esa

ECOSPHERE

SYNTHESIS & INTEGRATION

Guidelines for a graph-theoretic implementation of structural equation modeling

James B. Grace,[1,†] Donald R. Schoolmaster Jr.,[2] Glenn R. Guntenspergen,[3] Amanda M. Little,[4] Brian R. Mitchell,[5] Kathryn M. Miller,[6] and E. William Schweiger[7]

[1]*U.S. Geological Survey, National Wetlands Research Center, 700 Cajundome Boulevard, Lafayette, Louisiana 70506 USA*
[2]*Five Rivers Services, LLC at the U.S. Geological Survey, National Wetlands Research Center, 700 Cajundome Boulevard, Lafayette, Louisiana 70506 USA*
[3]*U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA*
[4]*Biology Department, University of Wisconsin-Stout, Menomonie, Wisconsin 54751 USA*
[5]*National Park Service, Northeast Temperate Network, Woodstock, Vermont 05091 USA*
[6]*National Park Service, Northeast Temperate Network, Acadia National Park, Bar Harbor, Maine 04609 USA*
[7]*National Park Service, Rocky Mountain Network, Fort Collins, Colorado 80525 USA*

Citation: Grace, J. B., D. R. Schoolmaster Jr., G. R. Guntenspergen, A. M. Little, B. R. Mitchell, K. M. Miller, and E. W. Schweiger. 2012. Guidelines for a graph-theoretic implementation of structural equation modeling. Ecosphere 3(8):73. http://dx.doi.org/10.1890/ES12-00048.1

Abstract. Structural equation modeling (SEM) is increasingly being chosen by researchers as a framework for gaining scientific insights from the quantitative analyses of data. New ideas and methods emerging from the study of causality, influences from the field of graphical modeling, and advances in statistics are expanding the rigor, capability, and even purpose of SEM. Guidelines for implementing the expanded capabilities of SEM are currently lacking. In this paper we describe new developments in SEM that we believe constitute a third-generation of the methodology. Most characteristic of this new approach is the generalization of the structural equation model as a causal graph. In this generalization, analyses are based on graph theoretic principles rather than analyses of matrices. Also, new devices such as metamodels and causal diagrams, as well as an increased emphasis on queries and probabilistic reasoning, are now included. Estimation under a graph theory framework permits the use of Bayesian or likelihood methods. The guidelines presented start from a declaration of the goals of the analysis. We then discuss how theory frames the modeling process, requirements for causal interpretation, model specification choices, selection of estimation method, model evaluation options, and use of queries, both to summarize retrospective results and for prospective analyses.

The illustrative example presented involves monitoring data from wetlands on Mount Desert Island, home of Acadia National Park. Our presentation walks through the decision process involved in developing and evaluating models, as well as drawing inferences from the resulting prediction equations. In addition to evaluating hypotheses about the connections between human activities and biotic responses, we illustrate how the structural equation (SE) model can be queried to understand how interventions might take advantage of an environmental threshold to limit *Typha* invasions.

The guidelines presented provide for an updated definition of the SEM process that subsumes the historical matrix approach under a graph-theory implementation. The implementation is also designed to permit complex specifications and to be compatible with various estimation methods. Finally, they are meant to foster the use of probabilistic reasoning in both retrospective and prospective considerations of the quantitative implications of the results.

Key words: Acadia National Park; Bayesian analysis; causal analysis; causal diagrams; guidelines; metamodels; statistics; structural equation modeling; structural equation metamodels; wetlands.

Received 16 February 2012; revised 11 June 2012; accepted 2 July 2012; **published** 16 August 2012. Corresponding Editor: A. McCall.

† **E-mail:** gracej@usgs.gov

Introduction

Structural equation modeling (when terms defined in the glossary in Box 1 are used for the first time, they are italicized) is a methodology increasingly used by those in the natural sciences to address questions about complex systems (Shipley 2000*a*, Grace 2006). It has several distinctive characteristics. First, since its origin (Wright 1920, 1921) its emphasis has been on providing a framework for learning about causal processes. Inferring cause-effect relationships has particular requirements that go beyond traditional statistics (Pearl 2009) and as a result structural equation modeling (SEM) characteristically involves a sequence of steps designed to support such inferences. Second, SEM is best understood as a framework for quantitative analysis that uses statistical techniques rather than a statistical method itself. A great variety of statistical techniques and tools have been used in the process of specifying and evaluating SE models and users of this methodology have been quick to incorporate new statistical techniques as they become available. Third, SEM permits the evaluation of networks of direct and indirect effects. As Wright noted and Pearl has reemphasized, *network models* are a natural device for considering causal relations. Many of the limitations of traditional *statistical models* can be understood by recognizing their inflexible architecture and inability to represent networks of causal relations (discussed in Grace 2006). Fourth, SEM incorporates both graphical and mathematical representations. Understanding causal relations is inherently difficult without the utilization of *graphical models*. The use of *graphical modeling* methods for the analysis of multivariate data permits the explicit expression of causal hypotheses. As a result of all these features, we feel that SEM has a unique and important role to play in quantitative science.

From a historical perspective, the first generation of SEM dates back to the early work by Sewall Wright (1920, 1921), who simultaneously invented *path analysis* and graphical modeling in pursuit of causal analysis of biological systems. This early work spread to econometrics (Haavelmo 1943) and the social sciences (Blalock 1964), laying the ground work for the emergence of the methodology that came to be known as structural equation modeling. The early implementation of this methodology was limited to the analysis of correlation matrices. The second generation of SEM was born with Jöreskog's (1973) synthesis of factor and path relations under the LISREL model. Through the comparison between model-implied and observed covariance matrices and maximum likelihood methods, this synthesis launched a tremendous growth of applications, involving both latent and observed variables, continuing to the present (e.g., the journal *Structural Equation Modeling*).

Aside from Wright's own studies, the use of SEM in the natural sciences was uncommon until the 1990s, when illustrations of its potential utility for ecology and evolutionary biology were published (e.g., Mitchell 1992, Wootton 1992, Brown and Weis 1995, Shipley 1995, Pugesek and Tomer 1996, Grace and Pugesek 1997). 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., Gotelli and Ellison 2006, 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 obser-

Box 1

Glossary of terms used in paper.

Bayesian reasoning—The use of Bayes theorem to consider within a network how knowing the states of certain variables influences our inferences about other variables.

causal diagram—A representation of the nodes (variables) and their connections (links) that are hypothesized to exist for a situation being modeled. The three distinctive properties of causal diagrams that distinguish them from other representations (such as metamodels and SE models) are (1) they may include nodes regardless of whether those nodes will be represented in any given SE model in order to aid logical analysis, (2) nodes included may be observed or hidden (latent), and (3) the functional forms for the linkages are not specified. Causal diagrams are representations designed to support causal analysis. The causal diagram is intended to permit the reading of (a) logical and testable implications, (b) identifiable parameters, (c) recognition of instrumental variables, (d) recognition of equivalent models, and (e) recognition of minimum admissible sets for determining causal effects. For further explanation refer to (Pearl 2012).

causal indicator—In SEM observed values of measurements are assumed to be indicators for a underlying process. Causal indicators are those that cause or contribute to the make up of the theoretical entity. In this case, the arrows representing the direction of causality point from the observed indicator to the theoretical variable (e.g., a composite variable). Causal indicators contrast with the more traditional *effect indicators*, which represent the effects of a latent entity. Causal arrows point from a latent variable to its effect indicators.

causal model—A model is causal if it makes a causal claim and if key assumptions can be supported. Fundamentally, causal claims, e.g., that $X \rightarrow Y$ (X affects Y) require manipulations for confirmation. So, if we can alter X without simultaneously altering other causes of Y that are related to X, and then observe that Y responds consistently, we have evidence to support a causal claim. While this assumption may go untested in many causal studies, the results from causal modeling can be understood to be explicit predictions that may be testable under future circumstances. Often, scientists will accept, based on collective prior knowledge, that certain statistical associations are very likely causal. That does not mean that they are completely understood or simple.

conditional independence—Two variables in a model can be declared to be conditionally independent if any bivariate association between them can be explained indirectly. Variables that are conditionally independent in a model are typically not connected by a directed or undirected direct linkage.

d-separation—The d-separation criterion for any pair of variables in a graph involves (1) controlling for common ancestors (variables that are causal antecedents) that could generate correlations between the pair, (2) controlling for causal connections through multi-link directed pathways, and (3) *not* controlling for common descendent variables. Collectively, these three rules combine to provide the guidance needed to identify the expectations implied by a graph. Practically speaking, the goal of d-separation is to identify cases where we need to test for a correlation between the residuals of two variables not connected by a direct path to ensure a model is consistent with the data.

effect indicator—When indicator variables represent the effects or manifestations of a latent entity they are considered to be effect indicators. Contrast with *causal indicators*.

endogenous variable—A variable that is predicted by another variable in a model, and therefore is a response variable.

equivalent models—Models that have idential statistical expectations but different causal assumptions.

Box 1. Continued.

exogenous variable—A variable whose variations are not explained in a model.

global estimation—In classical SEM, estimation involves comparing the entire covariance matrix to the covariances implied in the whole model. This approach is also referred to as a "full-information" method. Global estimation is in contrast to *local estimation*, in which solutions are determined for each node based on each node's parent variables, plus any directly correlated variables.

graphical models—A graph is basically a network made up of nodes and linkages. In the most basic case, a node is represented using a random variable and linkages represent conditional probabilities in some functional form (e.g., Gaussian responses with a linear link).

graphical modeling—A field of quantitative analysis that has greatly expanded in the past two decades. Graphical modeling is often thought of depending on a merger between graph theory and probability theory.

instrumental variables—For a pair of variables (X, Y), an instrumental variable Z is one that is correlated with the predictor of interest X, but free from any source of correlation with the predicted errors for Y. Instrumental variables are used in causal analysis to obtain unbiased estimations of effects.

interventions—An intervention is a direct manipulation of the values of an entity. Experimental studies are one example of an intervention. The concept of intervention also extends to the topic of prediction when an intervention is not part of a randomized replicated experiment, but when its outcome is predicted.

latent variables—Hypothesized variables for which we have no direct measurements, but whose existence may be revealed by associations among measured variables.

local estimation—In graph-theoretic SEM, instead of estimation being accomplished by comparing the entire covariance matrix to the covariances implied in the whole model (i.e., *global estimation*), estimation is local, at the level of each response variable and its parents and direct correlates. This is sometimes referred to as a "limited-information" method.

mediation—A key feature of SEM is the test of mediation, which relates directly to the study of causal relationships using path relations. In the test of mediation, we ask whether the effect of one entity (X) on another (Y) can be explained by a third variable (Z), e.g., if $X \rightarrow Z \rightarrow Y$ holds true (and if it does, then X and Y are conditionally independent given Z).

metamodel—A metamodel is a generalization of the modeling problem that is noncommital as to measurement, either with regards to the variables that will be included in the model or the functional forms of the relationships between variables. The purpose of the metamodel is to make explicit the relation between theoretical entities or constructs and the measurements that will be used to represent them. When used in SEM, the metamodel is sometimes referred to as the structural equation metamodel (SEMM).

network model—A general term that refers to models that can be envisioned as a collection of nodes connected by links, where all possible linkages are permissible (at least theoretically). SE models are one type of network model. Other terms one may encounter include Bayesian networks, probabilistic networks, causal networks, interaction networks, and neural networks. There are various degrees of overlap among some of these terms and to some extent their meaning is partially conveyed by their history of use. Since all of these listed types of networks exist only as quantitative relationships among variables, they can all be thought of as "models" as opposed to physical entities, such as traffic networks that are made up of roadways for vehicles.

Box 1. Continued.

path analysis—A term often used to refer to structural equation models that only contain observed variables (omitting latent variables). Generally, the phrase structural equation model, even if only observed variables are included, implies a more modern treatment of assumptions, while the term path analysis is more ambiguous in this regard unless one makes it clear that they are performing path analysis using SEM techniques.

prospective analyses—A forward-looking computation that considers what might happen under future conditions. Contrast with *retrospective analyses*.

query—A specific request for a quantity or set of quantities to be computed using the prediction equations of an SE model.

retrospective analyses—A backward-looking computation that considers how past or present circumstances may have led to current conditions. Contrast with *prospective analyses*.

statistical model—One that represents probabilistic associations without explicit consideration of whether the parameters derived can be justifiably interpreted as causal effects.

statistical specification—When we refer to the statistical specification of a model, we mean the explicit statement of all the information needed to permit estimation, such as variable and error distributions and linkage functions. Causal diagrams, in contrast, omit statistical details in order to focus on causal relations among variables.

structural equation modeling (SEM)—The process of developing and evaluating structural equation models.

structural equation models (SE models)—Probabilistic models containing or specifying multiple causal pathways. SE models are characterized by (a) attempting to satisfy the criteria for drawing causal inferences and (b) permitting endogenous variables to be functions of other endogenous variables, thereby potentially containing indirect effects.

structural model—The basic meaning of a structural model refers back to the original concept from econometrics that an equation that can be supported as representing a causal relationship is described as "structural" (Haavelmo 1995). Essentially, a causal probabilistic model is a structural model and structural equation models aspire to be fully structural (all coefficients can be given causal interpretations), though this is not always true in practice (i.e, some relationships in SE models may only represent probabilistic associations).

untestable assumptions—Assumptions in a causal diagram or model that cannot be tested with observational data. For example, for a pair of variables (X, Y), we may make the assumption that the direction of causation is $X \rightarrow Y$. Such an assumption is testable if we are able to manipulate X while holding constant other variables affecting Y that are correlated with X.

vational 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).

While interest in applying SEM has been increasing in the natural sciences, both the scientific and statistical ambitions of SEM practitioners have also been growing. An interesting and important source of new thought about SEM has been coming from the field of artificial intelligence. Pearl (2009) has suggested that the development of intelligent systems requires explicit consideration of causality and that structural equations are the natural language for representing and studying causal relations. He has further developed a coherent theory for explicating the requirements for causal reasoning under uncertainty. His approach has generalized SEM to the nonparametric level and he has systematically applied a graph-theoretic approach to causal analysis. He has also proposed new mathematical operators to support the extraction of causal interpretations from data and has synthesized these ideas into a suggested set of requirements for SEM to serve as an inference engine, with defined input requirements (Pearl 2012). Many of these ideas have yet to be incorporated into general SEM practice, though a few have (Shipley 2000*b*, 2003, 2009).

Another set of new possibilities for SEM has emerged from recent developments in Bayesian statistics. Historically, the second generation of SEM put forward by Jöreskog (1973) has been framed in terms of likelihood statistics; thus, there is a natural linkage to Bayesian estimation since Bayesian estimates are simply the likelihoods weighted by the prior probabilities. Adopting a Bayesian approach provides for three main opportunities. First it opens up the inference process to include a wider array of information sources through the priors. While it may be arguable how useful prior information will be for judging inferences for a set of data, prior information can be very useful when the goal is to forecast future observations. Second, it opens up the range of statistical specifications that can be estimated because of the flexibility of Markov chain Monte Carlo (MCMC) procedures (Gelman et al. 2004). Third, the use of *Bayesian reasoning* expands the range of possible applications of the information (Kjaerulff and Madsen 2008). This makes the transition from *retrospective analyses* (e.g., What seems to have caused the data we see?) to *prospective analyses* (e.g., What would happen if conditions changed or how would our inferences change if we learned new information?) easier. It is important in all this to keep in mind that our goal should not be to make the requirements for modeling intractable.

While technical innovations that can expand SEM's capacity for theory translation (Grace et al. 2010), causal analysis (Pearl 2009), and statistical specification (e.g., Lee 2007) have advanced, an overall strategy for their integration into the SE modeling process has not yet taken place. In fact, many of the requirements for causal analysis and advanced statistical analysis are incompatible with the classical implementation of SEM via the analysis of a covariance matrix. Thus, while the new possibilities represent an important opportunity for advancing SEM, there is a need for a new synthesis of these possibilities into a coherent whole. Such a synthesis could result in a third generation of SEM in which theory translation, causal inference, and prospective analyses are strengthened and integrated, allowing SEM practice to encompass a more complete scientific methodology.

The aim of this paper is to present and illustrate a more comprehensive SE modeling process that incorporates the above-described advances in logic, specification, and estimation into an integrated system. To facilitate integration, we use a graph-theoretic mathematical framework so as to generalize the great variety of possible statistical specifications and estimation techniques as part of a common methodology. To facilitate application, we expand on this new implementation through a set of guidelines for implementation.

Regarding the guidelines presented, we begin by considering how the goals of an analysis influence the model development. We then discuss the use of *metamodels* to represent the concepts of theoretical interest and the general hypotheses being investigated. Then we describe another newly-proposed device, the *causal diagram* (explained below), that can be useful for examining causal assumptions once theoretical ideas have been represented by measured or hypothesized variables. Completion of the full specification of SE models into a form where they can be quantitatively evaluated requires both an exact statement of the variables that will be used to represent entities of interest and the functional forms of relationships that link the variables. We describe how this process of *statistical specification* involves a number of decisions that are influenced by both the goals and theoretical background as well as the available data and its characteristics. A very wide array of choices for statistical specification are now possible in SE models and managing model complexity is an issue of concern that we discuss. Costs and benefits of Bayesian estimation for SE models are briefly discussed, as are some of the approaches to model evaluation that might be considered. Because models can range from simple to complex in their statistical specifications, we describe a number of options for how results from models can be summarized. In this presentation we seek to extend the traditional modeling process to include prospective/predictive analyses and illustrate this using queries (in this context, a *query* is a specific request for a quantity or set of quantities to be computed using the prediction equations of an SE model). Further, we provide advice for reporting on the process used, the results obtained, and the interpretation of results. Finally, we illustrate these guidelines by evaluating the linkages between human activities and wetland ecosystem conditions on Mount Desert Island, home of Acadia National Park; then by forecasting the possible consequences of interventions designed to protect the wetlands from nutrient runoff.

Materials and Methods

Structural equation modeling

Background.—The fundamental goal of SEM is to develop and evaluate models so as to learn about and represent underlying causal processes. A practical sense of the knowledge required for something to be considered a cause of something else is provided by Pearl (2009: 417) where he states, "Y is a cause of Z if we can change Z by manipulating Y." As will be described below, this definition foreshadows several things, including (1) the need for mathematical notations that permit expression of both the direction of causal influence and the expected effect of manipulations, (2) a fundamental distinction between causal/structural models and purely statistical ones (based solely on probability theory), and (3) a need for explicit consideration of causal assumptions and the testable and logical predictions they imply. While the essence of SEM is in its intent (the pursuit of causal understanding), it is also distinctive as a methodology because it permits and even requires representations of networks of relationships among variables. Classical statistical techniques do not permit sufficiently flexible representation of network relations to foster causal investigations. Further, SEM is not a purely statistical methodology, rather, it is a framework for quantitative scientific investigations that uses statistical principles along with scientific knowledge. Only the combination of statistical relations and causal assumptions can yield causal inferences (Pearl 2012).

Finally, while the aspiration of SEM is to learn about causal processes, the data by themselves do not provide enough information to demonstrate causation (unless complete experimental control of all processes of interest is achieved). Ultimately, it is the process of scientific investigation, the consideration of proposed mechanisms, and the accumulation of knowledge that provide the context for causal inference. Importantly, SEM results by themselves do not prove causation; instead, the pursuit of structural models facilitates the investigation of causal relations.

A graphical modeling representation.—From a mechanical viewpoint, in order to specify a network model we need three things, (1) a definition of graphical relations, including the nodes in the graph and the links connecting the nodes, (2) a vector of observed variables used to specify a particular instantiation of our graphical model, and (3) the statistical/mathematical functions used to convey the flow of information in the network (Kjaerulff and Madsen 2008). Thus, we can define network N as a superset comprising three sets such that

$$N = \{G, X, F\}, \tag{1}$$

where G is a graph set representing the network N in general terms, X is the set of variables used to operationalize G, and F is the set of functions used to link nodes. Further, $G = \{V, E\}$ where V is the set of nodes (or vertices) in the network and E the set of links between nodes (sometimes referred to as edges in graphical modeling). For the directed links in the set E, there is a recognized set of familial relations. Given a causally-ordered pair of variables $(u, v) \in E$, u is said to be a parent of v and v a child of u.

X, the set of variables, can include subsets that are random (observed or latent) variables (X_R) as well as those that are derived/computed (X_C), or decision/utility variables (X_D), including interventions/manipulations (Kjaerulff and Madsen 2008). X can also permit variables that are measured or conceptualized at different hierarchical levels. This generalization of the variable set substantially expands the representations possible in SE models. The inclusion of decision and utility variables also provides for an expanded capacity of networks to support probabilistic reasoning and expert system applications. The functions F used to pass information across links can be of various sorts and can include both conditional probabilities among discrete variables as well as a full array of probability density functions among continuous and categorical variables. Functional forms that link nodes can be linear, nonlinear, and more complex types.

Viewing the functional forms in F generically, the estimation problem for N can be described as estimating the joint probability distribution over the variables in the set X that describe the individual nodes v in the set of nodes V in graph G. In this context, the overall probability of a network given a set of X variables in V nodes can be represented by the combined conditional probability distributions P such that

$$P(X) = \prod P\left(X_v | X_{pa(v)}\right) \tag{2}$$

for all vs in V and where $X_{pa(v)}$ refers to the variables that are the parents of X_v. Thus, in a graphical modeling approach to SEM the estimation process is local and involves only limited or partial information from the whole network. This *local estimation* using "limited-information" methods differs from the matrix-level *global estimation* or "full-information" procedures characteristic of covariance-based SEM. The two fundamental advantages of local estimation are (1) it permits more complex specifications of responses and linkages than can be summarized by a covariance matrix and (2) a modularization of the estimation process that avoids propagating misspecification errors from one part of a network to other parts (Bollen et al. 2007).

A variety of estimation approaches can be applied to SEMs depending on several criteria, including the specific objectives of the analysis and the forms of the linkage functions in F. As stated earlier, SEM is a modeling framework for investigating multivariate causal relations and not a specific statistical technique. Second-generation SEM has been developed based largely on likelihood principles where we are interested in the likelihoods of various theoretical models, rather than null hypothesis testing where default preference is given to an independence model. As a result, there is a natural connection to the Bayesian approach since Bayesian estimates are simply the likelihoods weighted by the prior probabilities. Bayesian estimation of SE models using Markov chain Monte Carlo methods is now being used with increasing frequency (Arhonditsis et al. 2006, Lee 2007, Grace et al. 2011, Gimenez et al. 2012). Below we will say more about selecting the estimation method from amongst the available choices.

SEM as an inference engine.—The concept of SEM as an inference engine (Pearl 2012) envisions three kinds of inputs, (1) causal assumptions, (2) a set of queries of interest, and (3) data/information. The outputs from the analysis include (4) the logical implications that can be obtained from any given model independent of the data, (5) quantities produced from the analysis, and (6) the conditional (data-dependent) claims supported by the analysis. In this presentation we discuss these inputs and outputs in terms of the work flow steps in the modeling process, not as a rigid prescription, but to aid SEM applications. In the next section, we will briefly describe 10 steps to be considered when applying SEM. In our presentation we place special emphasis on practical criteria to consider when making modeling decisions, since this is one of the most commonly encountered gaps in the SEM literature. In this paper we illustrate the modeling process using a single example, thus greatly underrepresenting the range of possibilities. The rich topic of latent variable modeling has been recently addressed (Grace et al. 2010) and is not extensively explored here.

Guidelines for SEM

Step 1: Define the goals of the analysis.—As indicated in Fig. 1, the goals of an application are among the first things to consider when applying SEM. Often there may be a particular response or relationship one wishes to use as the focus of the investigation. For the example presented later in this paper, we are interested in the impact of human activities on ecosystem properties. In such a situation we may, for instance, wish to treat the natural unaltered forces controlling ecosystem properties as covariates because our primary interest is on isolating the effects of human activities. Such prioritizations may be quite important when deciding what is necessary to include in a model and what can be excluded so as to manage model complexity.

Decisions about statistical specifications may also be influenced by the goals of a study. When a study is focused on evaluating the linkages in a model, specification concerns are often directed

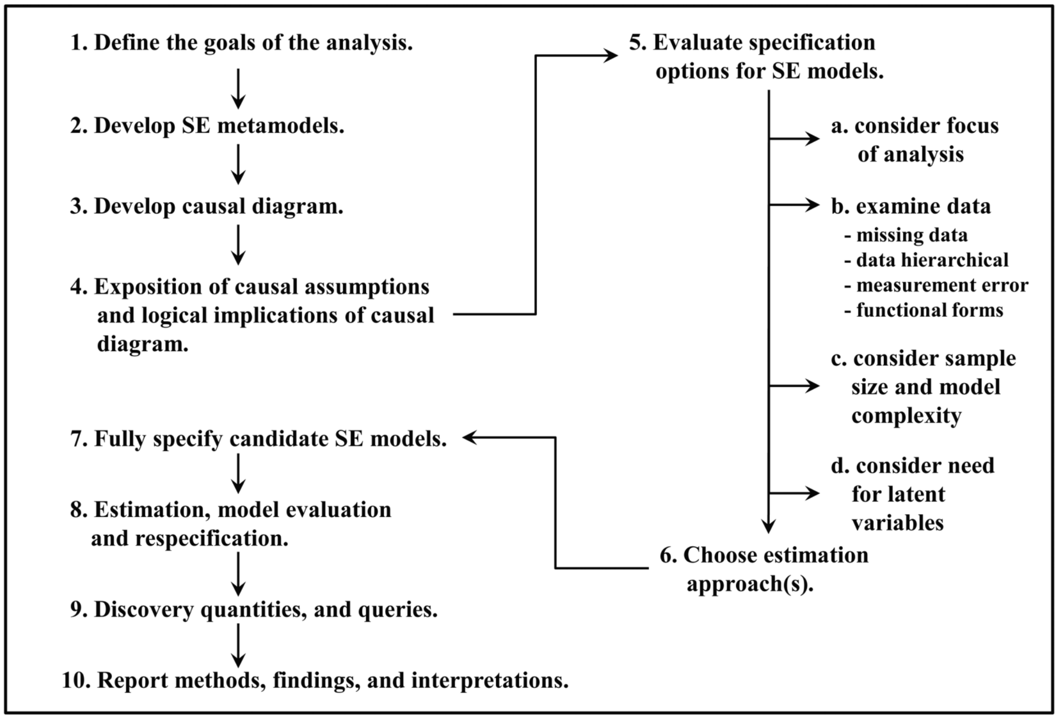

Fig. 1. Steps in the modeling process we recommend be considered as part of the guidelines for a graph-theoretic approach to SEM.

toward the parameters in the linkage equations rather than the predicted quantities (means and intercepts) of the nodes. If we are interested in predicting the states of the nodes, however, we may be quite concerned about distributional forms, since we may want the predicted scores to be compatible with what can be observed. For example, in the case of a quantity that is a proportion, it is not helpful if predicted values can range between $-\infty$ and $+\infty$ when observed values will fall between 0 and 1.0. In such a case, the form of the specification of quantities is important, since the specifications for prediction-oriented studies can be technically more demanding. As Edelaar et al. (2012) have recently pointed out, for certain data types only proper specifications can yield the correct inferences about linkages in the model.

One additional thing investigators should consider in any modeling enterprise is the definition of the system of interest and its boundaries. This decision has critical importance for how samples should be taken and inferences made. By explicitly stating the bounds of generalization, misinterpretation can be avoided. There tends to be a presumption in hypothesis testing that one is drawing inferences about some infinitely large (relative to the sample) homogeneous population. For the example presented in this paper, we instead wish to make inferences about the wetlands of Mount Desert Island so as to inform decision makers about potential threats. The results may be interesting to those investigating wetlands elsewhere (or other systems), but our study is not designed to yield quantitative inferences for the world at large. Further, we may view our sample as being a collection of unique entities from a defined area having a common set of properties, rather than samples from an infinitely large homogeneous population. The sphere of inference can also influence one's choices for modeling (e.g., the estimation and inference system) as well.

Step 2: Develop a structural equation metamodel as a device for representing theoretical expectations.—It is often recommended that researchers begin SEM with specification of the SE model. In this presentation, we argue that a more explicit treatment requires additional steps that precede model specification. The first of these is the evaluation and summarization of the relevant theory that will guide model development and the interpretation of results. One way to facilitate this is to develop what we call a "structural equation metamodel" or SEMM (Grace et al. 2010). A SEMM is a generalization of the modeling problem that is noncommital as to measurement, either with regards to the variables that will be included in the model or the functional forms of the relationships between variables. The purpose of the SEMM is to make explicit the relation between theoretical entities or constructs and the measurements that will be used to represent them. The study of causation requires a consideration of the theoretical content behind constructs (Pearl 2009). Deciding how a theoretical idea can be related to observations begins with a consideration of the validity of constructs and their hypothesized mechanistic interactions (Borsboom and Mellenbergh 2004, Grace and Bollen 2008). Theoretical constructs can be thought of as formalized concepts that are fundamental elements in theories. According to logical positivism, these postulated items do not really exist and theoretical concepts are merely economical devices used to explain observable phenomena while according to realists, theoretical entities are unobserved real phenomena (Borsboom and Mellenbergh 2004). We would argue that both kinds of constructs show up in theories. What is of importance in SEM is that the way we represent a construct using variables may differ depending on how we conceptualize the causal content of that construct.

There are several options for the finite specification of a construct. It is common in factor models (a core element of the "psychometric" tradition in SEM) that constructs are treated as coherent latent entities and are represented using one latent variable for each construct. The observed indicator variables that give us information about the latent entities are interpreted as responses or "effect indicators." However, in cases where a construct is an aggregate concept or "collection of things", it may be best modeled using *causal indicators* and the construct represented in the SE model using composite variables (Grace and Bollen 2008). Adding the step of developing an SEMM forces the investigator to explicitly describe the logic behind these specification decisions. It is expected that SEMMs will be particularly useful in studies where entities of theoretical interest are often heterogeneous collections, such as collections of species living together in a community.

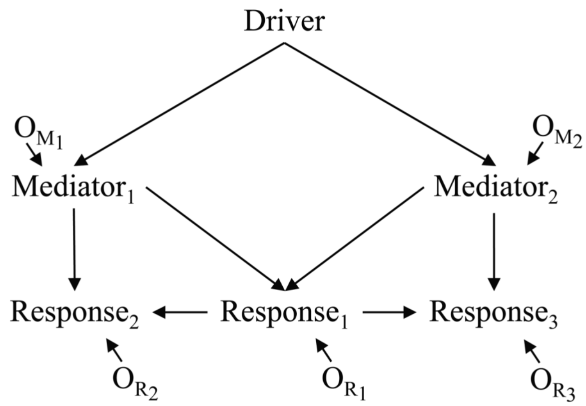

Fig. 2. Hypothetical causal diagram showing a hypothesized graph showing relations between a Driver, two Mediators, and three Responses of interest. The terms "O_x" refer to other factors that influence each endogenous variable. The purpose of the causal diagram is to facilitate careful causal analysis of a situation so as to aid decisions about what measurements are critical to obtain. Refer to the glossary of terms in Box 1 for a more complete definition and Pearl (2012) for further description.

Step 3: Develop a causal diagram.—A third possible step in SEM is to develop a "causal diagram." Causal diagrams are graphs summarizing hypothesized causal connections among variables (Pearl 1995, Greenland et al. 1999). These graphs are distinct from SEMMs because they specify the variables that may be included in the SE models. Their importance lies in the role they play in causal analysis. They are also distinct from fully specified SE models as well because causal diagrams are noncommital with regard to statistical particulars such as functional forms for equations (e.g., linear versus nonlinear relation, Gaussian versus Bernoulli responses). Despite this lack of commitment to a functional form, causal diagrams can be interpreted as a general type of probability model, allowing for the application of graph theory and deduction of causal assumptions and implications (Pearl 2012). To help clarify the concept further, we may consider a causal diagram as a kind of map of the modeling area (e.g., Fig. 2). For any given set of SE models we intend to specify, the diagram is intended to represent the causal domain without regard to how much of that domain we wish to use in creating particular models. We may, for example develop the outward areas of our diagram to encompass influences we know will not be included in our SE models, such as the potential causes of our *exogenous* variables. One of the purposes of these diagrams is to allow consideration of what variables need to be included in an SE model so as to permit estimation of particular causal effects of interest (i.e., what simplifications are permissible in SE model development). They can also be used as a formal tool for identifying potential biases and legal queries (those that don't violate causal requirements), such as predicting the effects of external interventions. Taking as given that a diagram is correct, one can see whether the causal effects of interest can be estimated from the available data or whether additional variables are needed for a particular purpose. A new addition to the possible steps in SEM, causal diagrams provide a major advance for attending to the logical requirements for causal interpretations.

Step 4: Exposition of causal assumptions and logical implications from the causal diagram.—As expressed by Pearl (2012), the causal diagram permits the reading of (a) logical and testable implications, (b) identifiable parameters, (c) recognition of instrumental variables, (d) recognition of equivalent models, and (e) recognition of minimum admissible sets for determining causal effects. This is a complex set of topics and we will not attempt to cover all the relevant concepts. Fig. 2 is a hypothetical causal diagram to aid our presentation (we will look at a real causal diagram for Acadia wetlands later in the paper). This diagram includes one driver variable (*Driver*$_1$), three response variables (*Response*$_1$–*Response*$_3$), and two mediators (*Mediator*$_1$ and *Mediator*$_2$). Because this is a simple example, we will only address a subset of the logical rules governing causal diagrams. Also, it is important to keep in mind that a causal diagram represents a model space and we may either not have measurements for all nodes in the space or deliberately choose to build models that are reduced form representations of the causal diagram, so we can view it as a template for model development.

There are several key concepts we need to mention before discussing causal assumptions or logical and testable implications in SE models. Since our goal is to obtain estimates of model parameters that convey causal effects, we must consider whether such estimates are *identifiable*. A parameter is identifiable if a true and unique estimate can be obtained from observed data. In the context of considering a causal diagram, the identification of parameters has been described from a statistical estimation perspective. This can be generalized to the identification of causal effects in nonparametric graphs as well (Brito and Pearl 2002). In practice, identification is also influenced by the data and limits of estimation methods, though here we are discussing rules that apply prior to the involvement of data or estimates.

The topic of *instrumental variables* is important for causal analysis. The relevant situation for considering instrumental variables is when some predictor of interest has a confounded relationship with the variable representing its effect. Confounding between a cause and its effect occurs, for example, when there is some backdoor correlation caused by an exogenous factor that influences both. Relative to Fig. 2, the causal effect of *Response*$_1$ on *Response*$_2$ is confounded by the joint effect of *Mediator*$_1$ on both. If we lack knowledge about *Mediator*$_1$, the reduced form of our causal diagram would be to have a correlation between the errors of *Response*$_1$ and *Response*$_2$. In such a case, classical estimation methods using data for *Response*$_1$ and *Response*$_2$ would yield a biased estimate for the causal effect. What can we do in such cases? This problem has interested economists since Philip Wright's first attempt to offer a solution in 1928 (see discussion of the history of this topic in Stock and Trebbi 2003). If we imagine a variable *Z* that is correlated with *Response*$_1$ but does not possess an uncontrolled (backdoor) relationship to *Response*$_2$ (i.e., does not have correlated errors with *Response*$_1$) then it is possible to compute the effect of *Z* on *Response*$_2$ and then calculate the proportion of that effect that can estimate the causal effect of *Response*$_1$ on *Response*$_2$. Z in this case is an instrumental variable for the task. The challenge in real-world applications of instrumental variables is with knowing that Z meets the requirements to serve as an instrumental variable. Causal diagrams can facilitate the considerations associated with that decision.

Equivalent models are those that have identical statistical expectations but different causal assumptions and interpretations. There typically exist equivalent SE models that are not valid causal models, but that cannot be rejected based solely on statistical model-fit criteria. An exposition of the causal assumptions in a model is given below. That said, software designed to facilitate the consideration of equivalent models that is general enough to cover the nonlinear situation would be a helpful addition to the enterprise of thoroughly considering equivalent models. At present, the TETRAD software (Spirtes et al. 2000) is perhaps the most complete option.

Fundamental to graphical considerations of the requirements for determining causal effects is the concept of *d-separation* (Pearl 1988). We can understand the context for a d-separation criterion by imagining that our goal is to write an algorithm to automate the finding of appropriate control variables for testing the *conditional independence* of any pair of variables not connected by a direct link (e.g., through the computation of partial correlations). These algorithms can be used to identify conditional independence implications in a causal diagram prior to model estimation. Now, we can state that the d-separation criterion for any pair of variables in a graph involves (1) controlling for common ancestors (variables that are causal antecedents) that could generate correlations between the pair, (2) controlling for causal connections through multi-link directed pathways, and (3) *not* controlling for common descendent variables. Collectively, these three rules combine to provide the guidance needed to identify the expectations implied by a graph. Practically speaking, the goal of d-separation is to identify cases where we need to test for a correlation between the residuals of two variables not connected by a direct path to ensure a model is consistent with the data.

Let us represent the variables in Fig. 2 using only their first letters plus subscripts. Our goal is to identify some set of variables S such that $M_1 \perp M_2|S$ (in words, M_1 and M_2 are conditionally independent if we know the variables in set S). If we assume the linear Gaussian case, the goal is to decide for our algorithm the conditioning variables for computing partial correlations appropriately. The d-separation criterion tells us that in this graph we expect $r_{M1,M2|D1} = 0$, i.e., the partial correlation between M_1 and M_2 is predicted to be zero. The criterion also tells us that if we were to use any common descendent of M_1 and M_2 as our control variable (e.g., R_1, R_2, or R_3) we would incorrectly generate a nonzero partial correlation between our target pair of interest. Thus, the d-separation criterion allows us to identify testable implications in causal diagrams and in SE models. There are several implied conditional independences and these constitute missing direct links in the diagram (Fig. 2). For example, there is no link from $D_1 \rightarrow R_1$ in our graph. The observed bivariate (zero-order) correlation between D_1 and R_1 will be determined by the effects of D_1 on R_1 through M_1 and M_2. The question of whether our graph is a correct causal connection can be partially tested by determining empirically if $r_{M1,M2|D1} = 0$. Furthermore, our ability to use the diagram to decide what models will yield causal versus confounded inferences depends on the assumptions, including both testable and untestable assumptions.

Elaborating on the topic of *untestable assumptions*, there are several assumptions in our example diagram that cannot be tested with observational data. The causal assumption that D_1 can have an effect on M_1 and not vice versa cannot be tested by simply observing the correlation between the two. As mentioned above, there exist equivalent models (those with identical independence claims but differing in causal assumptions) that offer alternative causal interpretations. One of the uses for a causal diagram is to help explicate predictions that could render these "untestable" assumptions testable through *interventions* (e.g., experimental manipulations). For example, Fig. 2 predicts that if we were to manipulate values of D_1, we would see some response in M_1, but not vice versa. This implied prediction helps to identify the information required if we need to test the causal validity of the directed pathways in our model. Interventions are incorporated into the language of structural equations in the form of the "do(•) operator" (Pearl 2009). For example, the statement $do(D_1 = d_1)$ describes an intervention where

Table 1. Biological characteristics selected as the top candidates for inclusion in an index of biotic integrity (Schoolmaster et al. 2012).

Biological characteristic	Bivariate correlation with HDI	Akaike weight for contribution to IBI
Typha abundance	0.71	0.95
Sphagnum abundance	−0.69	0.75
Perennial abundance	0.47	0.36
Forb abundance	0.60	0.36
Forb species richness	−0.55	0.25
Dicot abundance	0.55	0.23
Monocot richness	−0.66	0.20

the value of the variable D_1 is set to a specific value d_1. It is a prediction emerging from our analysis that the response of M_1 to the manipulation $do(D_1 = d_1)$ will be the same as that expected for the passive observation $D_1 = d_1$. The importance of the diagram is that it represents these various causal claims without reference to the data or to statistical quantities.

Step 5: Evaluate specification options for SE models.—A structural equation model differs from a causal diagram in that SE models are based on equations having explicit functional forms and estimable parameters. Here we consider a few of the main decisions that must be made when proceeding from a causal diagram to fully specified SE models. Developing candidate SE models for an analysis requires (at a minimum) consideration of (a) the focus of the analysis, (b) characteristics of the available data, including the sample size, and (c) an awareness of the data requirements for various models. There is a decided interplay between these three categories of information and the choices for model specification.

Step 5, part a: Consider the focus of the analysis.—There are at least four general motivations that we can pursue in SEM. One approach to constructing a model is "driver focused." For example, when the goal is to understand the consequences of changes in some driver or drivers for a system, we prioritize model building to maximize an understanding of the most important responses to a driver. For example, we might wish to model the main consequences of variations in wildfire history in the landscape for forest development (Laughlin and Grace 2006). Having this focus influences where one starts with the modeling venture, in this case, with a good characterization of the driver wildfire history. The emphasis then shifts to hypothesizing about and investigating the most significant consequences for the system, thinking of those as cascading consequences in a causal network.

A second possibility for an SEM application is to be "response focused." Here there is a priority for understanding and explaining variation in a particular property of a system. For example, Weiher (2003) has developed a model to understand controls on plant diversity in woodlands. In such cases, the response of interest (including its spatial and temporal properties) is first selected and then theory and data are used to build an understanding of the main factors influencing variations in diversity.

A third possibility is to be "mediation focused." Here, one begins with an observed relationship between variables, for example a correlation between the age of a woodland that is burned in a wildfire and the degree of post-fire recovery (Grace and Keeley 2006). A defining motivation of SEM is the investigation of mediating pathways so as to increase our causal knowledge through the investigation of indirect effects. In fact, one could argue that direct paths in models (this includes all the directed pathways in general linear models) are inherently in need of further evaluation if we are to confirm or uncover the causal connections between two variables. The addition of nutrients to plant communities most commonly results in a decline in diversity (e.g., Gough et al. 2000). What is the cause? We might imagine it is an intensification of competitive interactions. However, SE models have revealed that there are effects that are unrelated to competition, a potential key mechanism being acidification of the soil (Clark et al. 2007). Studies of *mediation* are foundational to the

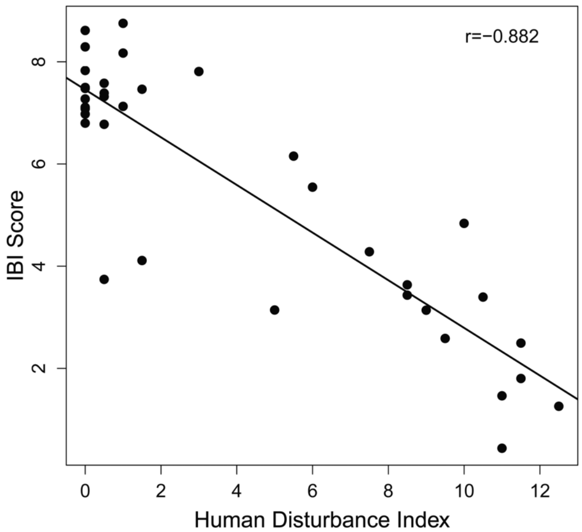

Fig. 3. Bioassessment results for Acadia National Park wetlands (Schoolmaster et al. 2012), expressed as a plot of scores for the Index of Biotic Integrity (IBI) against estimates of the Human Disturbance Index (HDI). The practical goal of the SEM example presented in this paper is to elucidate the causal connections between human activities (disturbances) and the biotic responses identified in the IBI analysis.

investigation of causal relationships (and outside the capabilities of the most commonly used framework for analyzing experiments, the ANOVA model). The objective of the wetland example in this paper, which is to illuminate how elements of human disturbance affect biological responses, is mediation-focused. We start with the observed net relationships between human activities and biotic indicators described in Table 1 and Fig. 3, then we seek to understand the causal connections underlying those net relationships. As an additional example, Gimenez et al. (2012) have developed SE models for understanding the causal pathways whereby environmental factors influence demographic parameters in avian population.

A fourth possible motivation for applying SEM is to evaluate a preexisting theory (i.e., to be "theory focused"). To be focused on the confirmatory testing of a priori theory is a major aspiration for SEM. It is also a key aspect of hypothesis testing when relying on observational data (Grace 2008). Theories that generate network expectations can be tested using SEM methods for model-data consistency. For example, Weiher (2003) performed a confirmatory test of a published model using a new set of data from a somewhat different ecological setting. The results suggested the need to generalize the existing theory to allow it to accommodate a greater diversity of mechanisms for specific circumstances. The introduction of SEMMs into the modeling process is meant to foster the translation and refinement of existing general theories into SE models, and to promote confirmatory testing of theories. As Pearl (2012) has emphasized, achieving statistical model-data consistency is neither sufficient nor even necessary by itself to support causal interpretations. Despite that caveat, evaluations of model-data consistency can be enormously illuminating and they frequently lead to new hypotheses suggested by model-data discrepancies.

Step 5, part b: Examine the available data.—We should aspire in the natural sciences to have SE models in mind when designing studies and deciding what data to collect. This is not always the case, in part because SEM is not part of the traditional training of natural scientists. In this paper we assume the available data may have been collected prior to model formulation. There will certainly be many situations where theory is sufficiently strong that reference to the available data in hand is not required for developing a model, only for estimating its parameters. This is more commonly true for the physical sciences than for the biological sciences, but still true often enough to be a reasonable expectation in some cases. The context of this paper presumes we do not have sufficient theory to proceed to prediction using models obtained purely from reference information.

In the majority of situations, examining the basic characteristics of data can provide vital information for making decisions about model specification. Data characteristics can give us clues about the mechanisms that have shaped those data. What is interesting is that the characteristics of those nodes and the bivariate associations *do not* tell us necessarily what the SE model will reveal about the network relationships. While *A* may correlate positively with *B* and *B* positively with *C* in a network model, the direct path from *A* to *C* can easily be negative (Grace 2006:57). For this reason, examining the characteristics of the variables and the bivariate relationships in an SEM analysis should not be regarded as a "data snooping" enterprise to be avoided. SEM is about what a set of nodes and their linkages represent as a whole. It is important to know what the raw materials are for the proper specification of a model.

It is also important for the application of SEM to understand what data characteristics imply about the needs of model specification. For example, zero-inflated count data may suggest a two-process mechanism associated with organism abundance. If the odds of colonization of a habitat by a species are different from the odds of the organism spreading within a site once colonized, indications of this may exist in the data and one may wish to use a model specification that matches this situation. At the same time, abundance data in samples are often proportion data and such metrics range only from 0 to 1, not from $-\infty$ to $+\infty$. As another example, threshold relationships may be important and estimation of thresholds in SE models may require non-standard specifications. Also, data are often collected in a hierarchical fashion, or some of the explanatory variables may be at a higher level than the response variables. Similarly, spatial or temporal non-independences may influence the data-generating process. Even if these influences are not of theoretical interest, it may be necessary to incorporate them into the model specification to control for their effects on confidence/credible intervals. The accommodations for these complexities that have been developed in statistical modeling (e.g., Gelman and Hill 2007, Zuur et al. 2009) can typically be brought into SE models, though an additional element of care is needed to accommodate the requirements for causal interpretations.

We need to make a practical point about the often-used assumption of a Gaussian distribution for parameters. First, it is important to keep in mind the differences between slope parameters and mean/intercept parameters in models. Slope parameters and residual errors may be Gaussian even if the variables themselves are not. Since SEM studies often emphasize the discovery of linkages (network structure) rather than prediction, the use of the classic linear Gaussian approximation may be more often justifiable than one might think given the distributions of the variables themselves. Generally, modeling response values themselves is more challenging, especially when predicted scores are important for the purposes of the study. There is a continuous increase in the availability of statistical specifications that can accommodate non-Gaussian errors, non-linear relations, non-independence, and random effects (e.g., Gelman and Hill 2007). These can be brought into SE models, depending on the quantitative detail desired, the priorities of the researcher, and the software used for estimation.

Step 5, part c: Consider sample size and appropriate model complexity.—For modeling where statistical inference to a larger population is an objective, a consideration of the number of samples available, and the design used to select sample locations, are first priorities. For cases involving a very large sample size, the challenge can be that a high level of power can detect very minor residual associations between variables and lead to models with very complex graphical relations. One solution to this problem is to emphasize only the larger effect sizes when presenting results and to decide in advance what magnitude of statistical association constitutes a significant scientific finding. On the other end of the spectrum, it is common in ecological studies for sample sizes to be small, especially where the unit of observation is large and complex (e.g., an ecosystem, such as a lake or meadow). In general, careful consideration of the relationship between sample size and model complexity is desirable. It has been suggested that to a degree, final model complexity adjusts to the statistical power because when power is low, only the most important relationships will be reliable enough to require inclusion (Anderson 2008). Despite this, there is merit in deliberately designing models whose complexity is appropriately matched to the available data.

An issue related to the relationship between model complexity and sample size is the choice of estimation method. Based on first principles, not all methods of estimation are equally defensible in small sample cases. The maximum likelihood (ML) estimation procedure used in covariance-based SEM is based on large sample theory. There have been studies to suggest that at small sample sizes, the use of ML in covariance-based SEM can lead to over fitting (Bollen 1989).

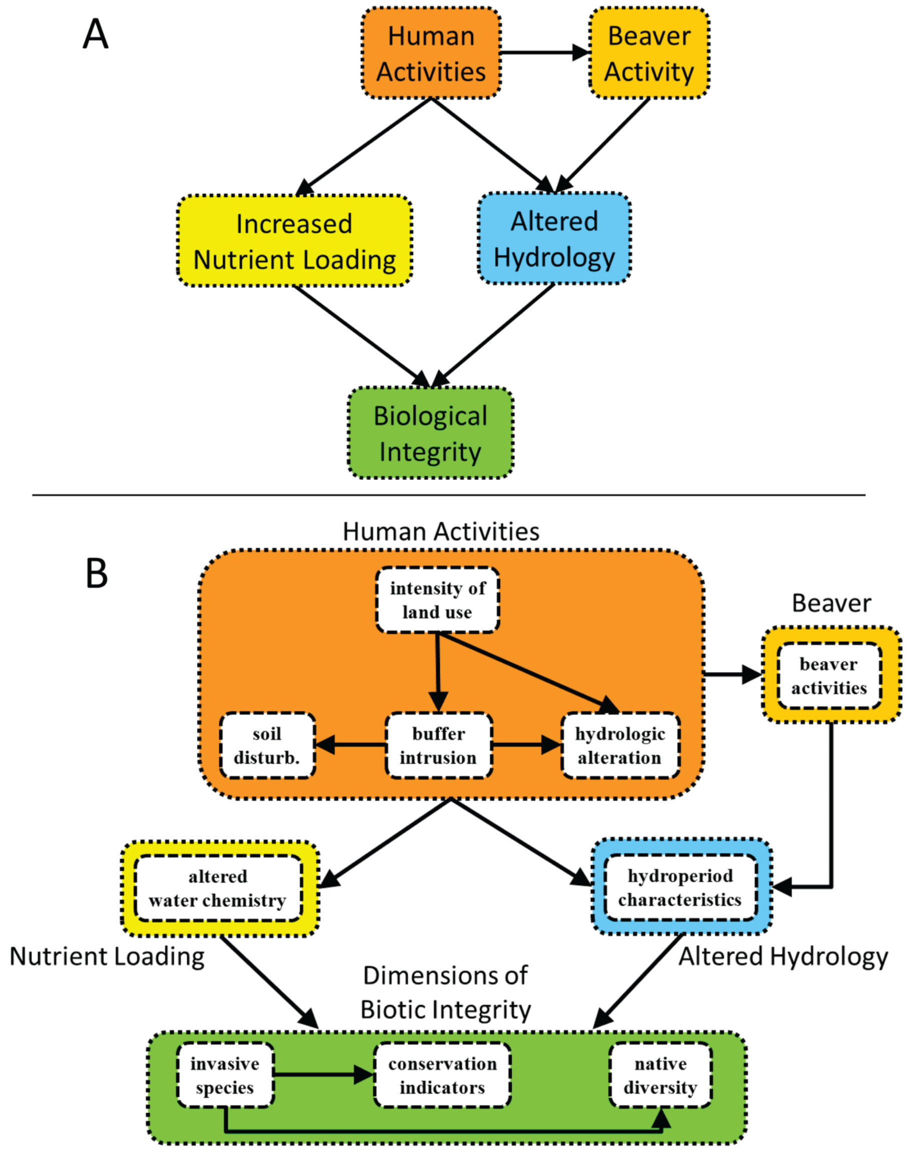

Fig. 4. (A) Upper-tier and (B) lower-tier structural equation metamodels (SEMMs) for the modeling of Acadia wetland plant communities. The nodes being represented are theoretical constructs that are defined (at this point in the analysis process) only at the linguistic level. Nodes are represented using rounded rectangles with dotted outlines to denote the fact that it has not yet been determined how observed, latent, or composite variables may be used to represent them in a SE model. The upper-tier metamodel (A) shows the hypothesized situation at the most general level. The lower-tier metamodel provides a more detailed view of the concepts that will be represented in SE models.

In contrast, Bayesian estimation is based on a finite sample perspective aimed at reducing uncertainty. Lee and Song (2004) have examined the relative performance of ML and Bayesian estimates as small sample sizes in SEM studies. They approached this problem in terms of the parameter d where $d = n/a$ and n is sample size while a is the number of parameters requiring estimation. Because their focus was on the small sample case, they investigated values of d ranging from 2 to 5 using simulation studies and compared ML with Bayesian MCMC esti-

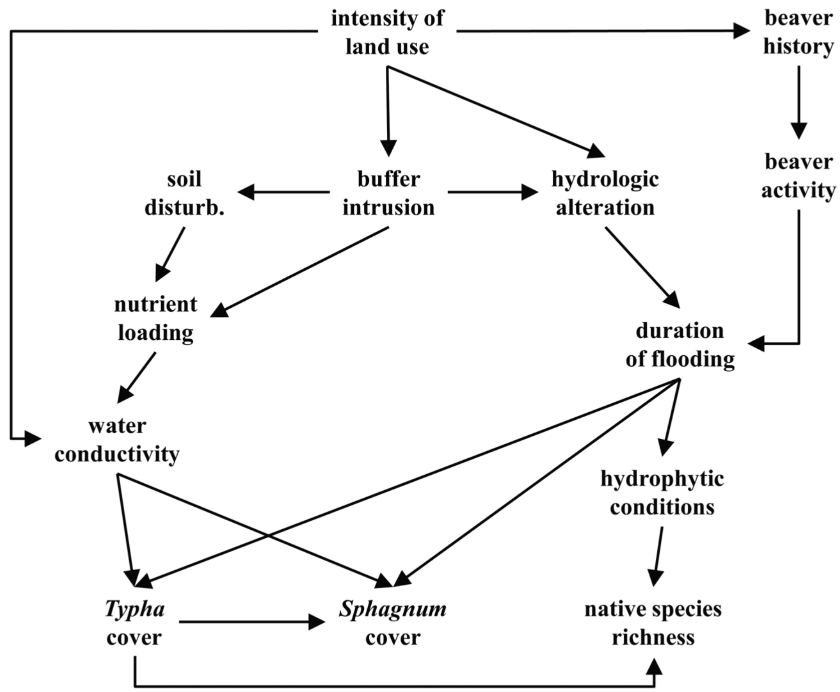

Fig. 5. Initial causal diagram for Acadia wetlands. This diagram is "overspecified" in that it includes nodes that will not be explicitly included in the SE models developed from it. Specifically, the node for "nutrient loading" represents a process for which no measures are available and "hydrophyte selection" represents a process (the elimination of flood intolerant species) that will not be included in the SE models for simplicity. These extra nodes are devices to aid decision making during the process of deciding what measurements need to be taken and what is to be included in the SE models to obtain causal interpretations. The other factors "O_x" that influence variables described in Fig. 2 are not shown here explicitly for visual simplicity.

mation results for two types of structural equation models. Their conclusion was that even at these small information values a Bayesian approach can be defended, while maximum likelihood is noticeably less reliable. This implies that at the very least SEM studies having low values of *d* may wish to use or validate their results against Bayesian estimates.

Step 5, part d: Consider the merits of including latent variables in your models.—Another major decision relates to the question, "Do we need to include latent nodes/variables in the model?" There are persuasive arguments for including latent variables. Latent variables can allow models to better reflect the underlying mechanisms that lead to manifest observations, thereby supporting causal interpretations through this more theoretically-sound representation. In addition, they can be useful in adjusting for measurement error and removing this common source of bias from the parameter coefficients (Grace 2006:80–82). Finally, latent variables provide a way to represent more abstract theoretical concepts (e.g., "biotic integrity"). Despite their appeal, modelers need to be cautious about including latent variables in models because their properties do not always or automatically match causal expectations (Grace and Bollen 2008). While latent variables can enhance the theoretical content of a model, they can also distort the meaning and foster incorrect interpretations as well. This complex of issues is discussed at length in Grace et al. (2010).

Step 6: Selection of estimation method.—Since the early 1970s, the default estimation method in SEM has been the application of maximum likelihood methods to the analysis of covariance relations. The revolutionary idea of working with the covariance matrix instead of the individual observations (Jöreskog 1971) made possible

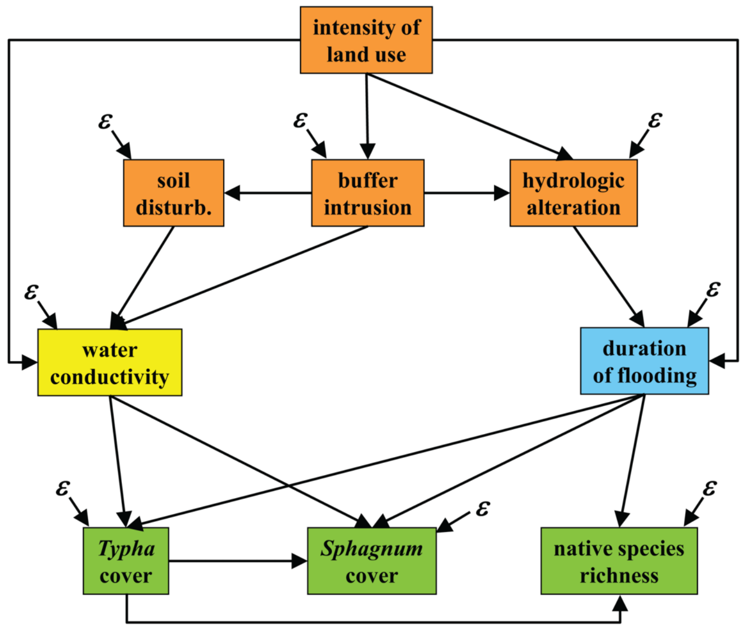

Fig. 6. Initial structural equation model. Color schemes are meant to match variables in the SE model with theoretical concepts in the metamodels (Fig. 4). Epsilons (ε) represent error terms, which connote the influences of factors uncorrelated with predictors on a variable.

estimation for a great variety of model types, in particular those involving latent variables and those with causal loops (nonrecursive relations). In this approach, observed covariances are compared to the model-implied covariances and a maximum likelihood-based minimization process leads to the selection of parameter estimates. Also derived is a measure of the total discrepancy in the model, which provides an overall measure of goodness of fit. This comparison of the observed covariance matrix against the model-implied matrix has the consequence that estimation and model evaluation are based on the entire model. This approach is still the dominant methodology in use today, it is applicable in a great variety of circumstances, and it is implemented in all the major SEM software packages. Its merits include the ability to produce rapid solutions for complex models, a fair degree of insensitivity to collinear relations, and an ability to confirm or reject the entire model. The primary compromise made to permit the elegant solutions provided by covariance-based methods is the assumption of linear Gaussian relations or some generalization of this specification that permits the data to be summarized effectively in matrix form. Over the years, a number of adaptations have been made to permit relaxation of the linear Gaussian assumptions while still relying on the analysis of the covariance matrix (Muthén 1984, Satorra and Bentler 1994, 2001). Today, the analysis of categorical, censored, and count data along with multi-level structures is possible using covariance analysis and approximating methods.

The matrix implementation of second generation SEM has the potential negative consequence of propagating incorrect estimations caused by misspecifications to the whole model (Pearl 2012). Further, the generation of model fit indices that summarize the entire matrix of discrepancies implies to some the potentially erroneous idea that adequate model fit automatically indicates the model meets causal criteria. Local estimation based on nodes and their parents permits an escape from these problems (though it generates a need for the kind of model checking for which the d-separation criterion was developed). What is required for such an approach, however, are estimation methods that are compatible with local evaluation. For path models having no latent variables and no causal loops, most estimation methods can be utilized in a local estimation process (e.g., Shipley 2000*b*). For SE models generally, however, a Bayesian estimation approach is one way to obtain local estimation for the full suite of model types (Lee 2007).

Bayesian estimation has become popular in the past decade in large part due to its implementation via Markov chain Monte Carlo methods (Geman and Geman 1984, Gelfand et al. 1990). Application of Bayesian methods to SEM is now seeing increased use (Ansari et al. 2000, Rupp et al. 2004, Dunson et al. 2005, Arhonditsis et al. 2006, Lee 2007, Grace et al. 2011, Gimenez et al. 2012). A characteristic of a Bayesian implementation of SEM is that the analysis is now examining individual data points at the equation level and not based on the covariance matrix summarization of the observations. This equation-level implementation along with the flexibility in estimation afforded by the MCMC approach now permits much more complex specifications. Bayesian estimation also permits the use of priors to inform the estimation process and to provide data augmentation procedures for use with missing data. As mentioned above, in the case of small-sample analyses, Bayesian methods have been shown to have a significant advantage over the maximum likelihood method (Lee and Song 2004). Bayesian estimation for a very limited set of models is now being implemented in some of the commercial software packages (Arbuckle 2011, Muthén and Muthén 2011). There is a cost to all this flexibility and precision, however, both in terms of training and in implementation. We believe that a fully Bayesian approach to estimating and evaluating SEMs will typically be motivated and justified by some specialized need until more automated procedures are developed.

Step 7: Specify candidate SE models.—The considerations described in the previous sections provide the basis for fully specifying the forms of equations that will be needed to permit the available data to provide valid estimates for the parameters in the SE models. To some degree, model specification will remain a bit of an art because of the great variety of possibilities. As a result, we can expect modeling advice to be in the form of guidelines and an incomplete set of examples. In the example presented later in the paper, we try to present the rationale behind the specification choices made, though we recognize that a complete presentation is not possible in a journal article.

It should be kept in mind that the strength of inference obtained from a modeling effort is influenced by how many models are examined. In highly exploratory applications, a great many models may be evaluated using the available data, generating an elevated opportunity for model selection to be influenced by chance variations in the data. In such situations, confirmation using a second data source, either for the whole model or for individual linkages, is highly desirable (depending somewhat on the size of the sample being examined). Applications involving a limited subset of candidate models can be thought of as comparisons among competing models and tend to inspire more confidence in the repeatability of the findings (all other things being equal).

Step 8: Estimation, model evaluation, and respecification.—At this stage, estimation of parameters is needed in order to evaluate testable implications of the SE model. We do not make a final interpretation of the quantitative estimates until we are confident the graph is consistent with the data. Thus, the process of evaluation involves (a) estimation of parameters, (b) checking of conditional independences to see if important linkages have been omitted, (c) consideration of whether the model is overspecified and can be simplified without loss of information, and then (d) cycling through that process again until model-data consistency is declared.

Details of the estimation options for SEMs with latent variables are described by Lee (2007). The primary methods described include maximum likelihood (ML), generalized least squares (GLS), and Bayesian MCMC solution procedures. In all cases, the estimates obtained can be used as a basis for checking model-data consistency. There are several different ways model-data consistency can be evaluated. In classical SEM, which typically relies on ML or GLS (or even two-state LS) to derive parameter estimates, discrepancies between the model-implied and observed covariance matrices provide indications whether or not "things add up", which is one way of evaluating consistency, usually in terms of a model fit statistic such as the model chi-square. There are

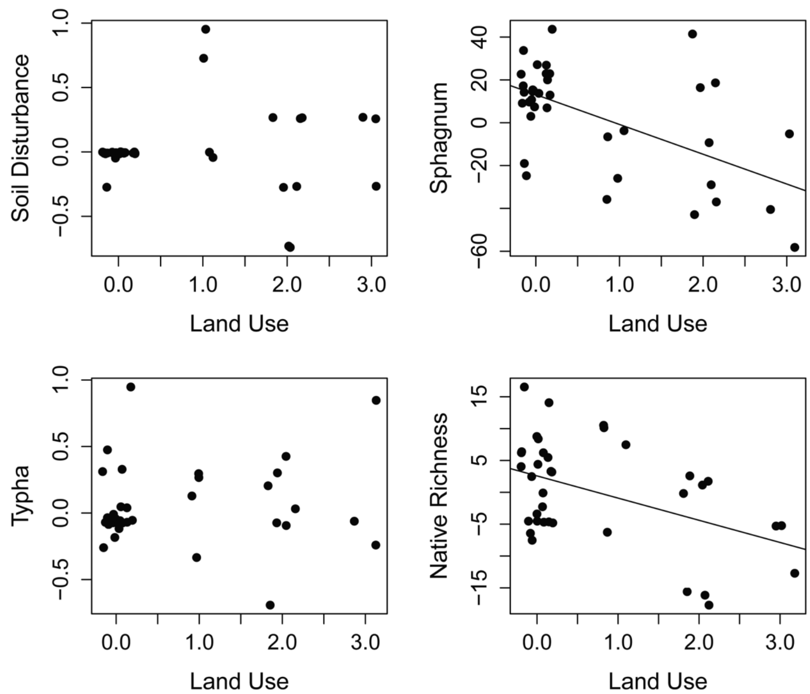

Fig. 7. Scatter plots of residual relationships in the initial SE model (Fig. 6). Residual relationships represent the associations evaluated if a path between unconnected variables was added. All *y*-variables in the plots shown are residuals (deviations from predicted scores based on paths explicitly included in the model). In this particular set of examples, the *x*-variable (Land Use) is categorical, ranging from 0 to 3 and exogenous in the model. For exogenous variables, their scores are not predicted by other variables and, thus, there is no difference between their raw scores and their residual scores.

several key assumptions being made in such evaluations, in addition to the usual statistical assumptions, and it is important to keep them in mind. First, we are assuming that the covariance matrix summarizes the data sufficiently well to avoid leading to incorrect conclusions. Data may be nonlinear or even exhibit discontinuous relationships (e.g., thresholds) and these relationships are not easily summarized in terms of covariances. Second, we are assuming that a measure of overall fit is sufficient to evaluate the linkages in a network. Third, it is sometimes assumed that finding close model-data fit automatically leads to the interpretation of parameters as causal estimates, which is not a generally supportable assumption. When Bayesian MCMC methods of estimation are used, a covariance discrepancy function is not computed and we must find another way to evaluate network-data consistency. Bayesian methods typically rely on quantities such as the deviance information criterion (DIC) or posterior predictive p (PPP) to summarize fit. These quantities are not satisfactory, however, for judging *G* (the graph representing a network in general terms) in any overall way. In the absence of a covariance matrix approach, linkages/edges *E* in the graph *G* must be judged individually, which requires the use of graphical modeling methods (e.g., Koller and Friedman 2009). We illustrate this approach in our example application (e.g., Fig. 7).

Judging the structure of a graph involves both the evaluation of the linkages included as well as the discovery of missing linkages. A graphical modeling approach to evaluating graphs recognizes that linkages can be of many different functional forms. Summary methods such as the computation of partial correlations or implied partial correlations or their derivatives afford limited information and can fail to diagnose nonlinear and discontinuous relationships. For this reason, a more complete inspection of residual relationships can be desirable. Committing to a search for missing linkages in the graph that allows for nonlinearities is currently a less streamlined process compared to that associated with classical implementations. On the other hand, the direct examinations required for complete evaluations are subject to fewer statistical assumptions. We use a graphical modeling approach to evaluating the graph in our example presentation by first evaluating the form of the residual relationships visually. It should be noted that classical analyses using covariance-based procedures permit double-checking of the network using graphical modeling methods, while piecewise solution methods such as Bayesian MCMC require them.

In our experience, respecification is often required, since many of the models initially considered fail to demonstrate model-data consistency. This is normal and constitutes a valuable learning process (one that is largely missing from classical statistical analyses using non-network models). The most common types of respecifications include (a) inclusion of additional directed pathways, (b) removal of direct pathways, (c) evidence for unanticipated correlations, either between variables or between their errors, and (d) changes in the functional forms used for linkages, such as those needed to accommodate nonlinear relations. Once a model has been provisionally accepted as the best model for the situation (provisional until additional data are obtained or additional diagnostics are performed), it is often useful to perform sensitivity tests on the model to ensure that there are no major nonindependences remaining in the model. Failure to include important missing relationships may, depending on the situation, introduce bias into the estimates.

Step 9: Discovery, quantities, and queries.—SEM analyses produce a variety of types of results. One interesting and important class of results relates to the topology of the graph (linkages and possible simplifications). These are particularly interesting because each pathway in a graph is generally thought to represent a separately distinguishable mechanism. Qualitative conclusions about the graph have implications that relate back to the causal diagram. Fig. 2 states that there are two main mechanisms whereby D_1 affects R_1 and one of the mechanisms operates through M_1 while the other operates through M_2. Of course we know that behind qualitative conclusions are quantities about which decisions have to be made (e.g., declarations of significance). These decisions are generally made in the previous step (*Model evaluation*), but their consequences permeate the rest of the process. If we decide that a direct path from D_1 to R_1 should be included in our graph, we may have discovered a new process if we previously had no expectation for such a connection. The discovery of evidence for new mechanisms via the identification of missing links is one of the most exciting aspects of SEM. Similarly, model comparisons that lead to conclusions about mechanisms supported through the inclusion of a hypothesized linkage are notable and often emphasized in SEM studies.

Aside from being interesting, conclusions about the topology of the graph are important for the sufficiency of the other results obtained. Deciding that a direct link between D_1 and R_1 should be included not only has logical implications (e.g., concluding that variations in D_1 that do not produce changes in either M_1 or M_2, will still have an effect on R_1), it also influences at least some of the quantitative estimates of the effects along the directed pathways connecting D_1 with R_2. This is the reason that one should not seek to interpret the quantitative values of effects until one has ensured the graph is supportable. As mentioned earlier, estimates of causal effects depend on the correctness of the graph—not necessarily the entire graph, but at least the part of the graph relating to the causal effects of focal interest.

An extension to the question of whether all important links are included in a diagram is the question of whether we might decide that some can be removed. We may anticipate processes represented by linkages in our diagram that do not turn out to be important in a particular study. This affords us the opportunity to simplify our graph by removing unimportant linkages (thereby setting the estimates of their quantitative effects to zero). Sometimes this simplification process can even allow us to remove nodes (variables) if their variations do not appear to have any consequences for other nodes. This is not the same as another kind of simplification sometimes called "node absorption" (Kjaerulff and Madsen 2008) in which a model is simplified by removing a node, not because its effects are unimportant but in order to remove it from overt consideration while absorbing its causal influences into the pathways passing through it.

Quantitative effects are often of major interest in SEM studies. Proportion of variance explained is often one example. While strongly dependent on the sampling scheme and not strictly a measure of invariant properties of the system being studied, estimates of variance explanation are often of interest in judging success in modeling variations among samples. Relative magnitudes are also commonly of interest. For example, there is currently keen interest in the relative importance of top-down versus bottom-up control in food webs, in part because of the implications for environmental protection needs (Riginos and Grace 2008, Byrnes et al. 2011). Grace and Bollen (2005) discuss some of the fundamental issues with standardized parameters and suggest a new approach where coefficients are standardized by ranges instead of standard deviations. For models containing linkages that are not simply Gaussian and linear, obtaining standardized coefficients that permit comparisons across paths requires additional considerations. This relates to the general problem for models with complex specifications that we may need to use results from queries in place of conventional quantities in order to make the needed interpretations.

The utility of queries in SEM relates to the fact that the implications of quantitative results are not always simply conveyed. We illustrate below ways that queries can be used to summarize relative sensitivities in models as one example of their use. Also, models will not always be based on simple linear functions. How does one summarize an effect when it is nonlinear? Using queries allows us to pose scenarios that permit illustration of the implications of the parameter values as well as alterations of those values. For complex nonlinear models it can be quite helpful for the reader if the investigator conveys the main points and most important findings by posing queries. Further, an emphasis on queries should become common practice as we place more emphasis on prospective reasoning from SEM results.

There are at least four kinds of queries. Queries that relate to future states given current conditions are predictions. These are statements about what value of Y we would expect to observe if in the future we observed a particular value of X (for the case where X is a cause of Y). An important alternative query about future observations is the intervention. Here we ask what value of Y we would expect if we were to cause X to change to a particular value x. Retrospective queries about what led to the currently observed states can be thought of as attributions. In the ecological example of wetlands that follows, we will use scenarios to summarize evidence about the relative importance of different influences that contributed to the observed responses. Finally, another query about the past is the counterfactual. A counterfactual claim refers to the retrospective question, "What would have happened if ... ?" For example, we might ask, what do our model results imply would have happened if a particular individual had been in the control group rather than in the test group. Thus, we can see that queries will play important roles in both retrospective and prospective considerations of our model and its parameter estimates.

Step 10: Report methods, findings, and interpretations.—It should be remembered when reporting on an SEM analysis that the description of methods should strive to be sufficiently complete that an analysis is repeatable, or at least reproducible. Bivariate relations are also informative, even if their relation to causal interpretations are not straightforward in many cases.

We should always remember in causal modeling that our inferences depend fundamentally on the assumptions embodied in the graph. The language one uses to convey their findings and interpretations should likewise reflect this conditionality. It is worthwhile declaring an awareness of the conditional nature of claims early in the paper. Then when one starts describing things like "total effects" in the Results section of a paper it will be understood by the reader that the author is not naive about the requirements for causal inferences.

Reporting estimated parameter values and expressions of uncertainty about them is a minimal requirement. Simply showing SE models as graphs with relative quantities expressed through the thicknesses of arrows is not sufficient. Unstandardized (raw) estimates of parameter values are fundamental quantities and should be reported. Depending on their interpretability for the situation (e.g., if complex transformations to the data have been applied), unstandardized results may be relegated to an appendix. Investigators often chose to describe their findings using standardized expressions of parameter estimates. The interpretation of these coefficients is subject to certain caveats, of which the investigators should be aware (Grace and Bollen 2005). We will show how queries can be used as an alternative means for generating measures of relative sensitivity.

Results

Illustration: The wetlands of Mount Desert Island and Acadia National Park

The data.—Acadia National Park is located on Mount Desert Island on the coast of Maine, occupying a major portion (roughly two-thirds) of the island, but with substantial private lands within its authorized boundary (Tierney et al. 2009). Mount Desert Island is a 24,000 ha granite bedrock island and includes the highest mountain on the Atlantic coast of the U.S. As a consequence of its mountainous topography, wetlands on the island are in relatively small catchments. The soils are shallow in the uplands while the wetlands are often peat-forming, receiving their water largely from acidic and low-nutrient inputs from rain and surface runoff (Kahl et al. 2000).

For this analysis, we used data from recent studies by Little et al. (2010) and Guntenspergen (*unpublished data*) on human activities and wetland plant communities on Mount Desert Island. In these studies, 37 nonforested wetlands were examined as part of an assessment of the relationships between degree of human development, biological characteristics, hydrology, and water quality. We added to the available information by quantifying degree and type of specific human disturbance activities in the vicinity of wetland catchments using a modification of the rating system developed for wetlands by the Ohio EPA (Mack 2001). This resulted in the ordinal rating of (1) the degree of human development in the watershed, (2) the degree of alteration of the natural hydrology, (3) human intrusion into the buffer zone (for a review of this subject, see Castelle et al. 1993; here we simply measured whether human developments came within 50 m of the edge of the wetland and if so, exactly how close.), (4) soil disturbance, and (5) habitat alteration. We then constructed a human disturbance index (HDI) for Acadia wetlands and used the HDI to select biological characteristics of the plant communities that represent sensitive responses to human disturbance (Schoolmaster et al. 2012). After screening metrics from a large set, a list of seven (Table 1) biological characteristics was chosen as candidates for constructing an index of biotic integrity (IBI) (Karr 1981, Karr 1991, Barbour and Yoder 2000). Below we use the guidelines presented in Fig. 1 to develop a SE model so as to provide an interpretive structure for the net relationship shown in Fig. 3. Step 10, which deals with reporting methods, findings, and interpretations will not require recapitulation and will be skipped in the example.

Step 1: Defining the goals of the analysis.—The analysis presented here has two goals. The first is to learn about the different mechanisms by which human activities on Mount Desert Island might impact the natural properties of wetlands. Starting from the results produced from a bioassessment of the wetlands (Table 1), we are motivated to understand the mediating factors and processes connecting the degree of human development to biological changes. The second goal of this study is to specify our SE model(s) in such a form that they can permit queries that will potentially guide the protection or restoration of wetlands in Acadia National Park. In this analysis, we capitalize on the range of disturbances on the island so as to perceive threats and consider remedies that can be applied to the wetlands with protected status that occur within park boundaries. This goal has the following implications: (1) Our system of primary interest is the wetlands in and around Acadia NP and we wish our results to generalize to that set. Inferences to other wetlands elsewhere is a (potentially large) side benefit. (2) While the sample of 37 wetlands is small in absolute terms, we note that the universe of inference, the complete set of wetlands in the park, is estimated to be somewhere around 1,600 (Guntenspergen, *unpublished data*). Additionally, we realize that the wetlands sampled may become sentinels for the system, thus they will in the future play a role as indicators. Detecting significant changes in these indicator wetlands becomes one inference problem and, of course, the extrapolation of conclusions from those wetlands to others another. (3) A third implication of the applied intent of this study is that it would be helpful if the analyses produced results and implied predictions in a form that can be directly compared to observed quantities. This motivates us to use realistic model specifications of counts, proportions, and thresholds so that predicted scores are directly comparable to present and future observations.

Step 2: Development of a structural equation metamodel.—In this step, we wish to begin the conversion of our linguistic understanding into formal concepts and relationships. In this case, the context is that human activities may lead to changes in environmental conditions that reduce biotic integrity in wetlands. A key question within this context is, "What are the mediators of those effects?" Knowing the mediators between human activities and biological changes is not only important for our understanding, it also may provide opportunities for interventions that could reduce adverse effects. Among the many possibilities, human activities are known to commonly alter both the hydrology and water quality in wetlands. For example, in the Florida everglades both alterations of hydrology and inputs of nutrients have been implicated as the two most important drivers of change in that system (Newman et al. 1998). Of course human activities are not confined to just the land surrounding wetlands but can occur in wetlands in the form of haying operations, cattle grazing, and prescribed burning. In the case of the wetlands of Acadia National Park, such direct physical disturbances are not known to be a current threat to the wetlands because of their protected status, so such physical disturbances are not considered here. In addition to human activities, the natural influences of beaver must be recognized (Little et al., *in press*). An additional construct included is the broad class of influences we refer to as environmental covariates. Covariates such as these can both mask and exaggerate the apparent effects of human activities (Schoolmaster et al. 2012). Explicit connections among five of the constructs are shown in the construct model (Fig. 4). The connections omitted (e.g., directly from human activities to biological integrity) represent the optimistic hypothesis that the important mediators will be represented by the variables available. Note also that environmental covariates are included in the figure as a cautionary reminder to consider the influences of this class of factors when specifying the causal diagram and SE models. We point out that the SEMM could be developed at a somewhat more detailed level, as long as the constructs are defined as theoretical/hypothetical entities. Once a model is specified at the level of variables, we are now at the level of the causal diagram, or if complete statistical specifications are included, the SE model.

Step 3: Development of a causal diagram.—A causal diagram for the Acadia wetlands is shown in Fig. 5. This represents a partial specification of the ideas in the SEMM based on the variables that might be included in SE models (Table 2). Regarding the specification of constructs, the construct related to human activities is clearly one that is multidimensional. The methods used to develop an index of human disturbance included quantifying the degrees of (a) human development, (b) alterations of wetland hydrology, (c) intrusion into the buffer zone around the wetland—i.e., how close alterations come to the water's edge, (d) soil disturbance adjacent to a wetland, and (e) habitat alteration (e.g., tree removal). Since habitat alteration was not seen as likely to have a direct causal effect on plants in the wetlands, this metric was not included in the diagram. For the others, a set of hypothesized causal linkages are represented. These hypothesized linkages represent causal assumptions that are embodied in the diagram. We can additionally consider these to be hypothesized relations that have testable implications and can be examined once structural equation models are specified, estimated, and evaluated.

Regarding the representation of the construct called biotic integrity, we selected three key metrics to represent dimensions of integrity (Fig. 5) based on their unique (nonredundant) information content about impacts of human activities (Table 1). Of the biotic responses the one ranked most uniquely indicative of human disturbance based on multimetric index analysis

Table 2. Theoretical constructs (see Fig. 4) and available data related to those constructs.

Theoretical constructs	Observed variables related to construct	Properties of variables
Human activities	intensity of land use	very low to very high; (0, 1, 2, 3)
	intrusion into the wetland buffer	buffer > 50 m to buffer < 10 m; (0, 1, 2, 3)
	hydrologic alteration	none to highly altered; (0, 1, 2, 3)
	soil disturbance	none, recovered, recovering, recent; (0, 1, 2, 3)
	habitat alteration	none, recovered, recovering, recent; (0, 1, 2, 3)
Biotic responses: ecosystem eutrophication	*Typha* cover	0–100%; semicontinuous
	total forb cover	0–100%; semicontinuous
	maximum vegetation height	0–max. value; continuous
	total dicot cover	0–100%; semicontinuous
	total perennial cover	0–100%; semicontinuous
	total native plant cover	0–100%; semicontinuous
	total fern cover	0–100%; semicontinuous
	total tree cover	0–100%; semicontinuous
Biotic responses: *Sphagnum*	total *Sphagnum* cover	0–100%; semicontinuous
Biotic responses: plant species diversity	monocot richness	0–maximum; counts
	native richness	0–maximum; counts
	forb richness	0–maximum; counts
	tree species richness	0–maximum; counts
	dicot richness	0–maximum; counts
	shrub richness	0–maximum; counts
	fern richness	0–maximum; counts
Mediators: water quality	surface water conductivity	continuous
	surface water pH	continuous
Mediators: hydrology	average water depth	$-\infty$ to $+\infty$; continuous
	SD of depth	continuous
	maximum depth	continuous
	minimum depth	continuous
	time soil surface flooded	0–365 days; proportional count
	time water below 30 cm	0–365 days; proportional count
Covariates	beaver	no sign, abandoned, active

(Schoolmaster et al. 2012) was the cover of cattails (*Typha* spp.). The simple bivariate correlation with the human disturbance index (HDI) was the strongest of any metric examined (0.71). Further, the presence and abundance of this plant in Acadia wetlands is of management concern because it is another highly ranked responder to human activities. Cover of this species serves as an indicator of the status of peatlands, a community type of special conservation value. Cover of *Sphagnum* possesses the second strongest simple bivariate correlation with HDI (−0.69). A third biotic response, the richness of native species, was selected as a sensitive indicator of highly disturbed wetlands.

There are several possible variables that could represent the two key mediation processes hypothesized to be important (Table 2). One of the hypothesized mechanisms that could lead to increases in plant production and *Typha* invasions, as well as a reduction in *Sphagnum* and indirectly a loss of diversity, is an increase in mineral nutrient concentrations in the water. While direct measures of nutrients were not available, for systems such as these that are naturally very low in mineral concentrations (because of the granitic nature of the bedrock), water conductivity can serve as a useful surrogate. Water pH could also be considered as an indicator of the mediating process. The other mediating mechanism hypothesized to be in operation is variations in water levels. Several hydrology variables were measured in the course of the study of these wetlands.

Step 4: Exposition of causal assumptions and logical implications from the causal diagram.—Box 2 provides a summary description of the theoretical basis (in the form of 16 causal assumptions and hypotheses) for interpreting the linkages included in the initial causal diagram (Fig. 5). In the absence of physical interventions (e.g., through experimental manipulations), the directed linkages represent untestable assumptions/implications, though they can also be thought of as predictions of the consequences of future manipulations. The assumption of sufficiency

Box 2

Causal assumptions/hypotheses embodied in the initial causal diagram (Fig. 5).

I. A high intensity of human development in the watershed of a wetland has several implications:
   (1) It potentially increases the probability that some activities will occur very close to the edge of the wetland (buffer intrusion).
   (2) It increases the probability that flows of water will be impacted. This is potentially a complex expectation. If alterations are primarily related to roadways and flood control, we expect water levels to be stabilized via a set of impoundments, culverts, and drainage ditches. This appears to be the nature of the hydrologic alterations for the majority of wetlands sampled, based on the pattern of association between degree of alteration and increased stabilization of water levels.
   (3) It increases the potential for soil disturbance adjacent to the wetland, but only when there is intrusion into the buffer zone.
   (4) It can be expected to affect the activities of beaver. One might expect an avoidance of heavily developed areas by beaver as a general expectation. As with any wild mammal that coexists with humans, its behavior reactions can be complex and there is some indication that beaver are attracted to some human-caused structures.

II. Only a limited variety of water quality variables were measured in the bioassessment of wetlands. These included electrical conductivity, a measure of total solutes, and pH. Of these two highly correlated measures, conductivity levels are a more general indicator of conditions important to plant communities in a soft-water landscape. The diagram anticipates several causal effects on conductivity effects.
   (5) Conductivity can be expected to be higher in wetlands with adjacent soil disturbance because loose soil easily transports solutes and nutrients.
   (6) Even when human developments do not intrude into the buffer zone around a wetland, human activities generate high nutrient loads generally and these can easily pass into wetlands via culverts and drainage ditches. Thus, there is likely to be increased conductivity in wetlands with high levels of human development independent of mediations through buffer intrusion and soil disturbance.

III. The hydrology of a wetland is influenced both by surface flows and ground water flows. Both of these may be impacted by human activities. Hydrologic alterations include roadways and dams that impede water flows as well as drainage ditches, culverts, and overflows that direct the flow of water. These are often added in combinations that are designed to stabilize hydrology. Ditches and tiles designed to drain wetlands are common, but apparently in the wetlands involved in this study such historical influences are not evident. Related to Assumption 2, it is our assumption in the initial causal diagram that if hydrology has been altered by human activities, that will be reflected in the metric "hydrologic alteration."
   (7) We hypothesize that when human activities intrude into the buffer zone, there may be additional hydrologic alterations added.

IV. The investigation of Acadia wetlands included measurement of hydrology for a subset of the wetlands. It was possible to benchmark these to average plant hydric affinity and then use missing data imputation methods to create a complete set of estimates for certain hydrologic variables. The hydrologic measurement most tightly associated with plant hydric affinities was the proportion of time the wetland was inundated (duration of flooding). The diagram assumes the following:
   (8) Degree of hydrologic alteration leads to stabilization of water levels and a longer duration of flooding.

Box 2. Continued.

(9) Beaver, which are famous as ecosystem engineers, also act to increase duration of flooding, both by increasing the average water depth in a wetland and by impounding flows. When beaver abandon a wetland, however, there can be an abrupt drop in water levels and shortening of hydroperiod.

V. *Typha* invasions have been studied extensively. While our predictive understanding is still incomplete, two things are well documented and embodied as causal assumptions in the diagram:

(10) Increased nutrient loading and increases in water conductivity increase the probability that *Typha* species will invade a wetland and also increase the abundance of *Typha* when they do colonize.

(11) Increased flooding duration promotes both *Typha* colonization and population development, at least up to a point (Grace 1987).

VI. *Sphagnum* peatland communities are special community types, generally being of high conservation value, in part because they require biotic control over the abiotic conditions. Low conductivity water sources are generally necessary, as are unbuffered, low pH inputs. Peatlands are also dependent on stable water levels and can be reduced both by drainage and excess flooding. The causal diagram permits the following:

(12) Increased conductivity will be detrimental to *Sphagnum* abundance.

(13) Increased flooding will lead to reductions in *Sphagnum*.

(14) High levels of *Typha* will crowd out *Sphagnum* over successional time.

VII. Plant species richness is highest where water levels are shallow and water levels fluctuate. The diagram assumes the following:

(15) Species richness of all groups, as indicated by forb species richness, will decline with increased duration of flooding.

(16) High levels of *Typha* will lead to a build up of plant litter and reduced germination levels, eventually resulting in reduced diversity of forbs.

for the included links leads to expectations of conditional independence among nodes not connected by direct linkages. This extends to a prediction of uncorrelated errors for the variables. For the causal diagram in Fig. 5, there are 45 possible links, 16 included links, and 29 implied conditional independences that constitute testable implications. These causal claims can be made prior to deciding on the specification of functional forms for the structural equations. That said, our ability to properly evaluate the evidence for linkages and omitted links does depend on our statistical specifications.

In this real-world application, we believe we have most of the measured variables needed to be able to identify the majority of effects implied by the causal diagram. One effect of concern because of limited data is the effect of beaver activities on duration of flooding. We are also concerned about beaver effects on biotic measures, even though we don't hypothesize those as direct effects in our initial causal diagram. If we develop a SE model that omits measurements of beaver activities, what are the options available? If the causal diagram in Fig. 5 is correct (and it may not be), we can develop a model that is a reduced form of the diagram and include a direct path from intensity of land use to duration of flooding. This path would represent a model simplification that absorbs hypothesized beaver effects into a net direct effect. The causal diagram helps us to describe and interpret this node absorption.

Any confidence we may have in our initial causal assumptions and the adequacy of our ability to isolate causal effects can be altered by unexpected findings resulting from the estimation and model evaluation process. We anticipate, based on experience, a good chance that our causal diagram will need to be altered based on empirical findings before a final SE model is usable for inference about the system. This should be considered normal for first time applications of causal modeling in many settings. For this reason, it is useful to explicitly emphasize that the causal assumptions that are not testable using observational data should be treated as predictions that might be evaluated given future interventions.

Step 5: Evaluation of specification options for the SE models: (a) Focus of the analysis.—The modeling effort here is "mediation focused" in that we wish to investigate the causal network of connections between human development, mediating changes in the environment, and biotic responses. Included is the ambition to arrive at a plausible understanding of how different aspects of human activities are interconnected. Also, we wish to consider whether biotic responses are independent or are causally linked. The causal diagram expresses the hypothesis that sufficient cattail cover may lead to reductions in *Sphagnum* and native species richness. The two primary mediator variables, conductivity and flooding, are hypothesized to convey the complete effects of human and beaver-caused activities on biotic responses. It is not expected that this simple result (which can be referred to as complete mediation) will necessarily be supported by the data. Additional influences of human disturbance on biotic conditions that do not pass through our two mediators are certainly possible.

(b) Consideration of the available data.—Data characteristics have important influences on model specifications. Properties of the variables that were considered when developing the causal diagram from the SEMM are given in Table 2. Examinations focused on (a) scales of measurement, (b) distributions of values, missing values, and outliers, and (c) bivariate relations with other variables included in the diagram.

(c) Consideration of appropriate model complexity.—In small sample studies, there can be a strong tension between ambitions and the need for simple models. A small number of samples will only permit reliable estimation of a limited number of parameters. The degree of confidence can influence decisions about model complexity as well. When there is strong theoretical knowledge, smaller samples can be tolerated. When uncertainty is high, larger samples are needed. Decision rules in this case are not hard and fast. In our wetland study, we felt it was very important to keep the model as simple as possible while accomplishing our most important objectives.

In striving to manage model complexity, we decided to include only a minimum number of variables in our SE model. Regarding human activities, we eliminated the variable for the degree of habitat alteration from inclusion in the modeling process because, while of importance to wildlife, it does not seem essential to understanding the three key biological responses being modeled (or at least does not appear to contribute any unique understanding of effects on the key biological responses). The other four variables that were components of the HDI appear to be sufficient measures of the nodes in the causal diagram and necessary to include for our purposes. We also chose to model only three key biotic responses, *Typha* cover, *Sphagnum* cover, and native species richness. To represent the eutrophication mediation process, we selected water conductivity over pH. These two indicators are highly correlated and among the two, conductivity is more readily interpretable in this case (since *Sphagnum* is known to produce organic acids, and can thereby have a causal effect on water pH). It is also known that pH can fluctuate on a diurnal basis due to metabolic influences (Wetzel 2001). Based on the strength and linearity of bivariate relations with a measure of community hydric affinity, we selected the duration of flooding (proportion of time soil surface inundated) as our indicator of critical hydrologic alteration.

Since sampling was not conducted relative to known environmental gradients, important covariates to control for were not obvious. Because covariates were not explicitly included in the model, examination of correlations among errors in the SE model is particularly important for finding their implicit effects and to avoid backdoor relationships that might bias the estimation of causal effects (Pearl 2012).

A measure of beaver activity was considered for inclusion in SE modeling. We chose to omit it in this first analysis for two reasons. One, the

Box 3

Equation specifications for final SE model. Variables were assigned codes and numbers for denoting equation coefficients as follows: land use, surr (var 1); buffer intrusion, buff (var 2); hydrologic alteraction, hyd (var 3); soil disturbance, soil (var 4); flooding duration, flood (var 5); water conductivity, cond (var 6); native species richness, natr (var 7); *Sphagnum* cover, sphag (var 8); *Typha* cover, typha (var 9). R code for implementing the final model is given in the Supplement.

Buffer intrusion:
buff[i] ~ dbin(buff.hat[i],n.buff)
logit(buff.hat[i]) ← b2.0 + b2.1*surr[i]

Hydrologic alteration:
hyd[i] ~ dbin(hyd.hat[i],n.hyd)
logit(hyd.hat[i]) ← b3.0 + b3.1*surr[i]

Soil disturbance:
soil[i] ~ dbern(soil.hat[i])
logit(soil.hat[i]) ← b4.0 + b4.1*buff[i]

Flooding duration:
flood[i] ~ dbin(flood.hat[i],n.wet)
logit(flood.hat[i]) ← b5.0 + b5.1*hyd[i] + b5.2*surr[i]

Water conductivity:
cond[i] ~ dnorm(cond.hat[i],tau.cond)
cond.hat[i] ← b6.0 + b6.1*soil[i] + b6.2*surr[i]

Native species richness:
natr[i] ~ dpois(natr.hat[i])
log(natr.hat[i]) ← b7.0 + b7.1*flood[i] + b7.2*surr[i]

Sphagnum cover:
sphag[i] ~ dbin(sphag.hat[i],n.sphag)
logit(sphag.hat[i]) ← b8.0 +b8.1*cond[i] + b8.2*flood[i] + b8.3*hyd[i] + b8.4*soil[i]

Typha cover:
typha[i] ~ dnorm(typha.hat[i],tau.typha)
typha.hat[i] ← b9.1*cond[i] + b9.2*step(cond[i]-psi.typha)*(cond[i]-psi.typha)

measures of beaver activity are complex and not reducible to a single variable. Some sites have not had beaver in recent years, others have active beaver, and a third group has had beaver abandon their lodges at some (various) time in the past several years. To avoid slipping into an unhelpful level of model complexity, we chose a simpler modeling option. Based on the causal diagram, we represented the hypothesized causal chain of land use to beaver to flooding using a direct arrow from land use to flooding. Thus, our SE model can be considered to be a reduced-form model relative to the causal diagram.

(d) Consideration of the merits for including latent variables in the SE model.—It could be useful in bioassessment modeling to include latent variables to represent biotic responses more generally. Eutrophication and hydrologic alterations can be expected to invoke a suite of biotic responses. That said, the small sample size in this study motivates us to keep our SE model as simple as possible and include only a minimum number of biotic responses. We relied on results from a multimetric indicator screening method that identifies the minimum set of metrics to represent human disturbance effects (Schoolmaster et al. 2012). As a result, we were able to select three key responses to the effects of human activity, thereby making it easier to keep our model simple. For the purposes of illustrating a broader range of modeling options, we ignore our own cautions about maintaining the simplest model

structure later in the paper when we return to consider latent variable modeling possibilities.

Step 6: Selection of estimation method.—In our evaluation of Acadia wetlands, we chose to use a Bayesian approach based on MCMC procedures implemented in WinBUGS via the R program (R Development Core Team 2008). Two major motivations drove this choice. First, the number of samples available for Acadia wetlands is small relative to the needs of matrix-level methods. For the initial model examined (Fig. 6) 16 pathways are specified, giving a ratio of samples to paths of 2.3. For maximum likelihood estimation we would prefer a minimum of five samples per parameter in most cases (though an even higher ratio would be desirable). Second, the objectives of our study include probabilistic reasoning and prediction. For the estimation of path relations, the linear methods and their extensions that are available in commercial software packages may be suitable approximations. However, predicting changes in the posterior distributions of quantities, such as *Typha* abundance, may benefit from greater flexibility in specification choices.

Step 7: Specification of candidate SE models.—Having decided on the graph (G) to use in our initial SE model (Fig. 6), the next task is to specify the functional forms for the linkages and the response forms for the variables (F). Our purpose in this paper is to present general modeling guidelines, not to develop an exposition on statistical forms. For this reason, we simply summarize the functional forms in Box 3 and provide references to the less common ones. For a general treatment of the subject, one can consult (Gelman et al. 2004, Gelman and Hill 2007, Ntzoufras 2009). Our SE model includes eight *endogenous* variables, and requires eight distributional descriptions, one for each response variable. The three endogenous disturbance variables, soil disturbance, buffer intrusion, and hydrologic alteration, were all estimated based on professional judgement. As is typical for such measurements, the data are rank order categorical responses, in this case with four levels (0–3). Despite their similarity as ordinal measures, detailed examination of the data and the relationships with their exogenous driver, intensity of land use, led us to model them in different ways. The response forms for human disturbance variables were as follows: buffer intrusion was modeled as a proportional odds (Ntzoufras 2009, Agresti 2010), as was hydrologic alteration, and soil disturbance was modeled as a binary Bernoulli response. Duration of flooding was modeled as a proportional odds response, water conductivity as a log-linear response, *Typha* as a threshold response with a single changepoint (Muggeo 2003), *Sphagnum* as a proportional odds response, and native richness as a Poisson response.

Step 8: Estimation, model evaluation, and respecification.—As stated above, estimation was performed using WinBUGS (Lunn et al. 2000) implemented through the R package R2winbugs (Sturtz et al. 2005). Implied conditional independences in the SE model were examined graphically as an initial step. To accomplish this, residuals were calculated and plotted. A few select results are shown in Fig. 7 to illustrate the process. In Fig. 7 we show the relationships observed between residuals for four of the endogenous variables and the exogenous variable Land Use. Measures of association were calculated for these relationships to quantify the linear relationships. However, the ultimate decision whether to include additional linkages in a revised model was based on evaluation of parameter significance for included links. Using this approach, evidence to support the inclusion of several new paths in our SE model was found. These new paths are highlighted in Fig. 8, the revised causal diagram, by showing them as dotted lines.

With paths added and the model reestimated, no indication of additional missing linkages was found. This included an evaluation of residual correlations among the three biological response metrics. At this point, all parameters were evaluated to determine whether there was a basis for model simplification (i.e., whether some processes/links were ignorable). It was determined that four paths in the model could be eliminated without loss of explanatory power, (1) the link from buffer intrusion to hydrologic alteration, (2) the link from duration of flooding to *Typha* cover, and (3) the links from *Typha* cover to *Sphagnum* cover and (4) native species richness (Fig. 8). R code showing the implementation of the final model is given in the Supplement.

Step 9: Discoveries, quantities, and queries.—We refer to the revised SE model as the provisional

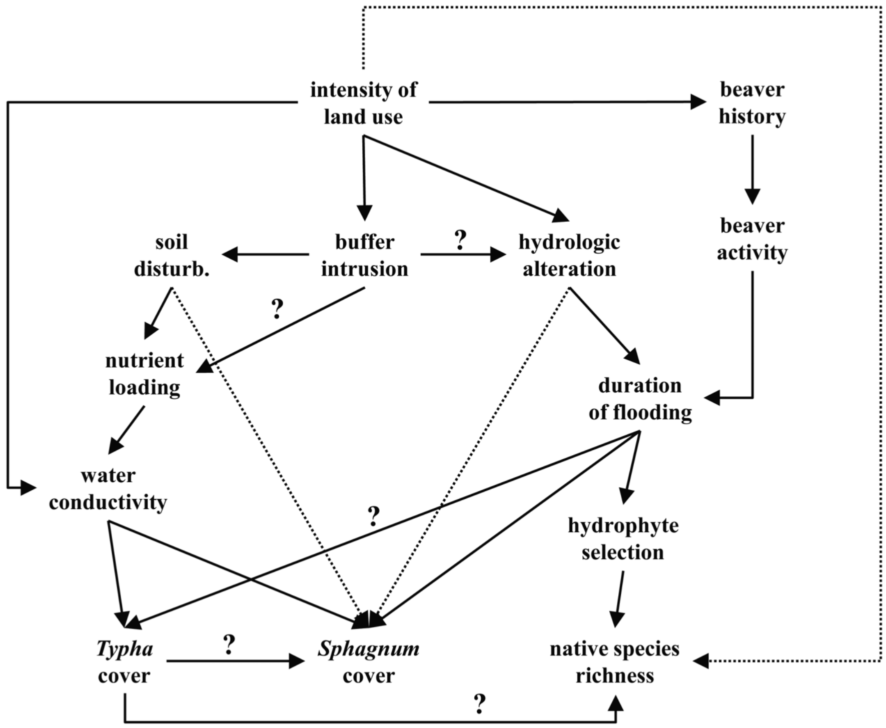

Fig. 8. Revised causal diagram (compared to the initial causal diagram in Fig. 5). Links represented using dotted lines are additions to the diagram based on discovered relationships. Question marks indicate linkages not found to be important in the analysis of the wetland data available.

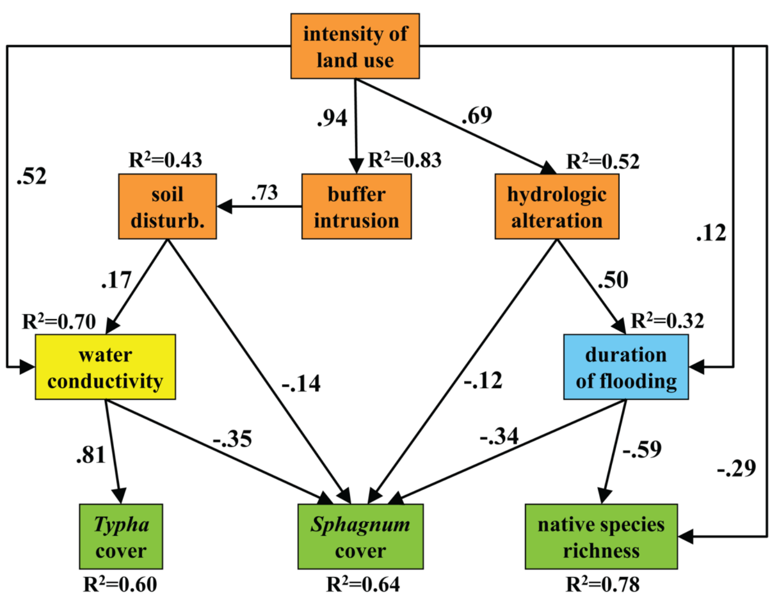

Fig. 9. Final structural equation model. Path coefficients are standardized using an adaptation of the range standardization method of Grace and Bollen (2005) and estimated using queries (see text and Supplement for further description).

Table 3. Parameter estimates for final SE model (Fig. 8). Parameters correspond to those given in Box 3.

Parameter	Mean	SD	2.5%	Median	97.5%
b2.0	−3.102	0.5396	−4.267	−3.066	−2.152
b2.1	2.44	0.4101	1.705	2.417	3.308
b3.0	−1.598	0.2167	−2.037	−1.593	−1.189
b3.1	1.139	0.1658	0.8235	1.136	1.473
b4.0	−5.003	1.731	−9.076	−4.743	−2.378
b4.1	2.006	0.7059	0.8682	1.922	3.619
b5.0	−1.352	0.02952	−1.41	−1.352	−1.295
b5.1	0.3711	0.01383	0.3441	0.3712	0.3985
b5.2	0.1696	0.02521	0.1206	0.1695	0.2191
b6.0	1.805	0.04075	1.725	1.805	1.886
b6.1	0.222	0.106	0.01447	0.2219	0.4313
b6.2	0.2267	0.0401	0.1469	0.2269	0.3058
b7.0	3.993	0.03813	3.918	3.993	4.067
b7.1	−0.002436	2.89E−4	−0.003006	−0.002435	−0.00187
b7.2	−0.1641	0.03585	−0.2352	−0.1639	−0.0945
b8.0	3.646	0.2783	3.1	3.645	4.194
b8.1	−1.133	0.1488	−1.423	−1.133	−0.8408
b8.2	−0.004445	3.718E−4	−0.005179	−0.004443	−0.0037
b8.3	−0.08306	0.02688	−0.1359	−0.08307	−0.0308
b8.4	−0.5694	0.1229	−0.8116	−0.5693	−0.3302
b9.1	0.04522	0.06729	−0.124	0.05807	0.1453
b9.2	1.354	0.3493	0.8161	1.301	2.201
psi.typha	1.901	0.1897	1.532	1.92	2.236

Note: Ninety-five percent credible intervals do not include zero for all parameters except for the initial slope (parameter b9.1) in the change-point model for *Typha*.

model, accept it as our best model based on current information, but leave open the possibility that further information could lead to modifications. As mentioned earlier, the functional forms specified in the provisional model are given in Box 3. Results obtained from examining the initial model and creating the provisional model are referred to here as discoveries, quantities, or queries. Regarding our discoveries, we interpret the additions of new linkages to our SE model as provisional discoveries of processes previously unanticipated (Fig. 9). All of these new connections can be classified as examples of "partial mediation." Our initial causal diagram hypothesized that water conductivity and measurements of the duration of flooding would be sufficient mediators to explain all of the effects of human activities on the biological responses examined. Clearly, that is not the case, as evidence for three additional mediating mechanisms was found. Controlling for duration of flooding, native richness was still found to be lower in areas with higher intensities of land use. Reference to the causal diagram (Fig. 8) shows that it is possible that either part or all of this effect is through influences on beaver. Further studies will be required to gain more clarity on what mediates the additional responsiveness of richness to land use. The other two additional mechanisms relate to lower *Sphagnum* cover where soil was disturbed and where hydroperiods were shortened. These results imply that *Sphagnum* is sensitive to human impacts through numerous mechanisms. Thus, it appears intervening for the purpose of protecting or restoring *Sphagnum* will be challenging. Furthermore, the mediators of the direct paths from soil disturbance and hydrologic alteration are unknown at present, implying the need for further research. Our discovery of a direct link between land use and native richness likewise adds to our uncertainty about how we would fully intervene to reduce human impacts on species diversity, though we could compute the benefit of shortening the duration of flooding by itself.

Quantitative estimates of parameters as well as the variation in their estimated values provide an additional body of important information obtained from the analysis (Table 3). The parameter estimates in their raw units are the most fundamental quantities produced by the estimation process. These are used extensively in the various queries that can be made. Quantities that represent variation in our estimates, such as standard deviations and other computations from the posterior distributions, give us the building blocks for computing the degree of confidence we have in our estimated means and medians as well as conclusions drawn from various queries. Summaries of fit between predicted and observed scores can also be seen as summary quantities (i.e., the R^2s in Fig. 9).

As stated earlier, queries can be used to summarize our results as well as to consider hypothetical situations. For example, a commonly reported summary of findings is the standardized path coefficient. Scientists often place emphasis in presentations on standardized coefficients because of their intuitive appeal. In the linear Gaussian case, standardized path coefficients can be derived simply from the unstandardized values. However, in models possessing more complex specifications, such as in the example presented here, classical standardized coefficients are not simple to interpret or compare. Therefore, we used a query approach to estimate quantities equivalent to standardized coefficients. To accomplish this, we developed two scenarios for each link in the model, one using the minimum value of a predictor and another using its maximum value. We then computed the changes in response associated with changing a predictor from its minimum value to its maximum. Computed results were then standardized by the maximum ranges of the response and predictor variables (Grace and Bollen 2005). The derived parameters obtained in this fashion (shown in Fig. 9) are predictions of the responsiveness of a variable relative to its maximum if one were to vary a predictor from its minimum to maximum value while holding all other predictors at their mean value. These represent a form of standardized path coefficient for comparing the signal strengths among paths.

A query of interest in this study relates to the question of what might be done to limit the development of *Typha* in the protected wetlands. To facilitate our presentation, we show the modeled relationship between water conductivity and *Typha* cover in Fig. 10. Evaluation of this relationship using a change-point model (one that postulates a threshold at which the slope of the response changes) revealed evidence for a threshold response. The median change-point estimate was a conductivity of 1.9 on a log10 scale, or approximately 80 meq, with a credible interval from 32 to 172 meq. Below the threshold, *Typha* appears to have difficulty colonizing (with three wetlands as clear exceptions) and the estimated slope of relationship is effectively zero. Above the threshold, nearly all the wetlands have been invaded by *Typha* and its abundance rises rapidly with increasing conductivity.

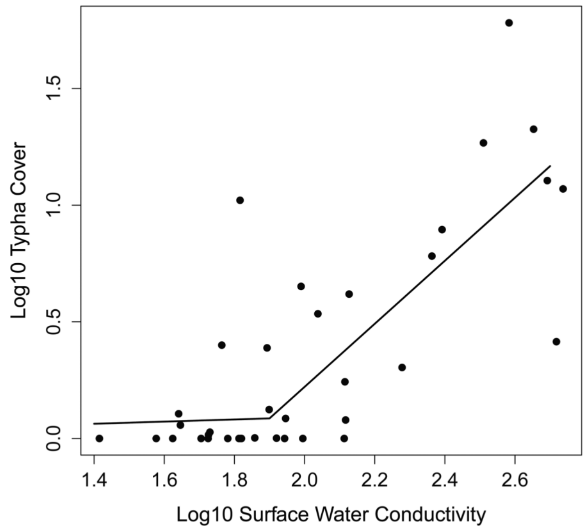

Fig. 10. Illustration of change-point (threshold) relationship between *Typha* cover (raw units between 0 and 100 percent) and surface water conductivity (Log[μS.cm^{-1}]). Change-point relationships represent ones that include thresholds where response slopes change abruptly. Analysis yielded an estimated threshold of 1.91 (95% CI: 1.4–2.2). Initial slope (below threshold) not significantly different from zero.

We can pose the query of what would be predicted to happen to *Typha* if we were to intervene on the sources of variation in conductivity. We illustrate our hypothetical intervention using a reduced-form representation of the predicted net effects of human activities on *Typha* invasion and abundance (Fig. 11). Our model for the observed system assumes that human activities lead to increases in water conductivity (including increases mediated through soil disturbance), which in turn leads to an increase in *Typha*. We present the results for three queries posed as scenarios. Scenario 1 is the no intervention option, which simply predicts for the larger population of wetlands on Mount Desert Island the relationship between water conductivity and *Typha* abundance. Scenario 2 represents the case

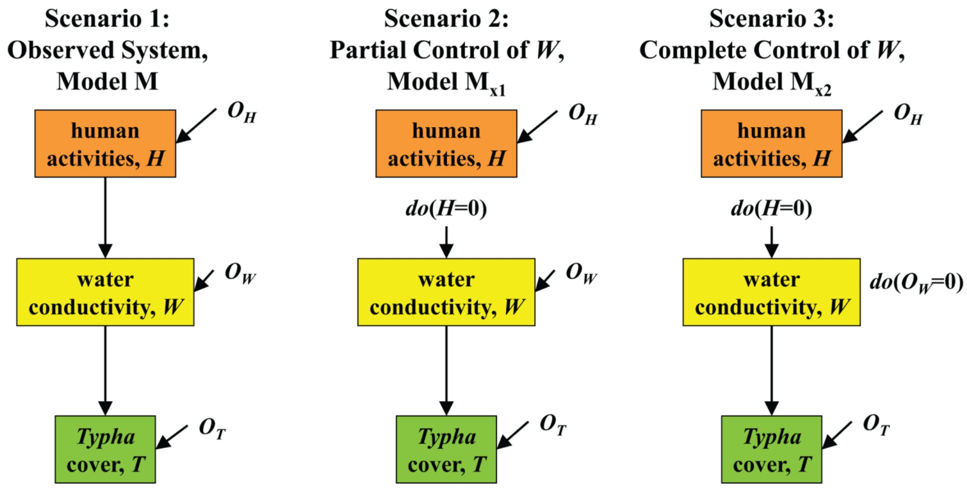

Fig. 11. Queries about predicted effects of interventions on water conductivities and cattail abundance. Scenario 1 is status quo; Scenario 2 is elimination of the human influences of buffer intrusion and soil disturbance; Scenario 3 is reduction of water conductivity to reference conditions (presumably through water treatment). The "O" variables refer to unknown "other" causes of variation. The operator "$do(H=0)$" refers to reducing the effects of land use and soil disturbance on conductivity to 0. The operator "do($O_W=0$)" refers to reducing the effects of other (unknown) factors on conductivity to 0.

where wetlands are protected from high conductivity inputs resulting from human activities. For the sake of simplicity, we assume that there are ways of accomplishing this through all mediating processes, including soil disturbance. If we refer to our three-variable submodel as Model M, a logical implication of Model M is that upon intervention, Models M_{x1} or M_{x2} will apply, depending on the degree of control we can impose on conductivity. Here we recognize the two possibilities of partial (Scenario 2) or complete (Scenario 3) control of conductivity. In Model M_{x1}, we represent an intervention where the values of water conductivity are now independent of levels of human activities, perhaps through the use of systems that trap nutrients in surface water runoff. In Model M_{x2}, we represent an intervention where the values of water conductivity are independent of human activities and also independent of other sources of variation, perhaps through the use of systems that directly regulate water quality. In the case where Model $M_i = M_{x1}$, we can represent the post-intervention distribution of W_i that is predicted from the pre-intervention model M as follows:

$$P_{\mathrm{M}}(W_i|do(H=0)) = P_{\mathrm{Mx1}}(W_i). \quad (3)$$

In the case where Model $M_i = M_{x2}$, we can represent the post-intervention distribution of W_i that is predicted as:

$$P_{\mathrm{M}}(W_I|do(H=0) \text{ and } (U_W=0)) = P_{\mathrm{Mx2}}(W_I). \quad (4)$$

Thus, if our model is an adequate causal model, we should be able to predict the distribution of values for *Typha* for the two interventions. Our goal for the computation will be to determine how much of the distribution of W_i will fall below the distribution for the threshold for an effect on *Typha*. Fig. 12 shows the observed distribution of *Typha* abundances, along with the predicted conductivity distributions for the two scenarios. As we can see, our queries for the two intervention scenarios predict moderate and major reductions in *Typha* abundance. Whether these are accurate predictions is a testable implication of our SE model. Finally, we note some other logical (and testable) implications of the observed system model (M). Most conspicuous is that human activities (including land use, buffer intrusions, and soil disturbance) that do not lead to changes in water conductivity will have no effect on the probability of *Typha* invasions or its abundance, a sometimes overlooked point.

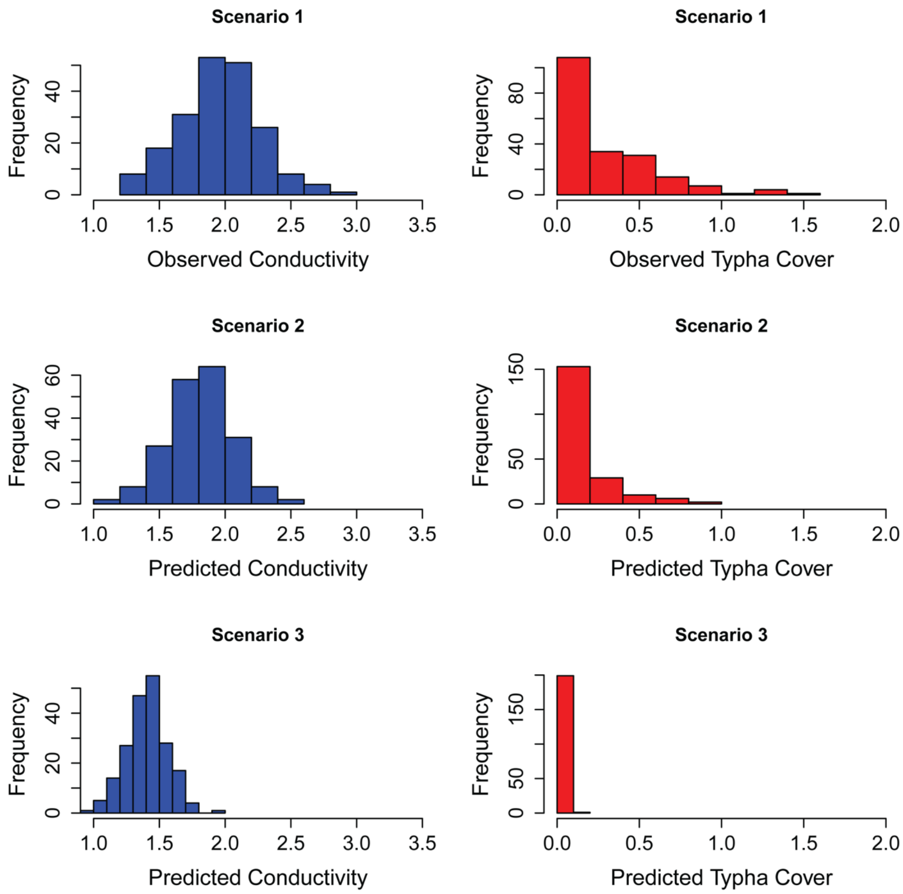

Fig. 12. Distributions associated with scenarios described in Fig. 11. Scenario 1: Observed distributions of values of conductivity, in Log(μS.cm^{-1}) units, and cattail abundance, in Log(percent cover + 1) units. Scenario 2: Predicted distributions for cases where effects of human activities are eliminated (Model M_{x1} in Fig. 11). Scenario 3: Predicted distributions for cases where conductivity is reduced to reference conditions (Model M_{x2}).

Discussion

The wetlands of Mount Desert Island and Acadia National Park

The guidelines proposed in this paper provided a system for learning about and quantifying the different pathways by which human activities on Mount Desert Island may impact the biotic properties of wetlands through abiotic modifications. Starting from an initial set of causal assumptions/hypotheses (Box 2), we first considered whether the data indicated that there were omitted linkages in our initial model. Residual relationships between variables suggested three candidate links for further consideration, one between soil disturbance and *Sphagnum*, one between hydrologic alteration and *Sphagnum*, and one between land use and native richness. For all three of these connections, we hypothesized a directed relationship (rather than an undirected association) involving effects of human activities or abiotic condition on biological conditions. We incorporated the additional predictors into submodels for *Sphagnum* and native richness and recalculated parameters. We judged addition of linkages to our model based on 95% credibility intervals for the terms in the submodels. All three of the detected residual relationships were found to be supported by the data. This led us to a revised causal diagram for the system (Fig. 8). Since the added linkages were not in the initial model, some discussion of their possible interpretations is called for.

The direct path from soil disturbance to

Sphagnum indicates an influence of this physical alteration that is independent from an effect mediated through water conductivity. Further research is needed to determine whether this is simply physical damage to the plant community or something else. The direct path from hydrologic alteration to *Sphagnum* further indicates a mechanism that does not involve the duration of flooding. Perhaps this results from ditches and dams that interrupt sheet flow and represent physical limits to *Sphagnum*, though again the explanation for this effect requires further research. A link between intensity of land use and native richness was also unanticipated. It is quite possible this linkage relates to historical influences not accounted for in the data. We have no good hypotheses to explain this finding at present. Collectively, these discovered linkages argue for the measurement of more potential mediators that could provide causal explanations for the mechanisms involved.

Aside from discovering evidence for additional mechanisms whereby human activities lead to reductions in *Sphagnum* and native richness, several hypothesized mechanisms were not supported by the available data. This result, by itself, does not necessarily mean a different sample of wetlands might not evidence the mechanisms driving the predicted pathways, but we must allow for the possibility that the paths not supported by this dataset may be unsupported generally. In developing a revised causal diagram for the system, we indicate the paths drawn into question (Fig. 8). One process not supported in this case is an effect of buffer intrusion on the degree of hydrologic alteration. As indicated in Box 2, we assumed that when human developments extend into the buffer of a wetland, we might expect additional attempts to stabilize water levels, such as the installation of bulkheads. While this might be true in other locations, it does not appear to be a characteristic feature of the wetlands sampled.

Another assumption/hypothesis not borne out by the data is a promotion of *Typha* by lengthening the hydroperiod. In earlier work on wetlands in the Everglades of Florida, Newman et al. (1996) showed experimentally that both increased flooding and nutrients stimulated *Typha* growth. It may be that further, nonexperimental findings by Newman et al. (1998) provide some clues that could refine our expectations for *Typha* at Mt. Desert Island. Newman et al. (1998) concluded based on comparative surveys that responses by *Typha* to lengthening of hydroperiods depended on sufficient nutrient availability. Thus, for the oligotrophic situation at Mt. Desert Island, we may only observe an effect of increased flooding when conductivities are high.

Typha is generally considered to be of concern in part because of its ability to competitively displace other species when conditions are eutrophic (Keddy 1990, Newman et al. 1996). In this study, we did not find any evidence to suggest a competitive suppression of *Sphagnum* or native species diversity where *Typha* abundance was greater. It may simply be that *Typha* levels were not high enough in this study to detect such suppressing effects. Therefore, we continue to expect that a negative impact of *Typha* on *Sphagnum* and other native species are a possibility in future studies.

Quantification of the relationships in our revised model presents a picture of the network of influences connecting the intensity of land use to biotic conditions (Fig. 9). Variance explanation for response variables (child nodes) was found to be good overall. In particular, the three biological properties were well explained, with R^2s ranging from 60% for *Typha* to 78% for native richness. A high level of variance explanation is not a required outcome for a successful model. Some phenomena are inherently stochastic and valid models may still show low variance explanation. Further, a high level of variance explanation is not a guarantee that we have a good causal model. All that said, since our ambitions for this study include being able to make useful prospective statements, good variance explanation is helpful. We do not wish to imply here that further predictive refinement would not be desirable, only that we are encouraged by these initial findings.

As described in the Methods section, we used queries to obtain standardized estimates of parameters (shown in Fig. 9) from the unstandardized parameters (Table 3). These coefficients represent the predicted responses in child nodes when a parent node is varied across its observed range in the study, where effects are proportional to the responding variables' ranges. Classic standardized coefficients (which are not ideal in the current model because of the complex specifications) represent predicted responses in units of standard deviations. In both cases, standardized coefficients are intended to convey predictive sensitivities in units that are at least somewhat comparable across pathways, though we recognize that standardized coefficients can be misinterpreted. That said, we can get a sense of information flow through the network from the coefficients. In this case, it is clear that the intensity of land use has a variety of kinds of influences on other parts of the system. Looking at the important mediator water conductivity, we can see that human impacts independent of those through buffer intrusion and soil disturbance are very important. We suspect this effect, which is represented by the direct path from land use to conductivity, has to do with mineral and nutrient inputs carried by channelized water flows (e.g., through drainages and culverts). We cannot rule out the importance of occasional extreme rainfall events, as extreme events are often the conveyors of major change. The other key mediator, duration of flooding, appears to be primarily influenced through recognized alterations in hydrology. The survey instrument used to score degree of hydrologic alteration specifically considers ditches, tiles, dikes, weirs, stormwater inputs, point sources, road beds, railroad tracks, and dredging activities. It appears that there is some other connection between land use and duration of flooding (indicated by the direct path), but this is less substantial. We can see that having a measure of the duration of flooding (or a surrogate, such as the average hydric affinity of the vegetation) is important for our modeling effort because duration of flooding is the one node not well predicted; therefore, absorbing this node (i.e., modeling without this variable) would lead to a substantial loss of information. Finally, it is apparent that *Sphagnum* responds to a variety of processes and no one of these is of dominant importance. In contrast, *Typha* abundance is explained reasonably well by a single predictor, conductivity.

Because of its importance to conservation management, we explored in greater depth the chain of effects controlling *Typha* (Figs. 10–12). Further work is needed to confirm our tentative conclusion that there is a threshold requirement for *Typha* response to conductivity. We estimated a change-point log10 conductivity of 1.91 (95% CI = 1.51–2.45). One wetland showed a substantially higher level of *Typha* than expected from its water conductivity level. Possible reasons for this are still under investigation, though beaver activity is suspected to play a role. Scenarios based on two kinds of interventions were explored using queries based on the estimated coefficients (Fig. 11). Further work will be needed to evaluate the practical remedies that are possible under field conditions, though at a minimum controlling soil disturbance adjacent to wetlands might be considered. Monitoring of water conductivity levels in wetlands and in inflows should give further information on management options. Simulating the potential consequences of intervening on the system suggests a considerable potential for avoiding *Typha* invasions if conductivities (and the nutrient inputs they imply) can be controlled (Fig. 12).

Overall, the results are consistent with the a priori expectations that key biotic conditions can be influenced through eutrophication and alteration of hydroperiods (Fig. 4). While this general expectation is supported, it is also clear that many of the details deserve further study. The model results suggest focused questions for further investigation. In particular, gaining a better understanding of the direct link between land use and conductivity in the model seems quite important. Developing a better understanding of how beaver fit into the system is also a conspicuous need. Aside from those needs for further study, the model can be viewed as a set of predictions that encourage further evaluation and refinement of our model of this important natural resource.

Updated guidelines for SEM

Our methodological objective in this paper has been to present an updated set of guidelines for SEM. Standard descriptions of the SEM process generally recommend the specification of the SE model(s) as the first step in the modeling process (Schumacker and Lomax 2004:57, Kline 2010:92). More advanced discussions of the subject (Bollen 1989: Chapter 3, Kaplan 2009: Chapter 10) provide a greater depth of advice for the numerous things that warrant consideration in order to specify an initial SE model that matches theory and represents hypotheses about the data-generating process. All of these treatments recognize that this is a complex decision process. Very few provide a detailed, explicit set of criteria for SE model specification and it often seems implied that previous SEM studies will have provided an a priori model for the current application. Grace and Bollen (2008) and Grace et al. (2010) have recommended beginning the modeling process with a SEMM. This pre-specification step is designed to strengthen the linkage between theory and models, but also to drive applications to a more explicit and careful consideration of specifications that involve latent variables. Generally, however, initial SE models are developed through an intuitive and nonexplicit process. In this presentation we have been motivated to be more explicit in providing suggestions that might guide the modeling process. A major influence on our presentation has been the ambition of incorporating new ideas relating to causal analysis (Pearl 2009), Bayesian implementations (Lee 2007), and probabilistic predictive networks (Kjaerulff and Madsen 2008) into our guidelines.

As a result of the perceived limitations of existing guidelines and the advances in quantitative modeling that have occurred in the past two decades, we have incorporated new devices, such as causal diagrams, and new principles for the investigation of causal relations into our recommendations. We have emphasized the use of graphical modeling methods for model evaluation, both because of their generality and because of the support they provide for explicating causal assumptions. Collectively, these updated guidelines describe a more general approach to SEM than currently practiced, subsuming the special case of the classic SE model

$$\eta = \Gamma\xi + \mathbf{B}\eta + \zeta,$$

where models are described as a series of latent regressions (regressions between latent variables), within the broader framework of the graphical model (Eq. 1)

$$N = \{G, X, F\},$$

where models are described as networks comprising graphs, variables incorporated in those graphs, and functional relations that link the nodes in the network. Our example involved a somewhat complex specification employing Bayesian MCMC methods. A key motivation in our presentation has been to demonstrate how queries can be used for a variety of purposes but especially how they open the door to prospective investigations that explore the quantitative implications that follow from the parameter estimates.

We chose a complex example in order to support a more detailed discussion of modeling possibilities. This is not to suggest that this level of complexity will be typical for most SEM studies. In fact, we would rate this example as an unusually complex one in terms of the linkage functions, one that cannot be implemented in any of the commercial SEM software packages available at this time. Many applications of SEM may not involve all of the steps we present in the guidelines, but considering them provides an enhanced support system compared to previous modeling guidelines. Similarly, choosing a Bayesian estimation approach for this example permitted us to develop a more complex specification for our model. However, Bayesian specification and estimation of SE models comes at a significant price in terms of the background knowledge required to specify the links, experience needed to do the analysis, and the time it takes to explore possible models and misspecifications.

The use of new devices, such as the causal diagram, deserves special mention, since its utility may not be obvious to those accustomed to statistical modeling. Its utility may be most easily grasped if one considers a case where we have a causal diagram prior to conducting an empirical investigation. If we have a well-developed diagram, it can help us to design the data collection enterprise to allow us to be both efficient and effective in achieving specific causal modeling objectives. The causal diagram also facilitates the treatment of causal analysis, as described above. Recent demonstrations of its potential use and contributions in epidemiology can be found in Greenland et al. (1999). Also deserving of special mention as a new emphasis in our guidelines is the use of queries. While this idea is not new, complex models will require queries in order to summarize their retrospective findings (those that draw inferences about what

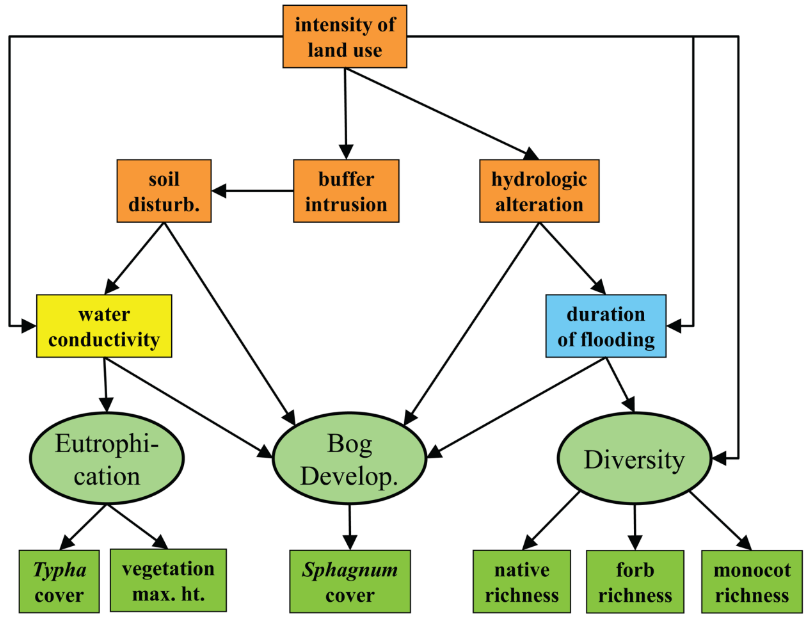

Fig. 13. Alternative structural equation model containing latent variables to represent the hypothesis that generalized system responses (Eutrophication, Bog Development, and Diversity) produce the observed biological metric values. This model is only shown for illustration purposes in this paper.

caused the data characteristics). Using queries automatically opens up the process for prospective analyses that consider potential future observations. Again, to our knowledge, none of the existing commercial software packages provide support for prospective explorations that emerge from the parameter estimates. We feel that SEM applications have generally failed to take full advantage of the opportunity to explore possible implications of the results obtained with prospective analyses. Extending the SEM process to routinely consider implications of results more completely might be one of the most important outcomes from use of the proposed guidelines.

Our presentation in this paper is limited in a number of ways. One of the most conspicuous is the lack of consideration of models that include latent variables, especially since latent variable modeling is a central capability of SEM. Strictly speaking, latent variables are variables for which we have no measurements, but whose influences on other variables need to be accounted for in a model. Modeling with latent variables is a rich and complex topic and we defer the reader to Grace et al. (2010) for a recent and more detailed discussion. It is worth mentioning here, however, that the degree of abstraction varies when it comes to the use of latent variables. For example, we might consider that some of the nodes in our model are subject to measurement error. It is possible to take steps when collecting data to obtain estimates for a node that would permit an estimate of the "true" latent value of a node based on multiple observed values. We might, at that point, develop a version of our model in which each node is represented by a latent variable and one or more observed indicators. Taking latent variable modeling a step further, we might hypothesize that our biotic response measures oversimplify the situation to an unrealistic level. We might consider that *Typha*, *Sphagnum*, and native richness are incomplete representations of the more general responses they represent. From this, we could create a model that incorporates latent nodes to represent these more general responses to human development (e.g., Fig. 13). Such a model could be used to provide a more complete explanation for the multitude of correlated biotic responses to human activities that are observed (Table 2). For the purposes of limiting this already long paper, we defer further discussion of dealing with latent variables in a graph-theoretic implementation of SEM to another time.

Conclusions

The use of statistical tools along with scientific/theoretical knowledge to develop causal inferences is a complex business. Statistical analysis alone is not sufficient for the task and a system for incorporating and explicitly considering causal assumptions and their testable implications is an additional requisite. At the same time, models are only caricatures of the real world and the variety of modeling possibilities is great, demanding both a flexible capacity for modeling and sufficiently developed guidelines that the modeling process is more science than art. It is noteworthy that the aspirations of the field of artificial intelligence are bringing a greater rigor to structural equation modeling. This push is being driven by the realization that artificially intelligent systems require a capacity for causal analysis and also that structural equations are the natural language for that causal analysis (Pearl 2012). Parallel to this fusion of ideas are two trends, one being advances in the capability of statistical modeling and another the desire to develop multivariate models that are better suited to the understanding of systems and the prediction of their behavior. This latter trend is supported by the evolution of graphical modeling methods that permit both the analysis of networks and the prospective application of probabilistic knowledge. The guidelines presented in this paper are intended to integrate ideas from all these domains under a graphical modeling paradigm so as to lead to a more efficient pursuit of causal relationships in systems and to contribute to what we believe will be a third generation of SEM that is both more rigorous and that serves a broader set of purposes.

Acknowledgments

We thank Acadia National Park and especially David Manski for support of data collection. D. Johnson provided assistance with some of the analyses. Conversations with Judea Pearl added to the guidelines presented. Jennifer Duke-Sylvester and Janine Powell provided editorial comments. We thank Randy Mitchell, Jarrett Byrnes, and Andrew McCall for very helpful comments that improved the manuscript significantly. This work was supported, in part, by funding from the USGS Status and Trends, Ecosystems, and Global Change Programs. The use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

Agresti, A. 2010. Analysis of ordinal categorical data. Second edition. John Wiley & Sons, Hoboken, New Jersey, USA.

Anderson, D. R. 2008. Model based inference in the life sciences: a primer on evidence. Springer, New York, New York, USA.

Anderson, T. M., J. Shaw, and H. Olff. 2011. Ecology's cruel dilemma, phylogenetic train evolution and the assembly of Serengeti plant communities. Journal of Ecology 99:797–806.

Ansari, A., K. Jedidi, and S. Jagpal. 2000. A hierarchical Bayesian methodology for treating heterogeneity in structural equation models. Marketing Science 19:328–347.

Arbuckle, J. L. 2011. IBM SPSS Amos 20 user's guide. Amos Development, SPSS. ftp://220.130.147.155/IBM_SPSS_Amos_User_Guide.pdf

Arhonditsis, G. B., C. A. Stow, L. J. Steinberg, M. A. Kenney, R. C. Lathrop, S. J. McBride, and K. H. Reckhow. 2006. Exploring ecological patterns with structural equation modeling and Bayesian analysis. Ecological Modelling 192:385–409.

Barbour, M. T., and C.O. Yoder. 2000. The multimetric approach to bioassessment, as used in the United States of America. Page 281–292 *in* J. F. Wight, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Proceedings of an International Workshop held in Oxford, UK, on 16–18 September 1997. Freshwater Biological Association, Ambleside, Cumbria, UK.

Beguin, J., D. Pothier, and S. D. Cote. 2011. Deer browsing and soil disturbance induce cascading effects on plant communities: a multilevel path analysis. Ecological Applications 21:439–451.

Belovsky, G. E., et al. 2011. The Great Salt Lake Ecosystem (Utah, USA): long term data and a structural equation approach. Ecosphere 2:art33.

Blalock, H. 1964. Causal Inferences in nonexperimental research. University of North Carolina Press, Chapel Hill, North Carolina, USA.

Bollen, K. A. 1989. Structural equations with latent variables. John Wiley & Sons, New York, New York, USA.

Bollen, K. A., J. B. Kirby, P. J. Curran, P. M. Paxton, and F. Chen. 2007. Latent variable models under misspecification: two-stage least squares (TSLS) and maximum likelihood (ML) estimators. Sociological Methods and Research 36:48–86.

Borsboom, D., and G. J. Mellenbergh. 2004. The concept of validity. Psychological Review 111:1061–1071.

Bowker, M. A., F. T. Maestre, and C. Escolar. 2010. Biological crusts as a model system for examining the biodiversity–ecosystem function relationship in soils. Soil Biology & Biochemistry 42:405–417.

Brito, C., and J. Pearl. 2002. A graphical criterion for the identification of causal effects in linear models. Pages 533–538 *in* Proceedings of the Eighteenth National Conference on Artificial Intelligence, Edmonton, Alberta, Canada, July 28–August 1, 2002. AAAI Press/The MIT Press, Menlo Park, California, USA.

Brown, D. G., and A. E. Weis. 1995. Direct and indirect effects of prior grazing of goldenrod upon the performance of a leaf beetle. Ecology 76:426–436.

Byrnes, J. E., D. C. Reed, B. J. Cardinale, K. C. Cavanaugh, S. J. Holbrook, and R. J. Schmitt. 2011. Climate-driven increases in storm frequency simplify kelp forest food webs. Global Change Biology 17:2513–2524.

Carnicer, J., L. Brotons, D. Sol, and M. de Caceres. 2008. Random sampling, abundance–extinction dynamics and niche-filtering immigration constraints explain the generation of species richness gradients. Global Ecology and Biogeography 17:352–362.

Castelle, A. J., A. W. Johnson, and C. Conolly. 1993. Wetland and stream buffer size requirements—a review. Journal of Environmental Quality 23:878–882.

Clark, C., E. Cleland, S. L. Collins, J. E. Fargione, L. Gough, K. L. Gross, S. C. Pennings, K. N. Suding, and J. B. Grace. 2007. Environmental and plant community determinants of species loss following nitrogen enrichment. Ecology Letters 10:596–607.

Dunson, D. B., J. Palomo, and K. A. Bollen. 2005. Bayesian structural equation modeling. Technical Report 2005-5. SAMSI Institute, Durham, North Carolina, USA.

Edelaar, P., D. Serrano, M. Carrete, J. Blas, J. Potti, and J. L. Tella. 2012. Tonic immobility is a measure of boldness toward predators: an application of Bayesian structural equation modeling. Behavioral Ecology doi: 10.1093/beheco/ars006

Forister, M. L., J. A. Fordyce, A. C. McCall, and A. M. Shapiro. 2011. A complete record from colonization to extinction reveals density dependence and the importance of winter conditions for a population of the silvery blue, *Glaucopsyche lygdamus*. Journal of Insect Science 11:art130.

Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith. 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association 85:972–985.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis. Second edition. Chapman & Hall, Boca Raton, Florida, USA.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge, UK.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6:721–741.

Gimenez, O., T. Anker-Nilssen, and V. Grosbois. 2012. Exploring causal pathways in demographic parameter variation: path analysis of mark–recapture data. Methods in Ecology and Evolution 3:427–432.

Gotelli, N. J., and A. M. Ellison. 2006. Food-web models predict species abundances in response to habitat change. PLoS Biology 4(10):e324.

Gough, L., and J. B. Grace. 1999. Predicting effects of environmental change on plant species density: experimental evaluations in a coastal wetland. Ecology 80:882–890.

Gough, L., C. W. Osenberg, K. L. Gross, and S. L. Collins. 2000. Fertilization effects on species density and primary productivity in herbaceous plant communities. Oikos 89:428–439.

Grace, J. B. 1987. The impact of preemption on competitive displacement between two *Typha* species along a water depth gradient. Ecological Monographs 57:283–303.

Grace, J. B. 2006. Structural equation modeling and natural systems. Cambridge University Press, Cambridge, UK.

Grace, J. B. 2008. Structural equation modeling for observational studies. Journal of Wildlife Management 72:14–22.

Grace, J. B., T. M. Anderson, H. Olff, and S. M. Scheiner. 2010. On the specification of structural equation models for ecological systems. Ecological Monographs 80:67–87.

Grace, J. B., and K. A. Bollen. 2005. Interpreting the results from multiple regression and structural equation models. Bulletin of the Ecological Society of America 86:283–295.

Grace, J. B., and K. A. Bollen. 2008. Representing general theoretical concepts in structural equation models: the role of composite variables. Environmental and Ecological Statistics 15:191–213.

Grace, J. B., S. Harrison, and E. I. Damschen. 2011. Local richness along gradients in the Siskiyou herb flora: R. H. Whittaker revisited. Ecology 92:108–120.

Grace, J. B., and J. E. Keeley. 2006. A structural equation model analysis of postfire plant diversity in California shrublands. Ecological Applications 16:503–514.

Grace, J. B., and B. Pugesek. 1997. A structural equation model of plant species richness and its application to a coastal wetland. American Natu-

ralist 149:436–460.
Greenland, S., J. Pearl, and J. M. Robins. 1999. Causal diagrams for epidemiologic research. Epidemiology 10:37–48.
Haavelmo, T. 1943. The statistical implications of a system of simultaneous equations. Econometrica 11:1–12.
Janssen, M. H., P. Arcese, T. K. Kyser, D. F. Bertram, and D. R. Norris. 2011. Stable isotopes reveal strategic allocation of resources during juvenile development in a cryptic and threatened seabird, the Marbled Murrelet (*Brachyramphus marmoratus*). Canadian Journal of Zoology 89:856–868.
Jonsson, M., and D. A. Wardle. 2010. Structural equation modeling reveals plant-community drivers of carbon storage in boreal forest ecosystems. Biology Letters 6:116–119.
Jöreskog, K. G. 1971. Simultaneous factor analysis in several populations. Psychometrika 36:409–426.
Jöreskog, K. G. 1973. A general method for estimating a linear structural equation system. Pages 85–112 *in* A. S. Goldberger and O. D. Duncan, editors. Structural equation models in the social sciences. Seminar Press, NewYork, New York, USA.
Kahl, S., D. Manski, M. Flora, and N. Houtman. 2000. Water resources management plan. National Park Service, Acadia National Park, Mount Desert Island, Maine, USA.
Kaplan, D. 2009. Structural equation modeling. Second edition. Sage, Thousand Oaks, California, USA.
Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6:21–27.
Karr, J. R. 1991. Biological integrity: a long-neglected aspect of water resource management. Ecological Applications 1:66–84.
Keddy, P. A. 1990. Competitive hierarchies and centrifugal organization in plant communities. pp. 266–287 *in* J. B. Grace and D. Tilman, editors. Perspectives on plant competition. Academic Press, New York, New York, USA.
Keeley, J. E., T. Brennan, and A. H. Pfaff. 2008. Fire severity and ecosystem responses following crown fires in California shrublands. Ecological Applications 18:1530–1546.
Kjaerulff, J. B., and A. L. Madsen. 2008. Bayesian networks and influence diagrams. Springer Verlag, New York, New York, USA.
Kline, R. B. 2010. Principles and practice of structural equation modeling. Third edition. Guilford Press, New York, New York, USA.
Koller, D., and N. Friedman. 2009. Probabilistic graphical models. MIT Press, Cambridge, Massachusetts, USA.
Laliberte, E., and J. M. Tylianakis. 2010. Deforestation homogenizes tropical parasitoid-host networks. Ecology 91:1740–1747.
Lamb, E. G., and J. F. Cahill, Jr. 2008. When competition does not matter: grassland diversity and community composition. American Naturalist 171:777–787.
Lau, J. A., A. C. McCall, K. F. Davies, J. K. McKay, and J. W. Wright. 2008. Herbivores and edaphic factors constrain the realized niche of a native plant. Ecology 89:754–762.
Laughlin, D. C. 2011. Nitrification is linked to dominant leaf traits rather than functional diversity. Journal of Ecology 99:1091–1099.
Laughlin, D. C., and J. B. Grace. 2006. A multivariate model of plant species richness in forested systems: old-growth montane forests with a long history of fire. Oikos 114:60–70.
Lee, S. Y. 2007. Structural equation modeling: a Bayesian approach. John Wiley & Sons, New York, New York, USA.
Lee, S. Y., and X. Y. Song. 2004. Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. Multivariate Behavioral Research 39:653–686.
Little, A. M., G. R. Guntenspergen, and T. F. H. Allen. 2010. Conceptual hierarchical modeling to describe wetland plant community organization. Wetlands 30:55–65.
Little, A. M., G. R. Guntenspergen, and T. F. H. Allen. In press. Wetland vegetation dynamics in response to beaver (*Castor canadensis*) activity at multiple scales. Ecoscience.
Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325–337.
Mack, J. J. 2001. Ohio rapid assessment method for wetlands v. 5.0, user's manual and scoring forms. Ohio EPA Technical Report WET/2001-1. Ohio EPA, Division of Surface Water, Wetland Ecology Group, Columbus, Ohio, USA.
Mitchell, R. J. 1992. Evolutionary and ecological hypotheses using path analysis and structural equation modeling. Functional Ecology 6:123–129.
Muggeo, V. M. R. 2003. Estimating regression models with unknown break-points. Statistics in Medicine 22:3055–3071.
Muthén, B. O. 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika 49:115–132.
Muthén, L. K., and B. O. Muthén. 2011. Mplus user's guide. Sixth edition. Muthén & Muthén, Los Angeles, California, USA.
Newman, S., J. B. Grace, and J. W. Koebel. 1996. Nutrient and hydroperiod effects on *Typha, Cladium*, and *Eleocharis*: implications for Everglades restoration. Ecological Applications 6:774–783.
Newman, S., J. Schutte, J. B. Grace, K. Rutchey, T.

Fontaine, M. Campbell, M. Pietrucha, and K. R. Reddy. 1998. Factors influencing cattail abundance in the northern Everglades. Aquatic Botany 60:265–280.

Ntzoufras, I. 2009. Bayesian modeling using WinBUGS. John Wiley & Sons, Hoboken, New Jersey, USA.

Pearl, J. 1988. Probabilistic reasoning in intelligent systems. Morgan Kaufmann, San Francisco, California, USA.

Pearl, J. 1995. Causal diagrams for empirical research. Biometrika 82:669–710.

Pearl, J. 2009. Causality. Second edition. Cambridge University Press, Cambridge, UK.

Pearl, J. 2012. The causal foundations of structural equation modeling. Pages 68–91 *in* R. H. Hoyle, editor. Handbook of structural equation modeling. Guilford Press, New York, New York, USA.

Prugh, L. R., and J. S. Brashares. 2012. Partitioning the effects of an ecosystem engineer: Kangaroo rats control community structure via multiple pathways. Journal of Animal Ecology 81:667–678.

Pugesek, B. H., and A. Tomer. 1996. The Bumpus house sparrow data: a reanalysis using structural equation models. Evolutionary Ecology 10:387–404.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reich, P. B., L. E. Frelich, R. A. Voldseth, P. Bakken, and E. C. Adair. 2012. Understorey diversity in southern boreal forests is regulated by productivity and its indirect impacts on resource availability and heterogeneity. Journal of Ecology 100:539–545.

Riginos, C., and J. B. Grace. 2008. Savanna tree density, herbivores, and the herbaceous community: bottom-up versus top-down effects. Ecology 89:2228–2238.

Riseng, C. M., M. J. Wiley, P. W. Seelbach, and R. J. Stevenson. 2010. An ecological assessment of Great Lakes tributaries in the Michigan Peninsulas. Journal of Great Lakes Research 36:505–519.

Rupp, A. A., D. K. Dey, and B. D. Zumbo. 2004. To Bayes or not to Bayes, from whether to when: applications of Bayesian methodology to modeling. Structural Equation Modeling: A Multidisciplinary Journal 11:424–451.

Satorra, A., and P. M. Bentler. 1994. Corrections to test statistics and standard errors in covariance structure analysis. Pages 399–419 *in* A. von Eye, and C. C. Clogg, editors. Latent variables analysis: applications for developmental research. Sage, Thousand Oaks, California, USA.

Satorra, A., and P. M. Bentler. 2001. A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66:507–514.

Scheiner, S. M., R. J. Mitchell, and H. S. Callahan. 2000. Using path analysis to measure natural selection. Journal of Evolutionary Biology 13:423–433.

Schoolmaster, D. R., Jr., J. B. Grace, and E. W. Schweiger. 2012. A general theory of multimetric indices and their properties. Methods in Ecology and Evolution 3:773–781.

Schumacker, R. E., and R. G. Lomax. 2004. A beginner's guide to structural equation modeling. Second edition. Lawrence Erlbaum Associates, Mahwah, New Jersey, USA.

Seabloom, E. W., J. W. Williams, D. Slayack, D. M. Stoms, J. H. Viers, and A. P. Dobson. 2006. Human impacts, plant invasion, and imperiled plant species in California. Ecological Applications 16:1338–1350.

Shipley, B. 1995. Structured interspecific determinants of specific leaf area in 34 species of herbaceous angiosperms. Functional Ecology 9:312–319.

Shipley, B. 2000*a*. Cause and correlation in biology. Cambridge University Press, Cambridge, UK.

Shipley, B. 2000*b*. A new inferential test for path models based on directed acyclic graphs. Structural Equation Modeling 7:206–218.

Shipley, B. 2003. Testing recursive path models with correlated errors using d-separation. Structural Equation Modeling 10:214–221.

Shipley, B. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90:363–368.

Spirtes, P., C. N. Glymour, and R. Scheines. 2000. Causation, prediction, and search. Second edition. MIT Press, Cambridge, Massachusetts, USA.

Stock, J. H., and F. Trebbi. 2003. Retrospectives: Who invented instrumental variable regression? Journal of Economic Perspectives 17:177–194.

Sturtz, S., U. Ligges, and A. Gelman. 2005. R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical Software 12:1–16.

Tierney, G. L., D. Faber-Langendoen, and B. R. Mitchell. W. G. Shriver, J. P. Gibbs. 2009. Monitoring and evaluating the ecological integrity of forest ecosystems. Frontiers in Ecology and the Environment. 7:308–316.

Tonsor, S. J., and S. M. Scheiner. 2007. Plastic trait integration across a CO_2 gradient in *Arabidopsis thaliana*. American Naturalist 169:E119–E140.

Vile, D., B. Shipley, and E. Garnier. 2006. A structural equation model to integrate changes in functional strategies during old-field succession. Ecology 87:504–517.

Weiher, E. 2003. Species richness along multiple gradients: testing a general multivariate model in oak savannas. Oikos 101:311–316.

Wetzel, R. G. 2001. Limnology. Third edition. Academic Press, San Diego, California, USA.

Wootton, J. T. 1992. Indirect effects, prey susceptibility, and habitat selection: impacts of birds on limpets and algae. Ecology 73:981–991.

Wright, S. 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. Proceedings of the National Academy of Sciences 6:320–332.

Wright, S. 1921. Correlation and causation. Journal of Agricultural Research 10:557–585.

Youngblood, A., J. B. Grace, and J. D. McIver. 2009. Delayed conifer mortality after fuel reduction treatments: interactive effects of fuel, fire intensity, and bark beetles. Ecological Applications 19:321–337.

Zuur, A. F., E. N. Leno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer Verlag, New York, New York, USA.

Supplemental Material

Supplement

R scripts for analyses summarized in the main text (*Ecological Archives* C003-009-S1).