CHAPTER 8

Structural equation modeling: building and evaluating causal models

James B. Grace, Samuel M. Scheiner, and Donald R. Schoolmaster, Jr.

8.1 Introduction to causal hypotheses

Scientists frequently wish to study hypotheses about *causal* relationships, but how should we approach this ambitious task? In this chapter we describe structural equation modeling (SEM), a general modeling framework for the study of causal hypotheses. Our goals will be to (a) concisely describe the methodology, (b) illustrate its utility for investigating ecological systems, and (c) provide guidance for its application. Throughout our presentation, we rely on a study of the effects of human activities on wetland ecosystems to make our description of methodology more tangible. We begin by presenting the fundamental principles of SEM, including both its distinguishing characteristics and the requirements for modeling hypotheses about causal networks. We then illustrate SEM procedures and offer guidelines for conducting SEM analyses. Our focus in this presentation is on basic modeling objectives and core techniques. Pointers to additional modeling options are also given.

8.1.1 *The need for SEM*

Consider a task faced by the US National Park Service (NPS), the monitoring of natural resources. For documenting conditions, they can use conventional statistical methods to quantify properties of the parks' ecosystems and track changes over time. However, the NPS is also charged with protecting and restoring natural resources. This second task requires understanding cause–effect relationships such as ascribing changes in the conditions of natural resources to particular human activities. Causal understanding is central to the prevention of future impacts by, for example, halting certain human activities. Active restoration through effective intervention carries with it strong causal assumptions. These fundamental scientific aspirations—understanding how systems work, predicting future behaviors, intervening on current circumstances—all involve causal modeling, which is most comprehensively conducted using SEM.

Ecological Statistics: Contemporary Theory and Application. First Edition. Edited by Gordon A. Fox, Simoneta Negrete-Yankelevich, and Vinicio J. Sosa.
Published in 2015 by Oxford University Press.

Modeling cause–effect relationships requires additional caveats beyond those involved in the characterization of statistical associations. For the evaluation of causal hypotheses, biologists have historically relied on experimental studies. SEM allows us to utilize experimental and observational data for evaluating causal hypotheses, adding value to the analysis of both types of information (e.g., Grace et al. 2009). While experimental studies provide the greatest rigor for testing individual cause–effect assumptions, in a great many situations experiments that match the phenomena of interest are not practical. Under these conditions, it is possible to rely on reasonable assumptions built on prior knowledge to propose models that represent causal hypotheses. SEM procedures can then be used to judge model–data consistency and rule out models whose testable implications do not match the patterns in the data. Whether one is relying on experimental or non-experimental data, SEM provides a comprehensive approach to studying complex causal hypotheses.

Learning about cause–effect relationships, as central as it is to science, brings with it some big challenges. We emphasize that confidence in causal understanding generally requires a sequential process that develops and tests ideas. SEM, through both its philosophy and procedures, is designed for such a sequential learning process.

Beyond testing causal hypotheses, we are also interested in estimating the magnitudes of causal effects. Just because we have properly captured causal relationships qualitatively (A does indeed affect C through B) does not guarantee arriving at unbiased and usable estimates of the magnitudes of causal effects. Several conditions can contribute to bias, including imperfect temporal consistency, partial confounding, and measurement error. The consequences of such biases depend on their context. In many ecological studies, SEM analyses are aimed at discovering the significant connections in a hypothesized network. In such studies, the relative strengths of paths are the basis for scientific conclusions about network structure. In other fields such as medicine or epidemiology, often the focus of a causal analysis may be on a single functional relationship that will be used to establish regulatory guidelines or recommended treatments (e.g., isolating the magnitude of causal effect of a drug on the progress of an illness). The general requirements for SEM are the same in both situations, but the priorities for suitable data for analyses and levels of acceptable bias may differ. Investigators should be aware of these distinctions and strive to obtain data suitable to their study priorities.

8.1.2 *An ecological example*

In this chapter, we use data from Acadia National Park, located on the coast of Maine (USA) for illustration. At Acadia, wetlands are one of the priority ecosystem types designated for protection. For these ecosystems, both resource managers and scientists wish to know how things work, what kinds of changes (favorable or unfavorable) they can anticipate, and what active steps might prove useful to protect or restore the wetlands.

Acadia National Park is located on Mount Desert Island, a 24,000 ha granite bedrock island. As a consequence of its mountainous topography, wetlands on the island are in numerous small watersheds. The soils are shallow in the uplands while the wetlands are often peat-forming. Additionally, they receive their water largely from acidic and low-nutrient inputs of rain and surface runoff, making them weakly-buffered systems (Kahl et al. 2000). For our illustrations, we use data from 37 nonforested wetlands recently studied by Little et al. (2010) and Grace et al. (2012) who examined the effects of human development on biological characteristics, hydrology, and water quality.

The studies measured various types of historical human activities in each wetland catchment area: (1) the intensity of human development in a watershed, (2) the degree of hydrologic alteration, (3) human intrusion into the buffer zone around wetlands, and (4) soil disturbance adjacent to wetlands. A human disturbance index (HDI) of the summed component measures was used to identify biological characteristics of plant communities that serve as bioindicators of human disturbance (figure 8.1a; Schoolmaster et al. 2012). Altered environmental conditions were also recorded, including water conductivity (as an indicator of nutrient loading) and hydroperiod (daily water depth). A subset of key biological characteristics was chosen to represent components of biotic integrity (Grace et al. 2012). We focus here on one biological property, native plant species richness, and its relationship with human disturbance as shown in figure 8.1b. For this example, we want to know how specific components of the human disturbance index might lower species richness and what might be done to reduce such impacts.

(a)

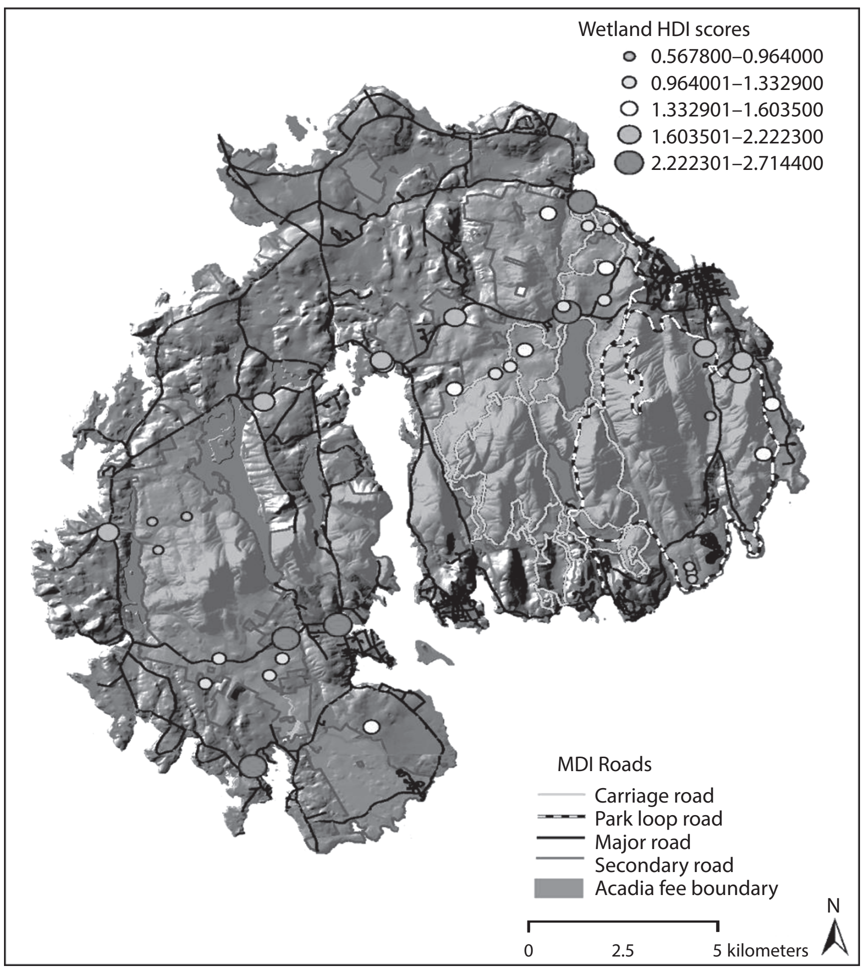

Fig. 8.1 (a) Map of Acadia National Park showing human disturbance index scores. (b) Native richness (species per plot) against land-use intensity scores.

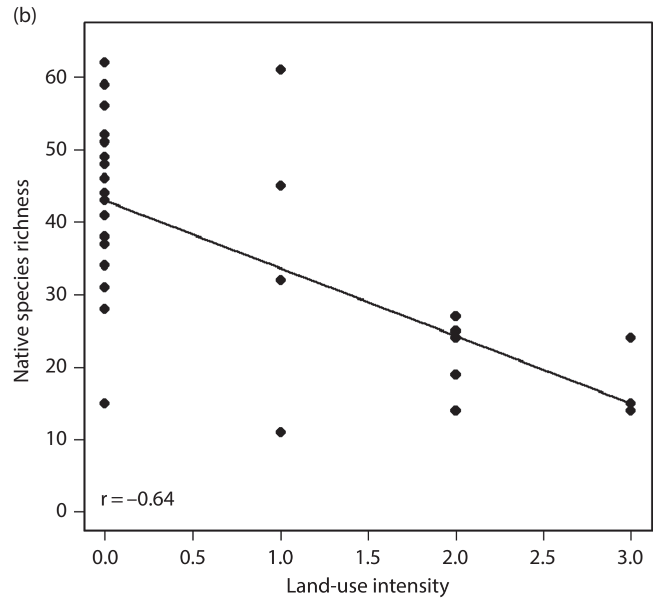

Fig. 8.1 (*continued*)

8.1.3 *A structural equation modeling perspective*

In our example, information from previous studies allows us to propose a hypothesis about how human activities and environmental alterations can lead to a loss of native species. In figure 8.2a we first represent our ideas in the form of a *causal diagram* (Pearl 2009) that ignores statistical details and focuses on hypothesized causal relationships. The purpose of a causal diagram is to (a) allow explicit consideration of causal reasoning, and (b) guide the development and interpretation of SE models. What we include in such a diagram is a function of our knowledge and the level of detail we wish to examine. Causal diagrams, distinct from structural equation models, are not limited by the available data.

Several causal assumptions implied in figure 8.2a are represented by directional arrows. The causal interpretation of *directed relationships* is that if we were to sufficiently manipulate a variable at the origin of an arrow, the variable at the head of the arrow would respond. In quantitative modeling, a relationship such as $Y = f(X)$ is assumed to be causal if an induced variation in X could lead to changes in the value of Y (see also the book Introduction). Generally, we must be able to defend a model's causal assumptions against the alternative that the direction of causation is the opposite of what is proposed or that relationships between variables are due to some additional variable(s) affecting both and producing a spurious correlation.

In this example, the assumptions expressed are: (1) Increasing land use in a watershed leads to more physical structures (ditches and dams) that control or alter hydrology. (2) Physical structures that influence hydrology can lead to changes in water-level variations (flooding duration). (3) Reduced water-level fluctuations (e.g., resulting from impoundment of wetlands) would create a plant community made up of the few highly flood-tolerant species. Collectively, these fundamental assumptions are only partially testable with observational data, since actual responses to physical manipulations are required to demonstrate causality unequivocally.

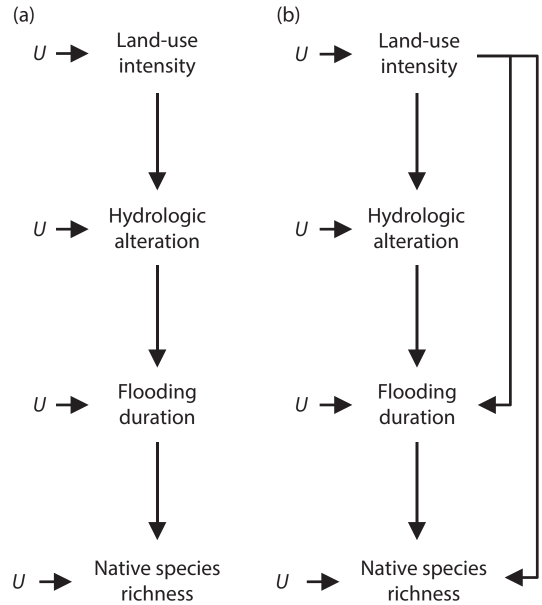

Fig. 8.2 (a) Simple causal diagram representing the hypothesis that there is a causal chain connecting land use to richness reduction through hydrologic alteration and subsequent impacts on flooding duration. The letter U refers to other unspecified forces. (b) Alternative diagram including additional mechanisms/links.

There are several ways data could be inconsistent with the general hypothesis in figure 8.2a. The direct effects encoded in the model might not be detectable. Also, the omitted linkages implied in the model might be inconsistent with the relations in the data. It is entirely possible that land use leads to changes in flooding duration, community flood tolerance, or native richness in ways not captured by the observed hydrologic alterations. Such additional omitted mechanisms (e.g., figure 8.2b) would result in residual correlations among variables not connected by a direct path. As explained later, these alternative possibilities are testable (i.e., these are "testable implications" of a model).

Generally, it is natural to think of cause–effect connections in systems as component parts of *causal networks* that represent the interconnected workings of those systems. *Structural equations* are those that estimate causal effects and a *structural equation model* is a collection of such equations used to represent a network (or portion thereof). Defined in this way, we think of causal networks as properties of systems, causal diagrams as general hypotheses about those networks, and structural equation models as a means of quantifying and evaluating hypotheses about networks. SEM originated as path analysis (Wright 1921); however, it has now evolved well beyond those original roots.

SEM represents an endeavor to learn about causal networks by posing hypotheses in the form of structural equation models and then evaluating those models against appropriate data. It is a process that involves both testing model structures and estimating model parameters. Thus, it is different from statistical procedures that assume a model structure is correct and only engage in estimating parameters. A key element of SEM is the use of graphical models to represent the causal logic implied by the equations (e.g., figure 8.2a).

SEM is a very general methodology that can be applied to nearly any type of natural (or human) system. At the end of the chapter we list and provide references for a few of the types of ecological problems that have been examined using SEM.

8.2 Background to structural equation modeling

The history of SEM and its mathematical details are beyond the limited space of this chapter, though a brief description of the equational underpinning of SE models is given in appendix 8.1. References to both general and specific topics related to this background are presented in the Discussion (section 8.4). Here we focus on fundamental principles related to the development and testing of causal modeling.

8.2.1 *Causal modeling and causal hypotheses*

Achieving a confident causal understanding of a system requires a series of studies that challenge and build on each other (e.g., Grace 2006, chapter 10). Any SEM application will have some assumptions that will not be explicitly tested in that analysis. Thus, SEM results will support or falsify some of the proposed ideas, while implying predictions that are in need of further testing for some of the other ideas. *SEM results should not be taken as proof of causal claims, but instead as evaluations or tests of models representing causal hypotheses.* With that qualifying statement in place, we can now ask, "What are the requirements for a causal/structural analysis?"

Structural equations are designed to estimate causal effects. We say "designed" to connote the fact that when we construct a SE model, we should be thinking in terms of cause–effect connections. More strictly, we are thinking about probabilistic dependencies as our means of representing causal connections. Careful causal thinking can be aided by first developing conceptual models and/or causal diagrams that focus on processes rather than just thinking about the available variables in hand. Each directed relationship in a causal model can be interpreted as an implied experiment. Each undirected relationship (e.g., double-headed arrow) connecting two variables is thought to be caused by some unmeasured entity affecting both variables. Further, in causal diagrams, the unspecified factors (U) (figure 8.2) that contribute to residual variances can be thought of as additional unmeasured causal agents (although they may also represent true, stochastic variation). Ultimately, our intent is to craft models that match, in some fashion, cause–effect relations. This is a more serious enterprise than simply searching for regression predictors. By our very intention of seeking causal relations, the onus is placed on scientists to justify causal assumptions. A strength of SEM is its requirement that we make these assumptions explicit.

The phrase "no causes in, no causes out" encapsulates the fact that there are certain assumptions embedded in our models that cannot be tested with the data being used to test the model. These untested assumptions include the directionality of causation. Such assumptions have to be defended based on theoretical knowledge; sometimes that is easy, sometimes it is more challenging. While links are not tested for directionality, we can still evaluate consistency in proposed direct, and *indirect effects*, as well as statements of *conditional independence*.

One point that is sometimes overlooked (and is commonly treated as implicit) is the assumption that causes precede effects. We should, for example (figure 8.2a), recognize that the plant diversity of today has been influenced by the flooding duration during some prior time period. Similarly, the flooding duration this year is influenced by hydrologic alterations made in prior years. It is not uncommon that the data may fail to strictly meet the precedence requirements desired for causal effect estimation. When proper temporal precedence does not hold, we must assume temporal consistency, meaning that current values of a predictor are correlated with values when the effect was generated. For example, we might only have data on flooding duration for a single year and have to assume that the variation among sites in duration was similar in past years. Such assumptions are not always reasonable. In such cases, one needs to develop dynamic SE models using time-course data (e.g., Larson and Grace 2004).

8.2.2 *Mediators, indirect effects, and conditional independence*

Arguably the most fundamental operation in SEM is the test of *mediation* (MacKinnon 2008). In this test we hypothesize that the reason one system property influences another can be explained by a third lying along the causal path. In our example (figure 8.2a), we hypothesize that one reason plant species richness is lower in areas with greater human land use is because of a series of processes involving hydrology that mediate/convey an effect. The ability to express causal chains and indirect effects is a distinguishing attribute of structural equation models (appendix 8.1). When we specify that flooding duration is influenced by land-use intensity through hydrologic alterations, we are making a causal proposition representing causal hypotheses that can be tested with observational data for model-data consistency. Tests of mediation are most powerful when a SEM analysis leads an investigator to conduct a follow-up study or to obtain additional measurements that permit possible mediators to be included in models.

Model–data consistency is critical for obtaining proper parameter estimates. First and foremost, variables not directly connected by a single or double-headed arrow in a model are presumed to exhibit conditional independence—that is, having no significant residual associations. Finding residual associations can suggest either an omitted direct, causal relationship or some unmeasured joint influence. Depending on model architecture, omitted links may result in biased estimates for some parameters. In figure 8.2a we pose the hypothesis that the effects of land use on flooding duration are due to hydrologic alterations. If we find that our data indicate that land use and flooding duration are not conditionally independent once we know the hydrologic alterations, either the intensity of land use influences flooding duration in ways unrelated to observable hydrologic alterations or an unmeasured process is causing the association. If land use and flooding duration are causally connected through two pathways (direct and indirect), both need to be included in the model to obtain unbiased estimates of effects along the causal chain (e.g., figure 8.2b).

Typically, after discovering a residual relationship (e.g., a significant correlation among residuals), we would revise our model either to include additional linkages or alter the structure of the model so as to resolve model–data discrepancies. It is critical that model revision be based on theoretical thinking and not simply by tinkering to improve model fit, otherwise our modeling enterprise is just a descriptive exercise rather than a test of a hypothesis or theory.

It can be helpful to know the minimum set of variables needed to be measured and modeled so as to properly specify a model, especially if one is working from a causal diagram. A general graphical-modeling solution to this problem, the *d-separation criterion*, has been developed by Pearl (1988). We omit describing this somewhat intricate concept and instead refer the reader to a more complete treatment in Pearl (2009).

8.2.3 *A key causal assumption: lack of confounding*

A classic problem in causal modeling is to avoid *confounding* (see chapter 7). Confounding occurs when variables in a model are jointly influenced by variables omitted from the model. Identifying and including the omitted variables can solve this problem, as represented in figure 8.3a. Here there is some factor U′ that jointly influences both intensity of land use and hydrologic alterations. If we are unaware of such an influence and estimate effects using a SE model that treats the two variables as independent, then, the directionality of linkages may still be causally correct, but our parameter estimate linking the two will be biased. An extended discussion of how confounding affects causal estimates can be found in Schoolmaster et al. (2013). Here, we consider only a single illustration.

Let us imagine a case where there is a planning process that determines which watersheds to develop and how many hydrologic alterations to install to support those potential developments. If planners assessed the topographic suitability for both land development and modifications of hydrology, we would need to include some measured variable to represent this decision process in our model if we are to avoid bias in our parameters. Figure 8.3b illustrates how we resolve this problem. By including a measured variable representing a planner's perceived topographic favorability in our SE model, we block the "back-door connection" between intensity of land use and hydrologic alterations (for more on the back-door criterion see Pearl 2009).

8.2.4 *Statistical specifications*

There is a relationship between how models are specified and how their parameters can be estimated. Options for statistical specification (response distributions and link forms;

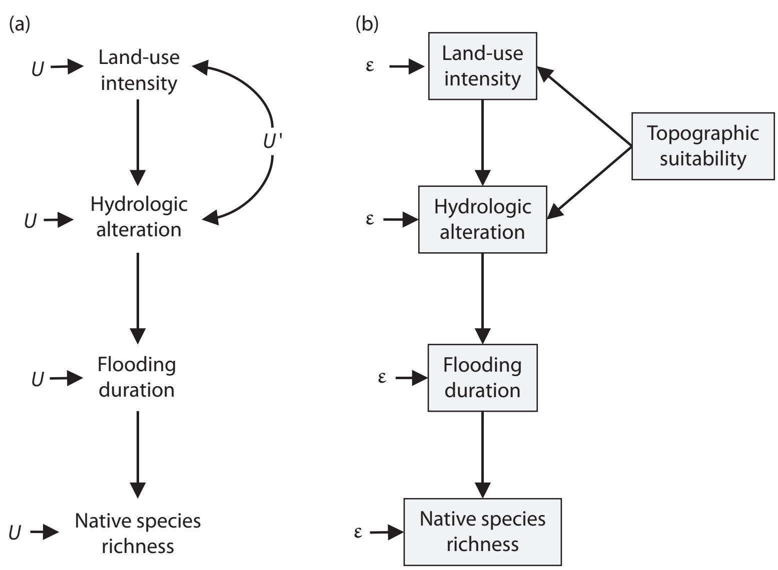

Fig. 8.3 (a) Causal diagram representing the case of a confounding effect by U′, an unmeasured factor that influences both ends of a causal chain, the effect of intensity of land use on hydrologic alteration. (b) A structural equation model that resolves the potential confounding by including measurements of the factor creating the confounding in (a).

e.g., Poisson responses with log-linear linkage) are well covered in conventional statistics textbooks and other chapters in this volume (see chapters 5, 6, 7, 13, and 14). In this chapter we provide a few examples of various response specifications, present some guidelines for the order in which specification choices might be considered (e.g., figure 8.5), and mention some of the criteria that may be used.

In any statistical model, including SE models, we must choose a probability distribution for each response (*endogenous*) variable and the form of the equation for relating predictors to responses. It is common to assume linear relationships with Gaussian-distributed independent errors, but we are not restricted to this assumption and a SE model can include any form for a particular causal relationship, including logistic, quadratic, and binary. Of course, for any functional form one must be cognizant of the statistical assumptions involved related to the data, model specifications, and estimation methods. The choices of model specification and estimation methods will depend on both the form of the data and the questions being asked. Each method has its array of specific assumptions and potential hazards and limitations, a topic too vast to cover in this chapter. We urge the reader to be cautious with any analysis, but especially when using unfamiliar procedures.

8.2.5 *Estimation options: global and local approaches*

There are two general approaches to parameter estimation in SEM, a single global approach that optimizes solutions across the entire model and a local-estimation approach that separately optimizes the solutions for each endogenous variable as a function of its predictors (figure 8.4). Much of the focus in SEM in the past few decades has been on *global estimation*, where data–model relationships for the entire model are summarized in terms of variance–covariance matrices (upper analysis route in figure 8.4). Maximum likelihood procedures (see chapter 3) are typically used to arrive at parameter estimates by minimizing the total deviation between observed and model-implied covariances in the whole model. Sometimes, alternatives to maximum likelihood, such as two-stage least squares, are used (Lei and Wu 2012).

Maximum likelihood global estimation typically relies on fitting functions such as

$$F_{ML} = \log|\hat{\Sigma}| + tr(S\hat{\Sigma}^{-1}) - \log|S| - (p+q). \tag{8.1}$$

Here, F_{ML} is the maximum likelihood fitting function, $\hat{\Sigma}$ is the model-implied covariance matrix, S is the observed covariance matrix, while $(p+q)$ represents the sum of the exogenous and endogenous variables. For a discussion of the statistical assumptions associated with estimation methods used in SEM, refer to Kline (2012). Global analyses have historically not used the original data, but instead only the means, variances, and covariances that summarize those data. This simplification allows for the estimation of a tremendous variety of types of models, including those involving latent (unmeasured) variables, correlated errors, and nonrecursive relations such as causal loops and feedbacks (Jöreskog 1973). (In appendix 8.2 we illustrate a simple application of this type of analysis.)

The alternative to global estimation is *local estimation*, estimating parameters for separable pieces of a model (figure 8.4, lower analysis route). Modern approaches to local estimation are implemented under a graphical modeling perspective (Grace et al. 2012). Consider the causal diagram in figure 8.2b. In graphical models we often talk about nodes and their links. The graph here has four nodes and five links; nodes that are directly linked are said to be adjacent. Causal relations within the graph can be described using familial terminology; adjacent nodes have a *parent-child relationship*, with parents having causal

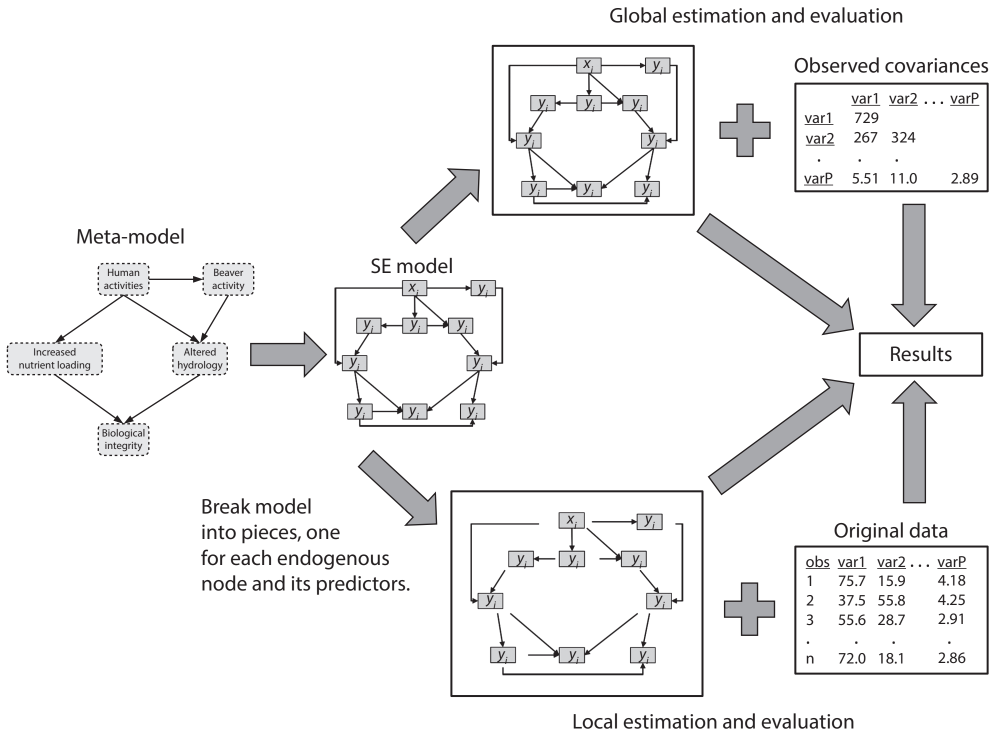

Fig. 8.4 Comparison of global- to local-estimation procedures. Starting with a meta-model based on a priori ideas, an SE model is defined. SE models can be analyzed either under a global-solution framework or through piecewise estimation of local relationships. While analytical procedures differ, both approaches represent implementations of the SEM paradigm.

effects on children. Three of the four nodes are endogenous because they have parents within the diagram; land-use intensity is *exogenous* because it does not have a parent. The node for hydrologic alteration has one parent (land-use intensity), while the nodes for flooding duration and native species richness both have two parents. While hydrologic alteration is an ancestor of native species richness, it is not a parent because there is no direct linkage. So, for the SE model, we have four equations representing the four parent-child relationships. These equations are of the form, $y_i = f(pa_1 + pa_2 + \cdots + \varepsilon_i)$, with one equation for each child node in the diagram and where y_i is any response variable, pa_i refers to the parent variables for each response variable, and ε_i is the residual variation. A local solution approach involves estimating the parameters for each of those four equations separately. Once that is done, there needs to be a separate analysis (and confirmation) of the conditional independence assumptions before the estimates are to be trusted.

Local estimation is a useful alternative because it permits great latitude for the inclusion of complex specifications of responses and linkages. It is also potentially advantageous because it avoids propagating errors from one part of a model to another, which can happen with global-estimation methods. Further, Bayesian estimation procedures optimize parameters locally in most cases, seeking optimal solutions for individual equations rather than the overall model. Bayesian estimation of SE models is increasingly popular (Lee 2007; Congdon 2010; Song and Lee 2012; for ecological applications see Arhonditsis et al. 2006; Grace et al. 2011, 2012). Here we present the local-estimation approach as an umbrella that permits a wide variety of statistical estimation philosophies, including Bayesian, likelihood, and frequentist methods (see chapter 1).

Despite philosophical preferences one may have for global-estimation versus local-estimation approaches, practical considerations are of overwhelming importance when considering the options, as we illustrate in section 8.3. Without question, the capabilities of available software are an important consideration and both software and instructional materials supporting SEM are continuously evolving. In the next section we provide further guidance for the choice of estimation method and how it relates to both model specification and modeling objectives.

8.2.6 *Model evaluation, comparison, and selection*

Few problems in statistics have received more attention than the issue of how models are critiqued, evaluated, and compared. This can ultimately be viewed in the context of a decision problem. The question is, "What variables should I leave in my model?" or, alternatively, "Which of the possible models should I select, based on the data available?" For models that represent *causal networks*, the question is a bit different. Here we wish to know, "Are the linkages in a structural equation model consistent with the linkages in the causal network that generated the data?" In this situation there should be theoretical support for any model that is considered, as we are not shopping for some *parsimonious* set of predictors; instead, we are seeking models representing realistic causal interpretations.

An important consideration in causal modeling is that it combines theoretical a priori knowledge with the statistical analysis of data. We bring some context to this enterprise by distinguishing between SEM applications that are model-generating versus model-comparing versus confirmatory. These applications represent a gradient from situations where we have relatively weak confidence in our a priori preference for a particular model to situations where we have great confidence in our a priori model choice. The companion ingredient is our degree of confidence in the data. For example, if we have a very large and robust sample, we must give the data strong priority over our initial theoretical ideas. Conversely some data sets are relatively weak, and we may have greater confidence in our views about the underlying mechanisms. The extreme example of this theory-weighted case is in system simulation models where data are used only to estimate parameters, not to critique model structure. Therefore, when dealing with models containing causal content, context and judgment matter in arriving at final conclusions about processes. Of course, it is important that one clearly notes any difference between statistical results and any final conclusions derived from other considerations. There is a parallel here to the issue of using informed priors in Bayesian estimation. In some cases it is appropriate to weight posterior estimates based on prior information, but we must make clear what the new data say when we arrive at conclusions. (This theme, that one needs to use judgment and not blindly rely on statistical procedures, is also touched upon in the introductory chapter and chapters 2, 3, 5, and 7.)

While model performance and model support have different nuances—assessments of explanatory power versus relative likelihoods, respectively—in this treatment we do not emphasize this distinction. When evaluating network models, there are always both implicit and explicit comparisons. Further, evaluating model predictions, explanations, or residuals informs us about both our specification choices and also whether we have overspecified our models. Ultimately, in causal modeling there are many different kinds of examinations (including comparisons to previous or subsequent analyses) that contribute to model selection and the inferences drawn.

The classical approach to evaluating SE models using global-estimation methods is based on the function shown in equation (8.1) (see section 8.2.5). That function goes to zero when there is a perfect match between observed and model-implied covariances. In contrast, when evaluating SE models using local-estimation methods, one evaluates individual linkages. This process of testing the topology of the model, while compatible with global-estimation methods, is more general, applies to any network-type model, and is essential with local-estimation methods. The first step in local estimation is usually to determine whether there are missing connections in the model, such as testing for conditional independence. Each unlinked pair of variables can be tested for a significant omitted connection (Shipley 2013). Information-theoretic methods, such as the Akaike Information Criterion, are commonly used for model comparisons, both for the global-estimation and local-estimation cases. chapter 3 covers the theory behind AIC methods.

8.3 Illustration of structural equation modeling

8.3.1 *Overview of the modeling process*

In this section we provide general, practical guidelines for SEM and illustrate core techniques and their application using our ecological example. Grace et al. (2012) present an updated set of guidelines for the SEM process (figure 8.5), which we briefly describe here. First, be clear on the goals of your study (step 1). The specific goals and the focus of an

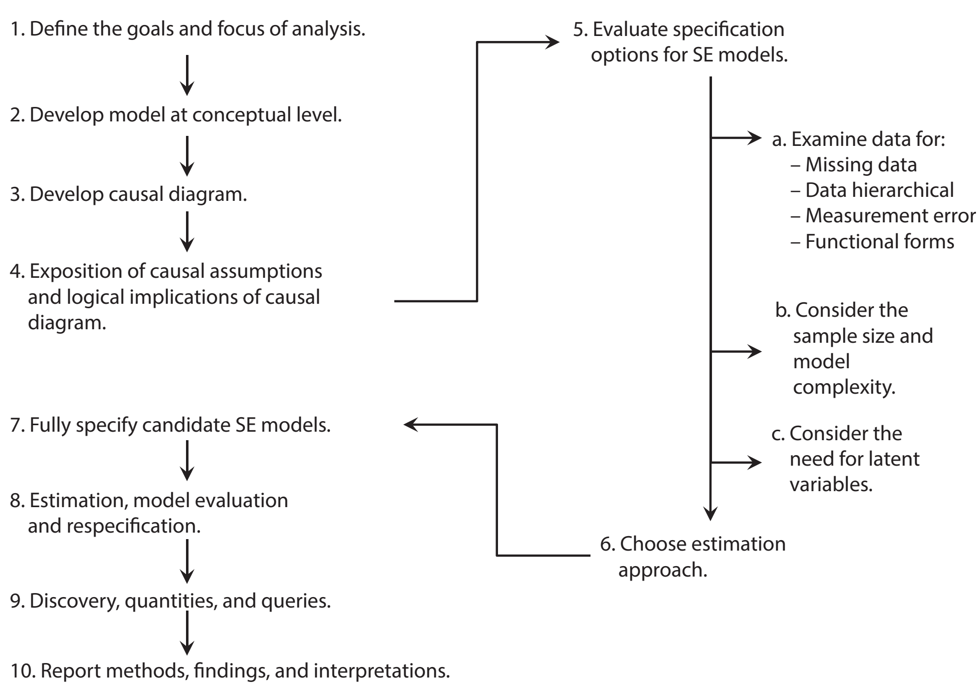

Fig. 8.5 Steps in the modeling process (from Grace et al. 2012).

analysis influence the data needed, model specifications, and estimation choices. Explicitly articulating the conceptual model (step 2), both verbally and graphically, is critical for conveying the logic that translates concepts into variables and ideas about processes into models made up of those variables (Grace et al. 2010). These goals can then be used to consider what is needed for drawing particular causal inferences as well as evaluating the testable *causal propositions* (steps 3 and 4).

A number of things need to be considered when developing a fully-specified model (steps 5–7). The characteristics of the data must be evaluated, both for the purpose of attending to data issues (Are there missing data? Do variables need transformation?) and for informing decisions about the equational representations (Are data hierarchical? Are non-linear relationships anticipated?). One must decide how complex to make the model. Model complexity is influenced by many factors, including objectives, hypothesis complexity, available measurements, number of samples, and the need for latent variables to represent important unmeasured factors. All of these choices influence the choice of estimation method, based on the criteria previously discussed comparing global versus local approaches. See Grace et al. (2010) for more background on model building.

For the next step (step 8), the processes of model estimation and model evaluation/comparison, it is ideal if there is a candidate set of models to compare. However, in SEM the issue of the overall fit of the data and model is of paramount importance. An omitted link is a claim of conditional independence between two unconnected variables, a claim that can be tested against the data. It is possible that all of the initially considered models are inconsistent with the data. In that case, you need to reconsider the theory underpinning the models and develop a revised hypothesis about the system, which can be subsequently evaluated. Once no missing links are indicated, the question of retaining all included links can be addressed. This is inherently a model comparison process. Only when a final suitable model is obtained are parameter estimates to be trusted. At that point parameter estimates, computed quantities, and queries of interest are summarized (step 9) and used to arrive at final interpretations (step 10).

8.3.2 *Conceptual models and causal diagrams*

The conceptual model for our ecological example (figure 8.6a) represents a general theoretical understanding of the major ways human activities impact wetland communities in this system. The conceptual model, termed a *structural equation meta-model* (SEMM; Grace and Bollen 2008; Grace et al. 2010), represents general expected dependencies among theoretical concepts. The SEMM provides a formal bridge between general knowledge and specific structural equation models, serving both as a guide for SE model development and as a basis for generalizing back from SEM results to our general understanding.

In the example, our general hypothesis (figure 8.6a) is that human activities primarily affect wetlands through changes in hydrology and water chemistry, especially elevated nutrient levels (Grace et al. 2012). Here we focus on two biological responses, cattail (*Typha*) abundance and native species richness (figure 8.6b). Cattails are invasive in this system and known to dominate under high-nutrient conditions. Thus, high cattail abundance is an undesirable state while high native species richness is a desirable state.

The causal diagram (figure 8.6b) is a statement of the processes operating behind the scene. This particular model does not portray the dynamic behavior of the system. Instead, it represents a static set of expectations appropriate to the data being analyzed. Given that simplification, we need to carefully consider the unmeasured (U) variables and associated processes if we are to avoid confounding (see section 8.2.3).

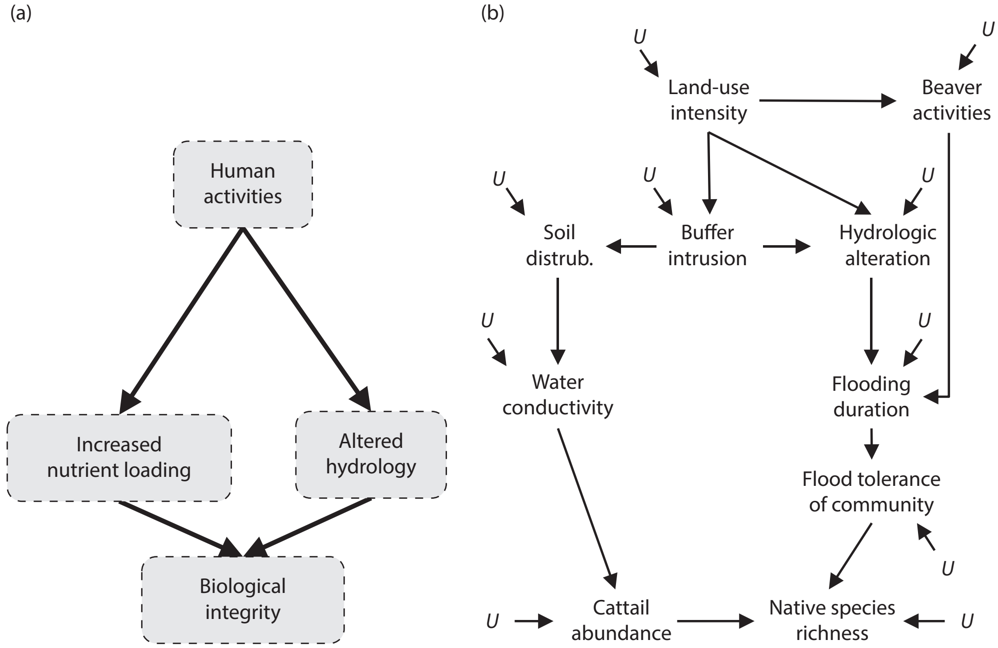

Fig. 8.6 (a) Meta-model representing general a priori theoretical understanding of how human activities most commonly affect wetland communities in cases such as this one. (b) Causal diagram representing a family of possible hypotheses about how specific activities might affect one particular component of integrity, native plant richness.

In this study, several environmental covariates were considered as possible confounders of relationships (distance from the coast, watershed size); ultimately none were considered to be sufficiently important for inclusion.

Causal diagrams can include variables that we did not measure such as hypothesized processes for which we have no direct measures (termed *latent variables*). A strength of SEM is the ability to include latent variables and evaluate their effects on observed quantities (Grace et al. 2010). Another alternative at our disposal is to absorb the effects of some variables in a *reduced-form model*. For example, we hypothesize that human activities may influence beaver populations and that the species pool for plants may be limited to flood tolerant species (figure 8.6b); however, we chose to not include beavers in our SE model (figure 8.7) because we lack appropriate data. Instead, the model has a direct link from intensity of land use to flooding duration to represent that process (thus, absorbing the node for beavers in the causal diagram). We also omit the variable “flood tolerance of community” even though data exist, because our sample size is small and we wish to keep our SE model as simple as possible. As all of this illustrates, the complexity of our SE models may be constrained for a variety of practical reasons.

8.3.3 *Classic global-estimation modeling*

In this section we demonstrate global-estimation approaches to model specification, estimation, and evaluation, including different implementations of SEM. We begin with a popular R library (`lavaan`; `latent variable analysis`) that implements SEM using maximum likelihood methods that seek global solutions. Further information about

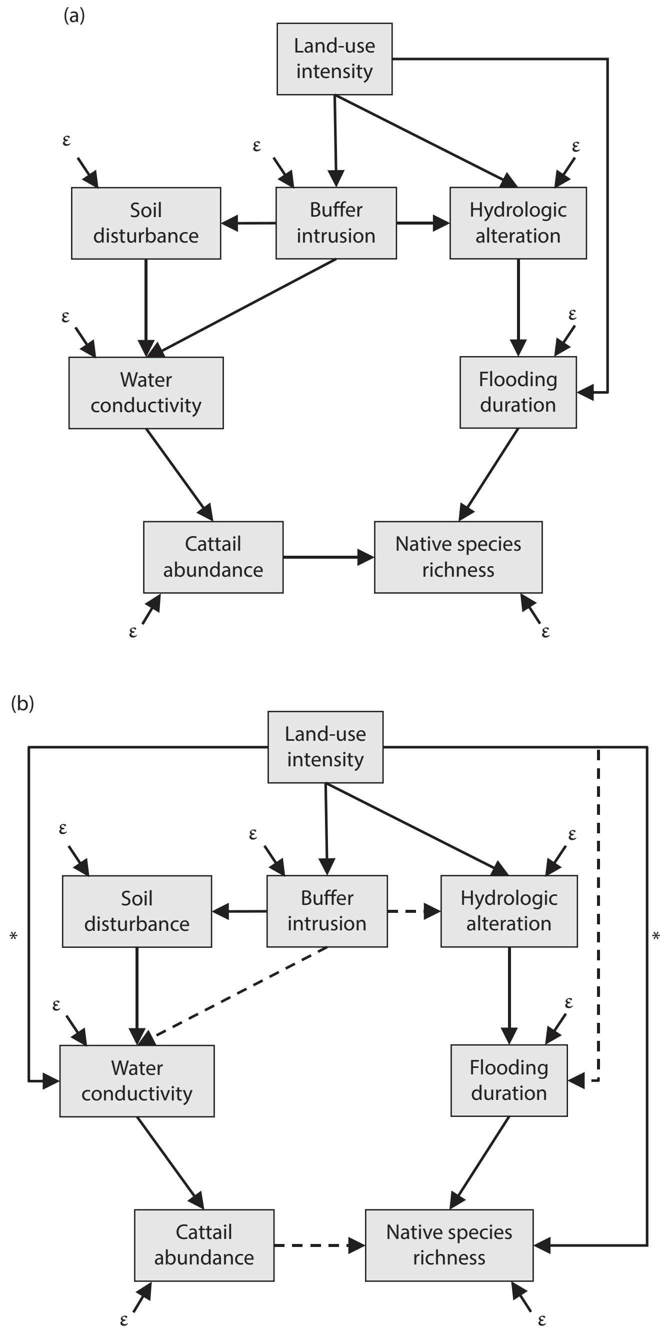

Fig. 8.7 (a) Initial structural equation model. (b) Revised model based on global analysis using `lavaan`. Paths with asterisks were added based on the discovery of non-independence. Paths represented by dashed lines were not supported (i.e., non-significant) based on the sample data.

`lavaan` can be found in Rosseel (2012, 2014). For simplicity of presentation, we assumed linear Gaussian relations throughout the model, the default setting of `lavaan`. We know that assuming Gaussian residuals is not appropriate for some of the variables in this model. For example, the variables representing human activities are all ordered categorical measurements. The `lavaan` library has an option for declaring ordered categorical variables that permits a more appropriate specification, though we do not use it in this demonstration for simplicity of presentation.

Once we begin the estimation process, a first task is to determine whether there are missing links that should be included for model–data consistency. When evaluating model fit, one should be aware that perfect fit automatically occurs when a model is saturated, i.e., there are as many parameters estimated as there are variances plus covariances. This is usually the situation when all possible links in a model are included. For any given model being evaluated, observed discrepancy is compared to a saturated model using a chi-squared statistic. The *model degrees of freedom* is the difference between the number of known covariances and the number of parameters estimated. The subsequent p-value represents the asymptotic probability that the data are consistent with the candidate model. Because in SEM our "default model" is our a priori theoretical model, not a null model, we use a logic that is the reverse from that used in null hypothesis testing. The hypothesized model is interpreted as being consistent with the data unless the p-value is small; in the case of a small p-value, we conclude that the data obtained are very unlikely given the model in hand. A chi-squared test is commonly used for evaluating overall fit and when comparing models differing by only a single link. However, when evaluating SE models we do more than use $p < 0.05$ for model rejection; instead, there are a number of different model fit assessment criteria. The literature relating to ways of assessing fit (and comparing SE models) is voluminous and well beyond what we can cover in this chapter; see Schermelleh-Engel et al. (2003) for further background. Ultimately, our goal is to detect and remedy any omitted associations, as their absence can substantially alter parameter estimates. In contrast, if a model includes unneeded links, their impacts on parameter estimates is generally small.

The `lavaan` code and some basic fit statistics for the first phase of analysis are shown in box 8.1, with the code presented in Part A. In this example, the very low p-value for the chi-squared test in Part B indicates a lack of fit. This lack of fit is reflected in the large residual covariances (differences between observed and model-implied covariances) shown in Part C. These discrepancies in turn are used to produce a set of *modification indices* that suggests ways of adding links to our model that would improve the fit (part D). These suggestions should not be used blindly, as some may make no scientific sense. The investigator must consider what plausible alternative hypotheses are worthy of consideration before re-estimating a new hypothesis.

In our model, several possible omitted linkages are suggested by the modification indices. Developing a revised model based on this kind of information can involve a bit of trial and error because modification indices are not perfect predictors of actual changes in model fit. So, some of the suggested modifications will reduce model–data discrepancy, but others will not. The reason for this paradoxical situation is that the raw material for the modification indices is the residual covariance matrix and large residuals can be created for a variety of indirect reasons. In our example, it appears that we should add a link between buffer intrusion and native species richness because that implied path had the largest modification index. However, there is no reason to think that buffer intrusion would have a direct effect. Ultimately, one must select theoretically supportable modifications and then make changes that seem most reasonable, continuing until a set of

Box 8.1 EXAMINING OVERALL GOODNESS OF FIT AND LOOKING FOR OMITTED LINKS IN INITIAL MODEL: R CODE AND SELECT RESULTS

```
# PART A: LAVAAN CODE
# creating data object for the analysis
semdat_8 <- data.frame ( landuse, buffer, hydro, flooding,
   richness, soil, cond, cattails)
# specify model
mod_8a <- ’buffer ~ landuse
            hydro ~ buffer + landuse
            flooding ~ hydro + landuse
            soil ~ buffer
            cond ~ soil + buffer
            cattails ~ cond
            richness ~ flooding + cattails’
fit_8a <- sem ( mod_8a,                # estimate model
   data = semdat_8)
summary (fit_8a, rsq = T,              # select results in Part B
   fit.measures = TRUE)
resid ( fit_8a,                        # select results in Part C
   type = "standardized")
modindices ( fit_8a)                   # select results in Part D


# PART B: INITIAL MODEL FIT RESULTS
lavaan (0.5-11) converged normally after 60 iterations
  Number of observations                          37
  Estimator                                       ML
  Minimum Function Test Statistic             36.776
  Degrees of freedom                              17
  P-value (Chi-square)                         0.004

# PART C: STANDARDIZED RESIDUAL COVARIANCES
          buffer hydro flooding soil cond cattails richness
landuse
buffer    0.000
hydro     0.000  0.000
flooding -2.037     NA  NA
soil      NA     0.466 -0.028  NA
cond      NA     0.863 -0.066  NA     NA
cattails  1.232  0.111  0.759  1.485  NA      NA
richness -1.502 -1.497 -0.303 -0.934 -1.214  -0.732  0.336
landuse   0.000  0.000   NA    0.211  1.311   1.237 -2.034  0.000
```

```
# PART D: SELECT MODIFICATION INDICES
Variable Pair          Implied Path          Modification Index
richness ~~ landuse    richness <-> landuse    8.941
richness ~ buffer      richness <- buffer     10.450
richness ~ landuse     richness <- landuse     8.941
richness ~ cond        richness <- cond        6.232
cond ~ landuse         cond <- landuse         5.513
```

defensible changes is obtained. Working through this we find that including links from land use to native richness and to conductivity (box 8.2, Part E) reduce model discrepancy to generally acceptable levels based on the chi-squared *p*-value (box 8.2, Part F), and in the process the other suggested modifications are resolved.

Model simplification, asking whether our model is parsimonious, is the next phase of evaluation. It turns out our revised model includes some links that may not be supported by the data (box 8.2, Part G *p*-values). One method for deciding whether a link actually

Box 8.2 ANALYSIS OF REVISED MODEL WITH LINKS ADDED: R CODE AND SELECT RESULTS

```
# PART E: LAVAAN CODE FOR MODEL WITH LINKS ADDED (added component
in bold)
mod_8b2 <- 'buffer ~ landuse
            hydro ~ buffer + landuse
            flooding ~ hydro + landuse
            soil ~ buffer
            cond ~ soil + buffer + landuse
            cattails ~ cond
            rich ~ flooding + cattails + landuse'
fit_8b2 <- sem ( mod_8b2, data = semdat_8)      # estimate model
# select results in Parts F and G
summary ( fit_8b2, rsq = T, fit.measures = TRUE)

# PART F: REVISED MODEL FIT
lavaan (0.5-11) converged normally after  67 iterations
  Number of observations                         37
  Estimator                                      ML
  Minimum Function Test Statistic            18.076
  Degrees of freedom                             15
  P-value (Chi-square)                        0.259
  RMSEA                                       0.074
  90 Percent Confidence Interval        0.000 0.180
  P-value RMSEA <= 0.05                       0.348
```

(*continued*)

Box 8.2 *(continued)*

```
# PART G: PARAMETER ESTIMATES
                   Estimate   Std.err   Z-value   P(>|z|)
buffer ~
    landuse         1.048      0.089     11.743    0.000
  hydro ~
    buffer          0.427      0.396      1.078    0.281
    landuse         0.966      0.468      2.065    0.039
  flooding ~
    hydro          29.854     11.784      2.533    0.011
    landuse        15.263     22.861      0.668    0.504
  soil ~
    buffer          0.211      0.040      5.289    0.000
  cond ~
    soil            0.179      0.102      1.748    0.081
    buffer          0.070      0.058      1.202    0.229
    landuse         0.163      0.064      2.549    0.011
  cattails ~
    cond            1.038      0.145      7.139    0.000
  rich ~
    flooding       -0.082      0.011     -7.571    0.000
    cattails        4.812      3.084      1.560    0.119
    landuse        -6.228      1.546     -4.029    0.000
```

can be removed is the *single-degree-of-freedom chi-squared test*, which is computed for two models that differ by only a single link/parameter. Either standard frequentist or likelihood ratio tests can be used to compare models (see chapter 1). A different approach that is preferred when comparing several alternative models is the use of information theory measures such as the Akaike Information Criterion (AIC). (See Burnham and Anderson (2002) and chapter 3 for more background on AIC.) Here we simply show the results (box 8.3) that led to our pruned final model (figure 8.7b).

8.3.4 *A graph-theoretic approach using local-estimation methods*

A graph-theoretic approach to SEM is non-parametric in the sense that the rules of causal modeling are compatible with any form of statistical specification (Pearl 2012; Grace et al. 2012). There is a great relaxation of restrictive assumptions that occurs if we can work with the original data instead of the derived covariance matrix as in the global-estimation approach. For our SE model (figure 8.7a) there are two aspects of specification we can now reconsider: (a) the distributions of variable responses, and (b) the form (linear or other) of relations between variables. Box 8.4 presents equations for our model that address these issues. Figure 8.8 shows the response distributions for the variables. For hypothesis testing, we are concerned about having residual variation that meets statistical assumptions. Often investigators using a global-estimation approach will use adjustment procedures to normalize residuals or will use resampling procedures to obtain bootstrapped parameter estimates. Local estimation permits us to do much more, as shown in box 8.4.

Box 8.3 MODEL SIMPLIFICATION: EXAMPLE R CODE AND SELECT RESULTS

```
# PART H: PRUNED MODEL ACCEPTED AS FINAL MODEL
mod_8b3 <- 'buffer ~ landuse
            hydro ~ landuse
            flooding ~ hydro
            soil ~ buffer
            cond ~ soil + landuse
            cattails ~ cond
            richness ~ flooding + landuse
            cattails ~ ~ 0 * richness'
fit_8b3 <- sem ( mod_8b3, data = semdat_8)
summary ( fit_8b3, rsq = T, fit.measures = TRUE)

# PART I: FINAL MODEL FIT
lavaan (0.5-11) converged normally after  66 iterations
  Number of observations                           37
  Estimator                                        ML
  Minimum Function Test Statistic              23.180
  Degrees of freedom                               19
  P-value (Chi-square)                          0.230
  RMSEA                                         0.077
  90 Percent Confidence Interval          0.000 0.171
  P-value RMSEA <= 0.05                         0.325

   PART J: PARAMETER ESTIMATES
                  Estimate  Std.err  Z-value  P(>|z|)
  buffer ~
     landuse         1.048    0.089   11.743   0.000
  hydro ~
     landuse         1.413    0.218    6.469   0.000
  flooding ~
     hydro          35.595    8.116    4.386   0.000
  soil ~
     buffer          0.211    0.040    5.289   0.000
  cond ~
     soil            0.222    0.097    2.289   0.022
     landuse         0.227    0.037    6.145   0.000
  cattails ~
     cond            1.038    0.145    7.133   0.000
  richness ~
     flooding       -0.080    0.011   -7.498   0.000
     landuse        -4.804    1.266   -3.795   0.000
```

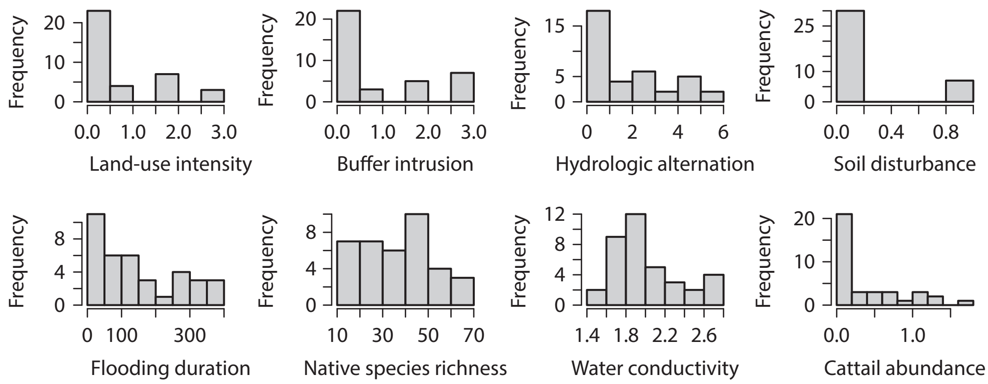

Fig. 8.8 Histograms for variables in SE model.

Box 8.4 INITIAL SPECIFICATIONS FOR LOCALLY ESTIMATED MODEL USING GLM AND LM FUNCTIONS IN R

```
# Buffer as a function of land-use intensity
glm_buffer <- glm ( buffer_prop ~ landuse, family = binomial)

# Hydrologic alteration as a function of land-use intensity
hydro_cat_pred <- ifelse ( landuse == 0, 0.739, 4.21)

# Soil disturbance as function of buffer intrusion
glm_soil <- glm ( soil ~ buffer, family = binomial)

# Flooding duration as function of hydrologic alteration and
   land-use

# intensity.
glm_flooding <- glm ( flooding_prop ~ landuse + hydro, family
  = binomial)

# Water conductivity as a function of soil disturbance and
    land-use intensity
lm_cond <- lm ( cond ~ soil + landuse)

# Native species richness as function of flooding duration and
   cattails
glm_richness <- glm ( richness ~ flooding + cattail,
   family = poisson)

# Cattail cover as change-point function of water conductivity
cattails_step1 <-lm ( cattail ~ cond)
cattails <- segmented ( cattails_step1, seg.Z = ~ cond, psi = 1.9)
```

One issue we address using local-estimation procedures relates to the fact that use of approximate methods for specifying response forms runs the risk of arriving at predicted scores that are not directly comparable to the data. For example, many of the variables in this model are bounded on one or both ends (e.g., species richness, percentage cover of cattails). To prevent predictions falling outside those bounds, particular distributions need to be specified for the responses (see chapters 3 and 6, and book appendix). Consider, for example, the relationships between land-use intensity and hydrologic alteration, and between water conductivity and cattails (figure 8.9). These data clearly do not fit linear models. When humans developed wetland areas, above some minimum value of land-use intensity they always put in structures (e.g., ditches, dams) to control the hydrology, resulting in a discrete (all or none) relationship (figure 8.9a). A more appropriate model representation is a two-level discrete response (figure 8.9c). For the case of cattails, they increase in abundance above some minimum threshold of water conductivity (figure 8.9b). This relationship can be specified with a change-point model (figure 8.9d) using local solution methods (box 8.4).

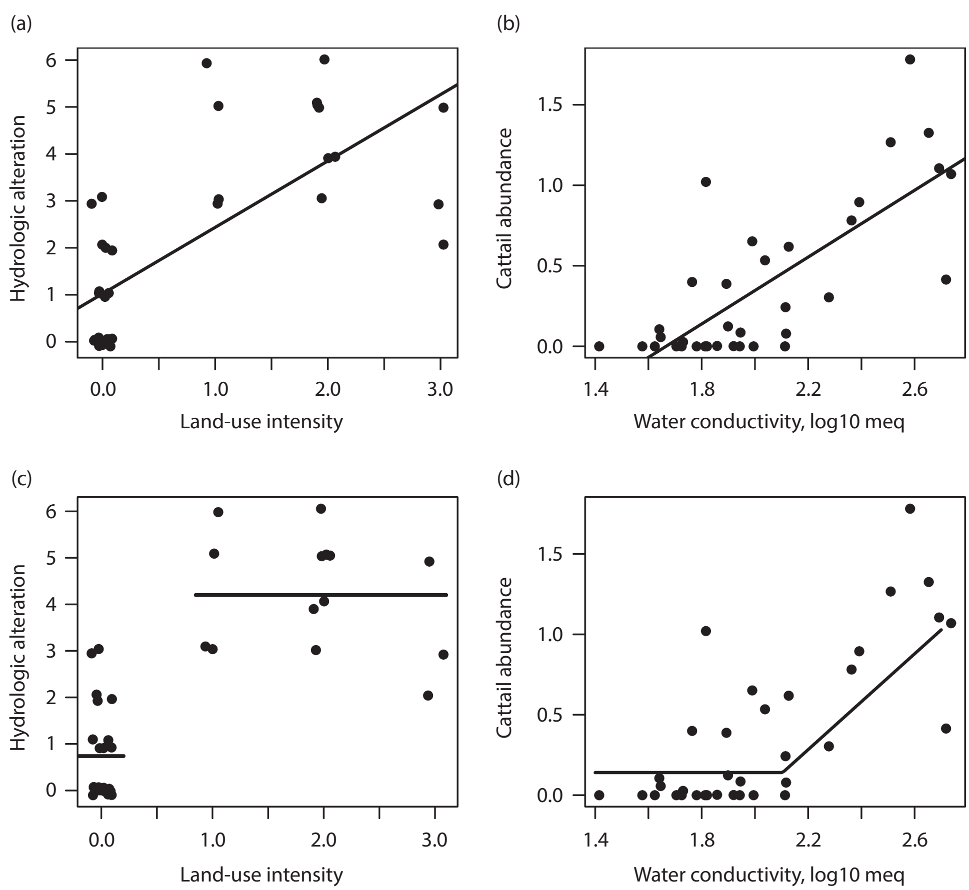

Fig. 8.9 (a) and (b) represent linear approximations of some key relationships, which are of the form typically supported by global-estimation methods (e.g., `lavaan`). (c) and (d) show more complex specifications of the same relationships as in (a) and (b), illustrating what can be included in locally estimated models.

Other response forms used in our model include a proportional odds specification (Agresti 2010) for the ordered categorical response of buffer intrusion (scored as one of four levels), and the proportion data of flooding duration (proportion of days flooded a year). For these variables, we used a logit link to represent the odds of observing maximum versus minimum values (box 8.4). Only two levels of soil disturbance were observed, so this was modeled as a Bernoulli outcome with a logit link (essentially a logistic regression). The degree of hydrologic alteration, although measured on a 6-point scale, behaved as a dichotomous response and was so modeled. Native species richness was modeled as a Poisson (count) variable while log water conductivity was treated as a Gaussian response. For details on the local-estimation procedures for these models, the reader should consult the appropriate R documentation. More justification of the forms used is in Grace et al. (2012). For more information on types of models, see chapter 6 and Bolker (2008).

Under a graph-theoretic approach for determining whether there are missing linkages, the key criterion is whether non-adjacent (unlinked) variables are conditionally independent. To determine whether a link between two variables should be added, we can use procedures that are illustrated in appendix 8.3. First, we obtained the residuals of the current model and then examined relationships among those residuals for unlinked variables. For variables with no predictors, the raw values were used in place of the residuals. Because we wished to consider all functional forms, we used both computational and graphical approaches. In our example, we detected residual associations (figure 8.10) that ultimately led us to a revised model (figure 8.11, box 8.5). This model is slightly different from that based on a global approach, specifically the inclusion of a link between buffer intrusion and native species richness, and the lack of a link between land-use intensity and richness (compare figures 8.7 and 8.10). The differences between the models arise from the use of linear Gaussian specifications in the global model, but more complex forms under local estimation (appendix 8.3).

8.3.5 *Making informed choices about model form and estimation method*

In practice, analyses are conducted by investigators with individual backgrounds, training, and scientific motivations. Analyses are also conducted with different software packages that have their own implementation of methods. Thus, one size does not fit all when it comes to SEM. There are two schools of thought as to which estimation method—global or local—is more appropriate. Pearl (2012) suggests that local estimation is more fundamental because it does not propagate errors caused by misspecification in one part of a model to other parts of a model. However, sometimes models are best seen as a single hypothesis. Consider psychology research where highly abstract concepts are represented exclusively using latent variables and the hypothesis is about how the latent machinery can cause the observed data patterns. In this case we can see an investigator preferring an estimation method that demands a simultaneous global solution. Biological problems with similar priorities might be expected for studies of behavioral ecology, life-history evolution, or organismal physiology. For example, Tonsor and Scheiner (2007) used SEM to relate physiological traits to traits involving morphology and life history, and ultimately to fitness. As the goal of that study was to investigate trait integration, a global-estimation procedure was necessary.

Global-estimation methods are inherently more constrained with regard to detailed statistical specification. With local-estimation methods we can implement highly specific and various response forms and complex non-linear linkages while avoiding propagating

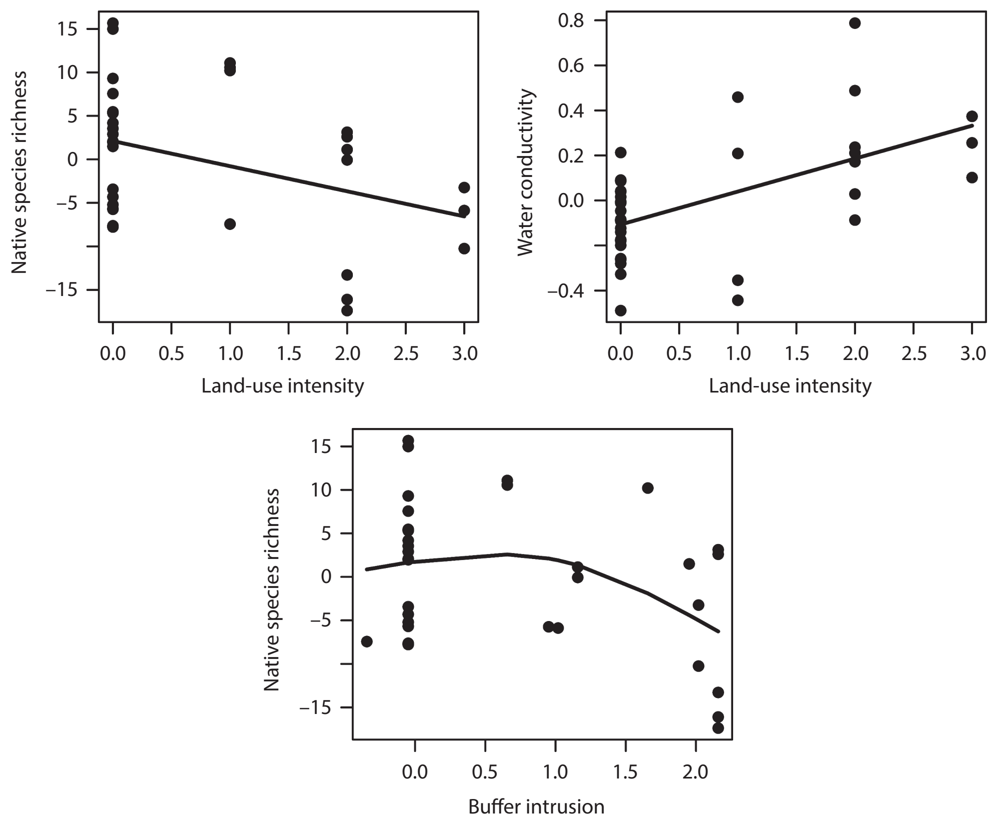

Fig. 8.10 Residual relationships between non-adjacent variables in initial model as revealed using local evaluation methods. The code for producing these graphs and associated results is presented in appendix 8.3.

misspecifications to the estimation of the whole model. In global estimation, declaring one response variable to be non-Gaussian causes the method of estimation of the entire model to change (e.g., from maximum likelihood to weighted least squares) and such changes can have undesirable features. Some software packages have the capacity to address statistical complexities within the global-estimation framework; however, these are always some form of approximate method, so local-estimation methods are generally more flexible in this regard.

When it comes to modeling with latent variables, global-estimation methods excel. By summarizing data and model implications as covariances, global-estimation methods permit estimation and evaluation of very complex latent hypotheses. To include latent variables with local estimations, one must use Bayesian MCMC methods. This approach permits greater flexibility (Lee 2007), but comes with a greater cost of time and expertise for setting up models, running them, and diagnosing problems.

Global estimation also facilitates modeling other types of relationships, including feedbacks, causal loops, and correlations among error terms. These types of relationships can be estimated for observed variable models using local-estimation methods, but with greater effort (appendix 8.3). Bayesian MCMC procedures do not readily permit estimation of models having causal loops. Thus, the choice between global-estimation and

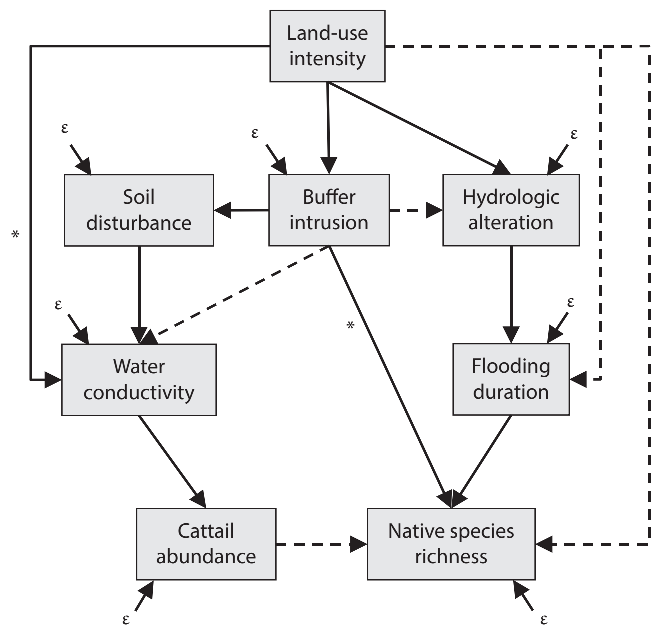

Fig. 8.11 Revised model based on local estimation and evaluation (see appendix 8.3). Paths with asterisks were added based on the discovery of non-independence. Paths represented by dashed lines were not supported by the data. Epsilons signify error variables representing the influences of unspecified factors on each endogenous variable in the model. The model differs slightly from that arrived at using global-estimation methods and linear relations (see figure 8.7b).

local-estimation approaches will depend on the structure of the model and its underlying theory.

Sample size requirements are a complex topic with a large and somewhat inconsistent literature. In general, careful consideration of the relationship between raw sample size (which is generally equal to the number of observations in SEM studies) and model complexity is very important in SEM. When power is low, one might consider reducing model complexity by including only the most important relationships (Anderson 2008). While general recommendations for standard sample size requirements are sometimes found in the older literature, it is inappropriate to require some fixed number of samples because sample adequacy depends on model complexity. Thus, the important issue is the number of samples per parameter (d). Our general advice is that a d of 20 is plenty, a d of 5 is on the low end, and a d of 2.5 is marginal. In our example, we had 37 samples and 9 estimated parameters in our final SE model, for a $d = 4.1$, a ratio of samples to parameters far less than ideal. Often the SEM analysis is intended to provide motivation for further studies that are stronger in numerous regards, including sample size. Because of the dependence of power on both sample size and model structure, it is important to consider the motivating theory and the possible model prior to data collection. Many things influence sample size, including the purpose of the study, the feasibility of large samples, and the need to use previously collected data, so one must be flexible while remaining

Box 8.5 PREDICTION EQUATIONS

```
# buffer response in logits
buffer_hat <- -2.9723 + 2.3232 * landuse
# transform from logits to proportions and rescale to (0:3)
buffer_hat_pr <- ( 1 / ( 1 + 1 / ( exp ( buffer_hat ) ) ) ) * 3

# hydro response: if landuse = 0, hydro = 0.739 else hydro = 4.21
hydro_cat_pred <- ifelse ( landuse == 0, 0.739, 4.21)

# soil response in logits
soilp_hat <- - 4.1617 + 1.6707 * buffer
# transform from logits to proportions and rescale to (0:1)
soilp_hat.pr <- ( 1 / ( 1 + 1 / ( exp ( soilp_hat ) ) ) )

# flooding duration response in logits
floodingp_hat <- -1.3146 + 0.4212 * hydro + 0.6845 * landuse
   - 0.55 * buffer
# transform from logits to proportions and rescale to (0:1)
floodingp_hat_pr <- ( 1 / ( 1 + 1 / ( exp ( floodingp_hat ) ) ) )
   * 365

# water conductivity response, in log-transformed units
cond_hat <- 1.8056 + 0.222 * soil + 0.2266 * landuse

# native species richness as poisson response; (log units)
richness_hat <- 3.998 - 0.00271 * flooding - 0.4556 * buffer3
# transform to linear units
richness_hat_tr <- exp ( richness_hat )

# cattail abundance as segmented/change-point response to
   conductivity
cattails_hat <- 0.14 + 1.4903 * ( ( cond - 2.104 ) > 0 )
   * ( cond - 2.104 )
```

skeptical about conclusions based on limited sample sizes. Not all estimation methods are equally defensible for small samples. Maximum likelihood global estimation is based on large sample theory and can lead to over-fitting when sample sizes are small (Bollen 1989). Local estimation is considered by some to be more tolerant of small sample sizes when a Bayesian MCMC approach is used (Lee and Song 2004).

8.3.6 *Computing queries and making interpretations*

Once final parameter estimates have been obtained for a SE model, it is time to use those estimates for interpretive purposes. A general technique for post-estimation analysis is the *query*. In the context of SEM, a query is a computation made using the prediction equations to yield some specific quantity or set of quantities. Such queries are extremely valuable in summarizing effects in models with non-linear linkages. They are also very useful for expressing the causal predictions implied by a model. Finally, queries allow us to use our hard-earned prediction equations to consider broad ecological implications that emanate from our scientific investigations.

There are four basic kinds of queries. Two are *retrospective*, looking backward in time. One of these is the query of *attribution*, which asks, "What prior conditions and processes led to the observed outcomes?" The second retrospective query is the *counterfactual*, which asks, "What would have happened if?" For example, what if wetland #12 had been exposed to a different set of conditions? There are also two *prospective* queries about future conditions. The *forecast* or prediction is a forward-looking query that extrapolates from current conditions and processes to future outcomes. Weather forecasts are a familiar example. *Interventions* are another kind of prospective query which ask, "What would happen if we changed (intervened upon) the conditions?"

When linear Gaussian models are used, information transfer through the network can be summarized by the multiplicative and additive properties of simple path coefficients (Wright 1921). We can calculate the strength of indirect and total effects, for example, by multiplying the coefficients along compound pathways. Using the `lavaan` model results (box 8.3, part J), the total effect of land-use intensity on native species richness can be computed as the sum of the direct (–4.804) and indirect ($1.413 \times 35.595 \times -0.08 = -4.024$) pathways (total effect = –8.828). When we use more complex model specifications, the computation of effects is more complex as well. In this situation, effects of varying land-use intensity from its minimum to maximum values on native richness through various pathways must be computed using the prediction equations given in box 8.5. Illustrations of such computations are given in the online supporting material for Grace et al. (2012).

One of the most common ways of expressing results is through the computation of standardized effects. An unstandardized effect represents the responsiveness of y to x in raw form. For the case of the total effect value of –8.828 computed above, this is a loss of nearly 9 species per unit increase in land-use intensity category. Such values have stand-alone interpretability and are the raw materials for our prediction equations. In contrast, standardized effects facilitate comparisons and can be used to talk about the relative importance of different variables and pathways. Classical standardization takes an unstandardized slope and multiplies it by the ratio of the standard deviations of x and y. Since slopes are in units of y/x, multiplying by $\mathrm{SD}(x)/\mathrm{SD}(y)$, puts the standardized coefficients into common units (standard deviations).

While standardized coefficients are widely used and advocated for (e.g., Schielzeth 2010), others have strongly criticized their use (e.g., Greenland et al. 1999), primarily because standard deviation estimates are strongly sample-dependent and using them in computations introduces multiple sources of error into the resulting estimates. In response to these criticisms, Grace and Bollen (2005) proposed an alternative method of standardization that uses the "relevant ranges" of the variables in place of their standard deviations. This changes the interpretation of a standardized coefficient into "the predicted change in y as a proportion of its range of expected values if we were to vary x across its range." Aside from providing a coefficient that has a more intuitive interpretation, this latter method lets the investigator decide the ranges over which slopes are relevant, giving further control to interpretation and comparison. This approach is especially relevant when comparing populations, where standard deviations will certainly not be constant. The use of relevant-range standardization is particularly useful in SEM when using complex model specifications because standard deviations are not good summary statistics for highly non-normal variables. Illustrations of the computation of both classical and relevant-range standardization are given in the online supplement to Grace et al. (2012).

The most basic forecast asks what would be the characteristics of a new sample from the same population. For example, we can ask, “If we took a random sample of wetlands from the same distribution and if cattails respond to conductivity as before, what would be the distributions of conductivity and cattail abundances?” Our interest here is extrapolating from a small sample. What we discover is that our simple extrapolation predicts a broader range of conductivity than seen in our sample but the current range of values for cattails covers the span we might expect from a much larger sample (figure 8.12).

Interventions are alterations of conditions, whether by humans or nature. An intervention-based query of interest is, “What would happen to the distributions of conductivity and cattail abundances if we could control influences on water conductivity?” We answer this question using a simplified model (figure 8.13). What we discover (figure 8.14) is that even when action is taken to prevent human influences, there is still an influence from other unspecified factors, which could be bedrock sources or effects or other past human activities.

The most difficult question to answer from a causal analysis is “What would have happened to a particular individual in our sample if in the past it had been subjected to different conditions?” Consider the scatter of points in figure 8.9b where wetlands have greater or lesser cattail cover than predicted. One wetland has 100% cover even though its conductivity is about 1.8, well below the estimated threshold. An informal examination of the evidence suggests that this wetland would have high levels of cattails even if it had never been influenced by human activities. Our evaluation here is informal; for formal analysis of such counterfactuals we refer the reader to Morgan and Winship (2007).

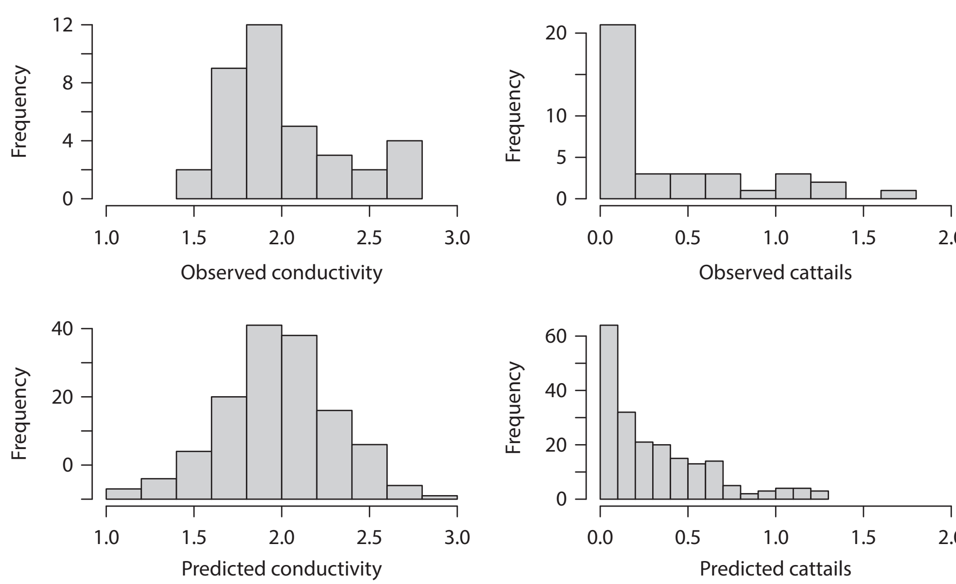

Fig. 8.12 Histograms for water conductivity and cattails, presenting both the observed distributions (upper) and the distributions that would be predicted for a new sample of 200 wetlands (lower).

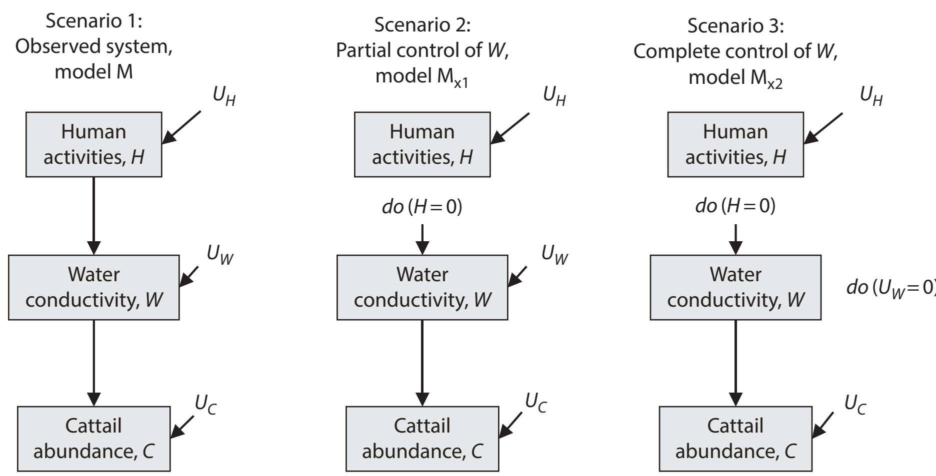

Fig. 8.13 Queries about predicted effects of interventions on water conductivities and cattail abundance. Scenario 1 is status quo; scenario 2 is the elimination of buffer intrusion and soil disturbance; scenario 3 is a reduction of water conductivity to reference conditions. The "U" variables refer to unspecified causes of variation. The operator "do(H = 0)" refers to reducing the values of land use and soil disturbance on conductivity to 0. The operator "do(U_W = 0)" refers to reducing the value of the unspecified influences on conductivity to 0. From Grace et al. (2012).

8.3.7 *Reporting results*

The guidelines in figure 8.5 (see also section 8.3.1) provide some advice for the general features of the study that should be reported. In particular, it is important to present the modeling rationale as explicitly and clearly as possible. Including a conceptual meta-model in your paper can help. Often, it is desirable to show both the initial and final SE models, unless the study was highly exploratory and presenting the initial model would just confuse the reader. In such cases it is important to declare that this was an exploratory, model-building exercise. When applications are model-comparing, however, SEM convention calls for describing each model examined and all modifications made (e.g., Laughlin et al. 2007).

A key aspect of any SEM analysis is to verify that the results are based on a model that is justified based on the data. Model-fit statistics, such as those presented in box 8.2 part F, are typically expected for cases of globally estimated models. For studies employing local methods, the number we are looking for is zero, i.e., no missing linkages that should be in the model. So, it is important to describe the examinations conducted and reassurances of model–data consistency.

Regarding model parameters and quantities, often final models and results are summarized graphically, though sometimes only the main findings (e.g., total effects) are summarized in the paper when many models are examined. Graphical presentations of final models typically include some representation of the parameters. If graphs are based on standardized parameters, then the primary unstandardized estimates and their properties (e.g., box 8.3) should be presented in a table or at least in online supporting materials. Readers typically find visual enhancement of SE model results, such as the inclusion of drawings or inset images, quite helpful (e.g., Alsterberg et al. 2013). Extrapolations such as those given in figure 8.14 have seldom been presented, but we feel this is an underutilized opportunity.

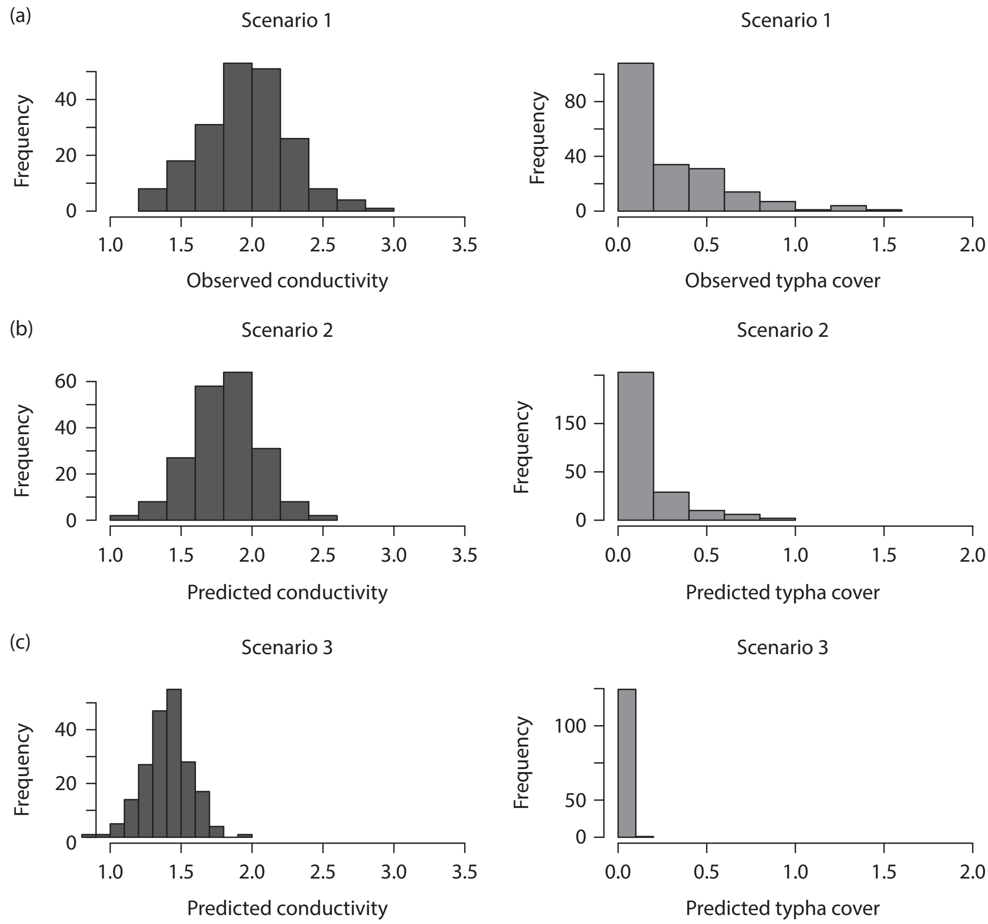

Fig. 8.14 (a) The observed distributions of water conductivity and cattail abundance (as cover). (b) The predicted distributions if effects of human activities are eliminated (Model M_{x1} in figure 8.13). (c) The predicted distributions if conductivity were completely controlled (Model M_{x2}). From Grace et al. (2012).

8.4 Discussion

The development of ideas and techniques that are emphasized in current treatments of SEM have been fueled by the needs of particular subject-matter disciplines primarily outside the natural sciences (specifically, in the human sciences). As a result of its multi-origin evolution, SEM has accumulated a large literature, much of which is not very accessible to biologists and ecologists. We believe a next-generation implementation is needed, both to incorporate transformative advances in methodology and to expand the practical potential of SEM (Grace et al. 2012).

The form of the presentation in this chapter differs from most other presentations of SEM currently available in that it is more focused on general capabilities than on technical details. This emphasis of ours is reflected in our consideration of both standard, matrix-based (global) and new, graph-theoretic (local) implementations. In this chapter we suggest fundamental principles for developing models and provide a formal exposition of causal diagrams as a useful step toward that goal, as causal diagrams support and encourage a rigorous consideration of causal assumptions. We contrast the classical SEM implementation using matrix techniques and global estimation with graph-theoretic methods using graphical analysis and local estimation. We further illustrate the use of both approaches, as each has its strengths and weaknesses. For example, global models are more constrained in the used of complex, non-linear, and non-Gaussian specifications, while local estimation permits a greater variety of specification forms. On the other hand, global estimation facilitates the inclusion of latent variables and non-recursive relationships. Another point of emphasis in our presentation is that obtaining parameter estimates is just the first step in exploring the implications of model results. Because of the tremendous flexibility and potential complexity of SEM, we only present core concepts and a limited illustration. In the following paragraphs, we advise the reader about additional issues and resources.

Bollen (1989) provides a historic treatment of the subject while Hoyle (2012a) summarizes recent advances. For those in the ecological and natural sciences, we recommend Mitchell (1992), Shipley (2000), and Grace (2006). For the measurement of theoretical entities see Raykov and Marcoulides (2010) for a psychometric perspective. Practical aspects of measurement in the social sciences are presented by DeVellis (2011) and Viswanathan (2005). For an ecologist's perspective, see Grace et al. (2010), especially with regard to the use of latent variables. Pearl (2009) deals with the theoretical development and use of causal diagrams. Practical presentations can be found in the field of epidemiology (Greenland et al. 1999). Recent summaries of specification complexities for global-estimation or local-estimation applications can be found in Gelman and Hill (2007), Zuur et al. (2007, 2009) , Edwards et al. (2012), and Hoyle (2012b), as well as in various chapters in this book. The relationship of data characteristics to specification is an important related topic (Graham and Coffman 2012; Malone and Lubansky 2012). For model evaluation and selection, West et al. (2012) have summarized choices under global estimation, while Shipley (2013) further illustrates the use of graph-theoretic methods for local estimation. The topic of interpreting the final model is one where a few technical aspects have received general treatment, though much of interpretation becomes discipline-specific. Grace and Bollen (2005) provide a general summary of coefficient types and their interpretations. More modern topics, such as queries and counterfactuals, are covered by Morgan and Winship (2007) and Pearl (2009).

There has been a notable expansion in the number and variety of applications of SEM in the natural sciences in recent years. SEM studies of trophic interactions (e.g., Lau et al. 2008; Riginos and Grace 2008; Laliberte and Tylianakis 2010; Beguin et al. 2011; Prugh and Brashares 2012), plant communities (e.g., Weiher 2003; Seabloom et al. 2006; Laughlin 2011; Reich et al. 2012), microbial communities (e.g., Bowker et al. 2010), animal populations (e.g., Janssen et al. 2011; Gimenez et al. 2012), animal communities, (e.g., Anderson et al. 2011; Belovsky et al. 2011; Forister et al. 2011), ecosystem processes (e.g., Keeley et al. 2008; Jonsson and Wardle 2010; Riseng et al. 2010), evolutionary processes (e.g., Scheiner et al. 2000; Vile et al. 2006), and macroecological relations (Carnicer et al. 2008) have been conducted. While SEM has been most commonly applied in observational studies, there have been numerous applications involving experimental manipulations (e.g., Gough and Grace 1999; Tonsor and Scheiner 2007; Lamb and Cahill 2008; Youngblood et al. 2009). To date, relatively few studies have applied Bayesian methods to ecological applications of SEM (e.g., Arhonditsis et al. 2006; Grace et al. 2011; Gimenez et al. 2012).

Finally, we reiterate that a confident, causal understanding requires pursuit via a series of studies that pose, test, and revise hypotheses. SEM, through both its philosophy and procedures, is designed to be used in that way. When you can turn your scientific understanding into a network hypothesis and evaluate it against data, a learning process is triggered that can change not only the way data are analyzed, but also how studies are designed and conducted. It is our experience that SEM can contribute significantly to this process, thus advancing our understanding of ecological systems. This chapter omits discussion of many possibilities and technical issues. We refer readers to our educational website http://www.nwrc.usgs.gov/SEM.html for more information.

Acknowledgments

We wish to thank Jessica Gurevitch, the editors of this volume, James Cronin, and E. William Schweiger, for helpful suggestions for improving the manuscript. JBG and DRS were supported by the USGS Ecosystems and Climate & Land Use Programs. The manuscript includes work done by SMS while serving at the U.S. National Science Foundation, though the views expressed in this paper do not necessarily reflect those of NSF. Any mention of software does not constitute endorsement by the U.S. Government.