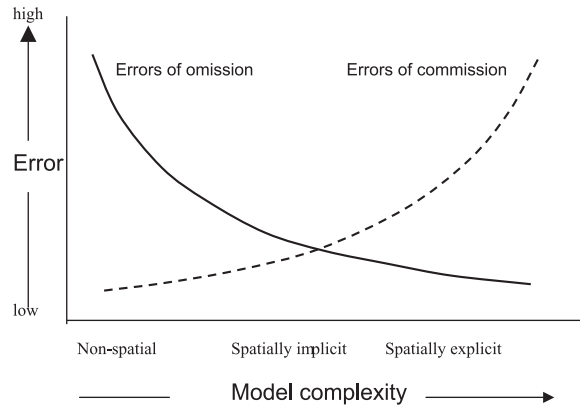


Fig. 1. Prediction errors of omission associated with excluding processes decrease as model complexity increases whereas errors associated with including processes (errors of commission) increase.

Estimation error is associated with parameters selected to calibrate the model, and is often indexed by the standard errors of regression coefficients. It is often difficult to isolate estimation error from model error because the choice of a particular model also dictates the number and kinds of parameters. Finally, process error is the natural variation of processes that are simulated as stochastic functions (e.g. weather, disturbances, mortality, etc.). All four sources lead to increased uncertainty in predictions due to both reduced accuracy (biased predictions) and reduced precision (more variable predictions).

## Trade-offs in model selection: a case study using the northern spotted owl

We examine trade-offs in realism and potential errors for the three approaches using a well-documented case study of the northern spotted owl (*Strix occidentalis*). At issue

Table 2. Sources of error in spatial or temporal predictions or extrapolations.

| Source | Definition |
|---|---|
| Measurement error | Variation associated with imperfect measurements at discrete points in space or time of processes or states that vary continuously over these scales |
| Model error | Uncertainty, inadequacy, or bias resulting from the choice of particular forms (equations) or components (variables, covariates) as a representation of the real system |
| Estimation error | Uncertainty or error resulting from statistical confidence in estimating the values of coefficients and parameters implemented as a model |
| Process error | Variability inherent to the system, especially expressed as 'natural' variation in variables or parameters of an implemented model |

was the tension between logging interests and demands to preserve old-growth forests as potential owl habitat (Thomas et al. 1990; FEMAT 1993, Marcot and Thomas 1997). Because of the huge spatial scale of the analysis (the Pacific northwestern US) and the long-term perspective required by the pace of forest dynamics and owl longevity, models played a crucial role in the assessment. McKelvey et al. (1993) provide an insightful discussion of the modeling process and document the sequence of increasingly complicated models used.

The simplest model was a non-spatial age projection model or Leslie matrix with three age classes. The model simulated population growth (birth and death), and required four parameters (three age-specific survivorship rates and a single fecundity rate). The model inferred the finite rate of population increase ($\lambda$). Because of the large amount of data used to estimate the demographic rates, it was possible to conclude with a level of certainty greater than measurement error that the population was decreasing. However, this approach ignored effects of owl dispersal on population dynamics.

The second approach invoked spatially implicit meta-population models to assess the effects of dispersal and extinction on landscape-level abundance of the spotted owl. Two metapopulation models were used that estimated dispersal success as a function of patch density and geometry – actual patch locations (and distance among patches) were not considered. The models used the same demographic parameters to describe fecundity and mortality as in the Leslie matrix, and added dispersal parameters. This model suggested that groups of habitat patches would support conservation efforts more effectively than isolated patches. However, this model did not provide insight into the effects of patch size, shape, or orientation on owl dynamics.

The third model was an individual-based simulator with a spatially explicit description of habitat patches. The model simulated habitat selection, breeding, and dispersal of individual male and female owls on a raster-based map of habitat types. This model used the same demographic parameters as the previous models, but added a dispersal process as a correlated random walk. A number of parameters were required to describe the spatial behavior of owls, such as their tendency to move in the same or different directions and their attraction to or avoidance of different habitats. In addition, the demographic parameters varied with habitat type. Thus, for the simplest case of a binary map ("habitat" vs "not habitat"), the model required 18 parameters; if three habitat types were simulated the total was doubled, and so on. This model provided insights into the effects of habitat amount and spatial complexity on owl dynamics as well as a tool to explore scenarios about owl dispersal behavior and source/sink dynamics.

From the perspective of model trade-offs (Fig. 1), the models perform as expected in terms of potential "errors of commission" (Fig. 2). The number of parameters increases dramatically with the spatially explicit model, and would be more dramatic if additional habitat types had been considered. Only four parameters in the spatially explicit model were estimated from data – the same four used in the Leslie matrix. The remainder of the parameters were estimated from expert opinion. Thus, the use of a spatially explicit model increased parametric complexity while simultaneously decreasing the proportion of parameters estimated from data. The "errors of omission" curve is difficult to estimate because all models omit information. We offer the "omissions curve" as a heuristic device to compare models. The curve decreases in number of processes omitted as model complexity increases (Fig. 2).

In terms of model uncertainty, McKelvey et al. (1993) illustrate the trade-offs involved in evaluating models. With the Leslie matrix, the solution is deterministic with an exact solution that depends only on model parameters. The advantage of this approach is that effects of parameter error can be evaluated analytically. In the spatially implicit models, the solution is also deterministic, but numerical solutions are required. With the individual-based simulator, the stochastic nature of demographics and dispersal required that model behavior be assessed from a set of replicate simulations. This approach provided a minimal expression of uncertainty associated with process error, and also provided the confidence limits typical of stochastic simulations. Although McKelvey et al. (1993) did not examine all sources of potential uncertainty (i.e. effects of measurement, model, and estimation error associated with dispersal), they went much further than most modeling
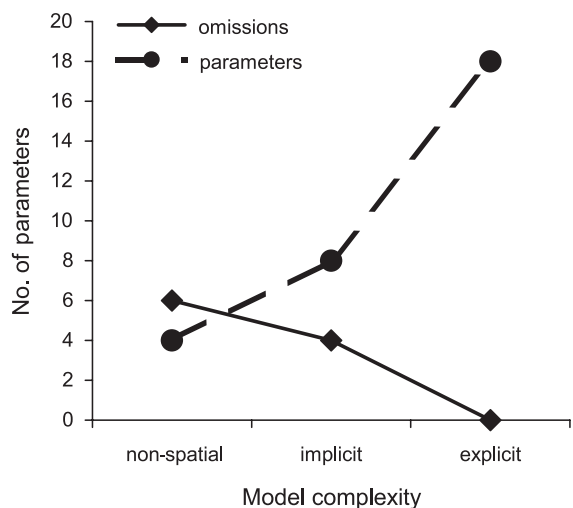


Fig. 2. The number of parameters included or omitted for each of three models of increasing complexity used in the spotted owl example. The total number of parameters increases and the proportion of parameters omitted from the analysis decreases as model complexity increases.

studies to examine effects of model sensitivity and uncertainty on management recommendations. It is important to emphasize that none of the spotted owl models is correct (all models are simplifications of reality), nor is there an obvious best model (model selection depends on the objectives). This case study illustrates the trade-offs between inferential richness due to the spatial complexity of dispersal and landscape context, and the cost of modeling this richness explicitly.

## Issues causing complexity in models: consequences and solutions

Addressing trade-offs between errors of omission and errors of commission, as highlighted in the case study, is the essential challenge when making spatial extrapolations. The adequacy of the extrapolation requires knowledge when a simple model is sufficient or when a more complex, spatial model is required. Four types of ecological situations are likely to require the complexity of either spatially implicit or explicit approaches: (1) non-linear processes; (2) threshold responses; (3) positive feedbacks; and (4) contagious processes that induce spatial fluxes of organisms, material or energy (Table 3). In each situation, it is the interactions with spatial structure that motivate the use of spatial models. These situations can, of course, occur simultaneously. For example, a pest outbreak driven by a contagious process (dispersal) can be reinforced initially by a positive-feedback mechanism (population growth) that leads to a threshold response (the eruption; Grassberger 1983, Cruickshank et al. 1999). Similarly, metapopulation models often show threshold responses to incremental losses in habitat; near this threshold, connectivity becomes increasingly important with dispersal providing a positive feedback that couples the fates of populations in adjacent habitat patches (Bevers and Flather 1999).

## Non-linear processes

A linear process has the property that the expected value of a function, f(x), is equivalent to the function evaluated when the independent values are set to their expected value, x. Non-linear systems do not have this property, thus requiring that the function be integrated across the distributions of x's. Although non-linearities are important in spatial systems, they are often ignored when making spatial extrapolations. For instance, a typical approach is to use area-weighted estimates of spatially distributed properties (Burke et al. 1991, 1997). Consequently, soil maps can be inadequate to predict phosphorous runoff into streams because the scale of the maps is too coarse (Rosenblatt et al. 2000).

## Threshold responses

A special case of non-linearity occurs when the function changes qualitatively at a critical value of the explanatory variable(s). These threshold responses include pest eruptions (Ludwig et al. 1978), watershed impacts of land cover change (Davenport et al. 1998), conversion of grasslands to shrublands (Archer 1989), and lake eutrophication (Hakanson 1999). Spatial patterns that alter connectivity are notorious causes of thresholds (Gardner et al. 1987, Levin and Durrett 1996, With et al. 1997). Attempts to model these thresholds result in highly sensitive responses because the exact location of the thresholds are difficult to specify or, once experienced, to generalize to other landscapes.

Table 3. Classification of ecological scenarios that invite model complexity, and the possible responses to this complexity. See text for explanation and examples.

| Issue[1] | Implication | Alternative responses |
|---|---|---|
| 1. process f($\underline{x}$) is non-linear | E[f($\underline{x}$)] $\neq$ f(E[$\underline{x}$]) | Integrate f($\underline{x}$) over the empirical distribution of x (its pdf) |
| 2. f($\underline{x}$) has a threshold effect at x* | forecasts are sensitive (if modeled) or incorrect (if not modeled); spatial extrapolations uncertain near threshold | (a) ignore if $\underline{x}$ is not near x* (b) parameterize piecewise above and below x* (c) simulate the effect explicitly |
| 3. f($\underline{x}$) has strong positive feedbacks | predictions are incorrect for some parameter combinations (regardless of location); forecasts get worse over time (errors amplify) | (a) ignore if weak (b) parameterize implicitly (c) simulate the feedback explicitly |
| 4. f($\underline{x}$) includes spatial fluxes, lateral transfers, or contagious processes $s_i$ for location i | predictions are incorrect for some locations; errors amplify over time and space | (a) ignore if weak (b) parameterize implicitly (c) simulate the spatial process explicitly |

[1] The issues are framed in terms of the geostatistical equation $z_i = f(\underline{x}_i) + s_i + \epsilon$, where f($\underline{x}$) is a function of a set of variables $\underline{x}$ measured location i and $s_i$ is a spatial neighborhood function local to i.

## Positive feedbacks

Feedbacks in ecological systems may be either positive or negative: negative feedbacks are stabilizing while positive feedbacks are often destabilizing. An example of a positive feedback involves nitrogen (N) status and productivity: high-N sites can increase plant production which generates more litter of higher quality; N mineralization then increases which leads to higher N status that feeds back to plant production. Forest models incorporating this feedback often give rise to complex behaviors, such as successional divergence to alternative stable states (Pastor and Post 1986). By contrast, models that only incorporate forest response to available resources result in an asymptotic approach to steady-state biomass or leaf area (Bormann and Likens 1979). Models that ignore positive feedbacks can be incorrect for soil texture, temperature, and precipitation conditions regardless of their spatial location. Simulation results will be unreliable and become more biased over time due to error amplification. This behavior suggests that the reliability of synoptic predictions depends on time: earlier predictions should be more accurate than later ones that include long-term feedbacks.

## Spatial interactions or contagious (neighborhood) processes

Some ecological processes are inherently spatial, such as lateral hydrologic fluxes (Beeson et al. 2001), propagule dispersal (Coffin and Lauenroth 1994), and contagious disturbances (Miller and Urban 2000). Ignoring such processes can lead to model predictions that are incorrect for some locations that are independent of other parameters at that location. Results from models that ignore important contagious processes are spatial autocorrelation in model errors (residuals; Legendre 1993). These errors may propagate spatially over time: an error at location i at time t can induce error at neighboring locations in the next time step.

## Potential solutions to addressing issues that invite model complexity

There are three approaches to dealing with model complexity: (1) ignore the complexity, (2) represent the complexity in a simplified manner, and (3) explicitly model the complexity. Ignoring complexity is appropriate when there are small effects on model results. For example, in many terrestrial ecosystems, the dominant hydrological and biogeochemical fluxes are vertical; thus lateral fluxes are often assumed to be unimportant, and spatially implicit models are effective for short-term predictions (Parton et al. 1987).

A second approach is to address complexity implicitly. Positive feedbacks can be modeled implicitly for multi-plicative interactions among dependent variables or higher-order forms such as quadratics. Similarly, spatial interactions can be captured implicitly. The approach of indexing dispersal success in terms of patch size and isolation is well established in ecology (Whitcomb et al. 1981 after MacArthur and Wilson 1967). The neighborhood effect of predation and nest parasitism on birds in forest edges can be parameterized implicitly in terms of edge:area ratios or similar indices (Temple 1986).

The third approach is to deal with the complexity explicitly, a necessity for addressing effects of many disturbances (Romme et al. 1998) or any system displaying attributes of self-organized criticality (Maddox 1992). Forecasts developed from simple models will be misleading when systems are near a threshold. For some applications, especially mapped forecasts of community formation as a result of dispersal (Pacala and Silander 1990), spatial interactions must be simulated explicitly. An effective approach may be a combination of the three listed above. For example, one can use a complex model to determine the range of conditions under which a threshold is operative. Given these threshold values, one could reformulate two different models: one for use above and the second for below the threshold.

## Evaluating the appropriateness of a given approach

A priori consideration of the four issues that invite model complexity can aid in both model selection and testing that begin with a spatially referenced data-set independent of model development and parameterization. Predictions from a non-spatial model can be compared to these data, allowing residual errors to be inspected. The adequacy of a given model can be evaluated using prediction errors provided by formal statistics (Draper and Smith 1966). These methods have been used infrequently for evaluating the appropriate level of ecological model complexity (Gardner and Urban 2003).

Statistical approaches can be illustrated within a regression framework. Residuals associated with model predictions are evaluated as $e_i = Y_i - \hat{Y}_i$, where $Y_i$ is the value observed at location i; $\hat{Y}_i$ is the model prediction and $e_i$ is the error (residual) associated with the prediction. The total prediction error is the sum of these errors squared, corrected for the degrees of freedom (mean squared error, MSE). Because variance in the data is the sum of squared deviations from the mean (CSS corrected for sample size), predictability of the model (as in regression) is measured as $R^2 = (CSS - MSE)/CSS$. Model evaluation is most often concerned with increasing $R^2$ and reducing the magnitude of model residuals e.

A model can be deemed adequate if four criteria are satisfied: (1) $R^2$ is sufficiently large; (2) significant trends

in residuals are absent; (3) prediction errors are spatially independent (i.e. residuals are not autocorrelated); and (4) residuals are normally distributed (mean $= 0.0$). The first criterion ($R^2$) has considerable latitude because useful explanatory power depends on the application; if the other conditions are met, even a model with low $R^2$ can provide useful insights about trends in spatial extrapolation or forecasting. The other three criteria have implications for extrapolation (and forecasting); violating these criteria propagates error and produces unacceptable predictions. Model residuals with a significant trend with respect to predictor variables, the predicted variable itself, or spatial location violate criterion 2. Similarly, residuals that are asymmetrically distributed or have a nonzero mean (violation of criterion 4) suggest an inadequate model fit and invite model reformulation.

Spatially dependent errors are of special concern. Spatial trends (criterion 2) appear as trends with respect to broad-scale location variables, such as longitude or latitude. These broad-scale spatial trends are distinct from finer-scale errors that appear as autocorrelated residuals (criterion 3). In terms of a geostatistical model, locational trends are a part of the general forcing $f(\underline{x_i})$, while fine-scale errors are associated with the neighborhood function $s_i$ (Urban et al. 2002). The conventional way to fit the neighborhood function $s_i$ is to first detrend the data by removing broader-scale forcings. Broad-scale trends in the residuals imply an unaccounted constraint expressed spatially (e.g. a gradient) or an unaccounted spatial process (e.g. advection by prevailing winds). Such errors can be remedied by including the latent variable(s) in a spatially implicit model. Fine-scale errors (i.e. autocorrelated residuals) also suggest unaccounted constraints or processes; these errors can be parameterized implicitly or accounted for explicitly with a neighborhood function.

A regression framework can provide useful guidelines to decide when a model "needs" to include spatially explicit neighborhood functions. Consider the case where the model includes a neighborhood function expressed in terms of a location "1 step away" from the location of interest: $z_i = f(\underline{x_i}) + g(z_{[i+1]}) + \varepsilon$. The regression may not detect the effects of neighbors (g) for two reasons: (a) the data are adequate but the neighborhood effect is weak; or (b) the neighborhood effect is real, but the data are inadequate to estimate it. In both cases, the fit is statistically non-significant. The former case rejects a spatial model, in effect arguing that there is no "error of omission" while the latter case rejects the spatial model because of the "error of commission" implied by the poor fit. In either case, standard model-fitting and evaluation techniques inform the decision of the sufficient level of model complexity.

## Prospectus

Ecologists have access to a variety of techniques for reducing sources of error or uncertainty into model projections, including bootstrapping methods (appropriate for measurement error) and Monte Carlo methods for propagating estimation and process error. However, a general framework for correcting four sources of error at once is not available, although some approaches such as hierarchical Bayesian Markov chain Monte Carlo methods appear promising (Wikle et al. 1998, Clark 2003). Here we reinforce previously noted cautions in extrapolating from fine to broad scales: use the simplest possible model that an application will allow. In addition, we emphasize the importance of being forthright in admitting the potential for error associated with each extrapolation approach, and in particular the error propagation that spatially explicit models present.

## References

Archer, S. 1989. Have southern Texas savannas been converted to woodlands in recent history? – Am. Nat. 134: 545–561.

Bachelet, D., Neilson, R. P., Lenihan, J. M. et al. 2001. Climate change effects on vegetation distribution and carbon budget in the United States. – Ecosystems 4: 164–185.

Baker, W. L. 1989. A review of models of landscape change. – Landscape Ecol. 2: 111–133.

Band, L. E., Tague, C. L., Groffman, P. et al. 2001. Forest ecosystem processes at the watershed scale: hydrological and ecological controls of nitrogen export. – Hydrol. Processes 15: 2013–2028.

Bascompte, J. and Sole., R. V. 1996. Habitat fragmentation and extinction thresholds in spatially explicit models. – J. Anim. Ecol. 65: 465–473.

Beeson, P. C., Martens, S. N. and Breshears, D. D. 2001. Simulating overland flow following wildfire: mapping vulnerability to landscape disturbance. Special issue: wildfire and surficial processes. – Hydrol. Processes 15: 2917–2930.

Beven, K. J. and Kirkby, M. J. 1979. A physically based, variable contributing area model of basin hydrology. – Hydrol. Sci. Bull. 24: 43–69.

Bevers, M. and Flather, C. H. 1999. Numerically exploring habitat fragmentation effects on populations using cell-based coupled map lattices. – Theor. Popul. Biol. 55: 61–76.

Borcard, D., Legendre, P. and Drapeau, P. 1992. Partialling out the spatial component of ecological variation. – Ecology 73: 1045–1055.

Bormann, F. H. and Likens, G. E. 1979. Pattern and process in a forested ecosystem. – Springer-Verlag.

Botkin, D. B. 1993. Forest dynamics: an ecological model. – Oxford Univ. Press.

Burke, I. C., Kittel, T. G. F., Lauenroth, W. K. et al. 1991. Regional analysis of the Central Great Plains. – BioScience 41: 685–692.

Burke, I. C., Lauenroth, W. K. and Parton, W. J. 1997. Regional and temporal variation in net primary production and nitrogen mineralization in grasslands. – Ecology 78: 1330–1340.

Clark, J. S. 2003. Uncertainty and variability in demography and population growth: a hierarchical approach. – Ecology 84: 1370–1381.

Clark, J. S., Macklin, E. and Wood, L. 1998. Stages and spatial scales of recruitment limitation in southern Appalachian forests. – Ecol. Monogr. 68: 213–235.

Clark, J. S., Carpenter, S. R., Barber, M. et al. 2001. Ecological forecasts: an emerging imperative. – Science 293: 657–660.

Coffin, D. P. and Lauenroth, W. K. 1994. Successional dynamics of a semiarid grassland: effects of soil texture and disturbance size. – Vegetatio 110: 67–82.

Cruickshank, I., Gurney, W. S. C. and Veitch, A. R. 1999. The characteristics of epidemics and invasions with thresholds. – Theor. Popul. Biol. 56: 279–292.

Dale, V. H., Pearson, S. M., Offerman, H. L. et al. 1994. Relating patterns of land-use change to faunal biodiversity in the central Amazon. – Conserv. Biol. 8: 1027–1036.

Dale, V. H., Joyce, L.A., McNulty, S. et al. 2001. Climate change and forest disturbances. – Bioscience 51: 723–734.

Davenport, D. W., Breshears, D. D., Wilcox, B. P. et al. 1998. Sustainability of piñon-juniper woodlands – a unifying perspective of soil erosion thresholds. Viewpoint. – J. Range Manage. 51: 231–240.

Deutschman, D. H., Levin, S. A. and Pacala, S. W. 1999. Error propagation in a forest succession model: the role of fine-scale heterogeneity in light. – Ecology 80: 1927–1943.

Draper, N. R. and Smith, H. 1966. Applied regression analysis. – John Wiley and Sons, Inc.

Eastman, J. L., Coughenour, M. B. and Pielke, R. A. 2001. The regional effects of $CO_2$ and landscape change using a coupled plant and meteorological model. – Global Change Biol. 7: 797–815.

FEMAT. 1993. Forest ecosystem management: an ecological, economic, and social assessment. Report of Forest Ecosystem management Assessment Team, Portland, OR, USA.

Gardner, R. H., and Urban, D. L. 2003. Model validation and testing: Past lessons, present concerns, future prospects. – In: Canham, C. D., Cole, J. C. and Lauenroth, W. K. (eds), Models in ecosystem science. Princeton Univ. Press.

Gardner, R. H., O'Neill, R. V., Mankin, J. B. et al. 1980. Comparative error analysis of six predator–prey models. – Ecology 61: 323–332.

Gardner, R. H., O'Neill, R. V., Mankin, J. B. et al. 1981. A comparison of sensitivity analysis and error analysis based on a stream ecosystem model. – Ecol. Model. 12: 177–194.

Gardner, R. H., Cale, W. G. and O'Neill, R. V. 1982. Robust analysis of aggregation error. – Ecology 63: 1771–1779.

Gardner, R. H., Milne, B. T., Turner, M. G. et al. 1987. Neutral models for the analysis of broad-scale landscape pattern. – Landscape Ecol. 1: 19–28.

Grassberger, P. 1983. On the critical behavior of the general epidemic process and dynamical percolation. – Math. Biosci. 63: 157–172.

Hakanson, L. 1999. On the principles and factors determining the predictive success of ecosystem models, with a focus on lake eutrophication models. – Ecol. Model. 121: 139–160.

Hansen, A. J., Garman, S. L., Marks, B. et al. 1993. An approach for managing vertebrate diversity across multiple-use landscapes. – Ecol. Appl. 3: 481–496.

Heuvelink, G. B. M. 1998. Uncertainty analysis in environmental modelling under a change of spatial scale. – Nutrient Cycl. Agroeco. 50: 255–264.

Heyerdahl, E. K., Brubaker, L. B. and Agee, J. K. 2001. Spatial controls of historical fire regimes: a multiscale example from the interior west, USA. – Ecology 82: 660–678.

Hogeweg, P. 1988. Cellular automata as a paradigm for ecological modeling. Appl. Math. Comp. 27:8 1–100.

Keitt, T. H., Bjornstad, O., Dixon, P. et al. 2002. Accounting for spatial pattern when modeling organism-environment interactions. – Ecography 25: 616–625.

King, A. 1991. Translating models across scales in the landscape. – In: Turner, M. G. and Gardner, R. H. (eds), Quantitative methods in landscape ecology. Springer-Verlag, pp. 479-517.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm. – Ecology 74: 1659–1673.

Legendre, P., Troussellier, M., Jarry, V. et al. 1989. Design for simultaneous sampling of ecological variables: from concepts to numerical solutions. – Oikos 55: 30–42.

Levin, S. A. 1992. The problem of pattern and scale in ecology. – Ecology 73: 1943–1967.

Levin, S. A. and Durrett, R. 1996. From individuals to epidemics. – Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci. 351: 1615–1621.

Lieth, H. and Whittaker, R. H. (eds). 1975. The primary productivity of the biosphere. Springer.

Lu, L., Pielke, R. A., Liston, G. E. et al. 2001. Implementation of a two-way interactive atmospheric and ecological model and its application to the central United States. – J. Climate 14: 900–919.

Ludwig, D., Jones, D. D. and Holling, C. S. 1978. Qualitative analysis of insect outbreak systems: the spruce budworm and forest. – J. Anim. Ecol. 47: 315–332.

MacArthur, R. H. and Wilson, E. O. 1967. The theory of island biogeography. – Princeton Univ. Press.

Maddox, J. 1992. Forest fires, sandpiles and the like. – Nature 359: 35.

Marcot, B. G. and Thomas, J. W. 1997. Of spotted owls, old growth, and new policies: a history since the Interagency Scientific Committee report. – USDA Forest Service General Technical Report PNWGTR-408. PNW Forest and Range Experiment Station, Portland, OR, USA.

McKelvey, K., Noon, B. R. and Lamberson, R. H. 1993. Conservation planning for species occupying fragmented landscapes: the case of the northern spotted owl. – In: Kareiva, P., Kingsolver, J. G. and Huey, R. B. (eds), Biotic interactions and global change. Sinauer, pp. 424–450.

Melillo, J. M., Borchers, J., Chaney, J. et al. 1995. Vegetation ecosystem modeling and analysis project-comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate-change and $CO_2$ doubling. – Global Biogeochem. Cycles 9: 407–437.

Miller, C. and Urban, D. L. 1999. Interactions between forest heterogeneity and surface fire regimes in the southern Sierra Nevada. – Can. J. For. Res.- Rev. Canad. De Recherche Forest. 29: 202–212.

Miller, C. and Urban, D. L. 2000. Connectivity of forest fuels and surface fire regimes. – Landscape Ecol. 15: 145–154.

Neilson, R. P. and Running, S. W. 1996. Global dynamic vegetation modelling: coupling biogeochemistry and biogeography models. – In: Walker, B. and Steffen, W. (eds), Global change and terrestrial ecosystems. Cambridge Univ. Press, pp. 451–465.

O'Neill, R. V. 1973. Error analysis of ecological models. – In: Nelson, D. J. (ed.), Radionuclides in ecosystems. National

Technical Information Service, Springfield, Virginia, USA, pp. 898–908.

O'Neill, R. V. 1979. Natural variability as a source of error in model predictions. – In: Innis, G. S. and O'Neill, R. V. (eds), Systems analysis of ecosystems. International Co-operative, Fairland, Maryland, USA, pp. 28–32.

Pacala, S. W. and Silander, A. J. Jr. 1990. Tests of neighborhood population dynamic models in field communities of two annual weed species. – Ecol. Monogr. 60: 113–134.

Parton, W. J., Schimel, D. S., Cole, C. V. et al. 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. – Soil Sci. Soc. Am. J. 51: 1173–1179.

Parysow, P., Gertner, G. and Westervelt, J. 2000. Efficient approximation for building error budgets for process models. – Ecol. Model. 135: 111–125.

Pastor, J. and Post, W. M. 1986. Influence of climate, soils moisture, and succession on forest carbon and nitrogen cycles. – Biogeochem. 2: 3–27.

Peters, D. P. C. 2002. Plant species dominance at a grassland-shrubland ecotone: an individual-based gap dynamics model of herbaceous and woody species. – Ecol. Model. 152: 5–32.

Pielke, R. A. 2002. Overlooked issues in the U.S. National Climate and IPCC assessments. – Clim. Change 52: 1–11.

Pimentel, D. and Skidmore, E. L. 1999. Rates of soil erosion. – Science 286: 1477–1477.

Pitelka, L. F., Gardner, R. H., Ash, J. et al. 1997. Plant migration and climate change. – Am. Sci. 85: 464–473.

Reynolds, J. F. and Acock, B. 1985. Predicting the response of plants to increasing carbon dioxide: a critique of plant growth models. – Ecol. Model. 29: 107–129.

Romme, W. H., Everham, E. H., Frelich, L. E. et al. 1998. Are large, infrequent disturbances qualitatively different from small, frequent disturbances? – Ecosystems 1: 524–534.

Rosenblatt, A. E., Gold, A. J., Stolt, M. H. et al. 2001. Identifying riparian sinks for watershed nitrate using soil surveys. – J. Environ. Qual. 30: 1596–1604.

Rossi, R. E., Mulla, D. J., Journel, A. G. et al. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. – Ecol. Monogr. 62: 277–314.

Shugart, H. H. 1984. A theory of forest dynamics: the ecological implications of forest succession models. – Springer Verlag.

Solomon, A. M., Prentice, I. C., Leemans, R. et al. 1993. The interaction of climate and land-use in future terrestrial carbon storage and release. Water Air Soil Pollut. 70: 595–614.

Temple, S. A. 1986. Predicting impacts of fragmentation on forest birds: a comparison of two models. – In: Verner, J., Morrison, M. L. and Ralph, C. J (eds), Modeling habitat relationships of terrestrial vertebrates. Univ. Wisconsin Press, pp. 301–304.

Thomas, J. W., Forsma, E. D., Lint, J. B. et al. 1990. A conservation strategy for the northern spotted owl. – Report to the Interagency Scientific Committee to address the conservation of the northern spotted owl. Portland, OR, USA.

Tilman, D. 1994. Competition and biodiversity in spatially structured habitats. – Ecology 75: 2–16.

Turner, M. G., Gardner, R. H., Dale, V. H. et al. 1989. Predicting the spread of disturbance across heterogeneous landscapes. – Oikos 55: 121–129.

Urban, D., Goslee, S., Pierce, K. et al. 2002. Extending community ecology to landscapes. – Ecoscience 9: 200–212.

Wessman, C. A. 1992. Spatial scales and global change: bridging the gap from plots to GCM grid cells. – Annu. Rev. Ecol. Syst. 23: 175–200.

Whitcomb, R. F., Lynch, J. F., Klimkiewicz, M. K. et al. 1981. Effects of forest fragmentation on avifauna of the eastern deciduous forest. – In: Burgess, R. L. and Sharpe, D. M. (eds), Forest island dynamics in man-dominated landscapes. Springer, pp. 125–205.

Wikle, C. K., Berliner, L. M. and Cressie, N. 1998. Hierarchical Bayesian space-time models. – Environ. Ecol. Stat. 5: 117–154.

With, K. A., Gardner, R. H. and Turner, M. G. 1997. Landscape connectivity and population distributions in heterogeneous environments. – Oikos 78: 151–169.