Comprehensive evaluation of model uncertainty in qualitative network analyses

J. Melbourne-Thomas,[1,4] S. Wotherspoon,[2,3] B. Raymond,[1,3] and A. Constable[1,3]

[1]*Antarctic Climate and Ecosystems Cooperative Research Centre, University of Tasmania, Private Bag 80, Hobart, TAS 7001 Australia*
[2]*Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 129, Hobart, TAS 7001 Australia*
[3]*Australian Antarctic Division, Department of Sustainability, Environment, Water, Population and Communities, 203 Channel Highway, Kingston, TAS 7050 Australia*

Abstract. Qualitative network analyses provide a broad range of advantages for formulating ideas and testing understanding of ecosystem function, for exploring feedback dynamics, and for making qualitative predictions in cases where data are limited. They have been applied to a wide range of ecological questions, including exploration of the implications of uncertainty about fundamental system structure. However, we argue that questions regarding model uncertainty in qualitative network analyses have been under-explored, and that there is a need for a coherent framework for evaluating uncertainty. To address this issue, we have developed a Bayesian framework for interpreting uncertainty that can be applied when comparing and evaluating the characteristics and behavior of alternative model formulations. Specifically, we recognize that results from previously developed simulation approaches to qualitative modeling can be interpreted as marginal likelihoods that translate to Bayes factors for model comparison. We then test and extend our Bayesian interpretation of qualitative model results to address comparisons both between and within alternative models. With the use of examples, we demonstrate how our Bayesian framework for interpretation can improve the application of qualitative modeling for addressing uncertainty about the structure and function of ecological networks.

Key words: Bayes factor; Bayesian framework; network analysis; qualitative model; uncertainty.

INTRODUCTION

Models are valuable tools for exploring potential food web and ecosystem responses to human impacts and environmental change, but the development of models for complex ecological systems is constrained by data availability, mechanistic understanding, and manpower. Qualitative network modeling provides a method for exploring the structure and function of ecological systems without the need for a fine level of detail in functional representations or parameterization. Particular advantages of the qualitative approach to network modeling include the ability to rapidly formulate ideas about system function, the ability to capture feedback effects in qualitative predictions, and as a precursor to inform the development of more detailed quantitative models.

The qualitative approach was originally developed by Levins (1974) and Puccia and Levins (1985) to formalize the analysis of feedbacks in network models. Since then, qualitative network modeling has been broadly applied in ecology, for example to develop hypotheses for species-level interactions (Dambacher et al. 1999), to explore life expectancy changes in perturbed communi-

ties (Dambacher et al. 2005), to determine indicators for exploited systems (Dambacher et al. 2009, Metcalf et al. 2011), to evaluate eradication scenarios for invasive species (Raymond et al. 2011), and to explore alternative stable states in ecological systems (Marzloff et al. 2011). However, we are aware of only three cases where uncertainty about the structure of qualitative network models has been treated explicitly (Montaño-Moctezuma et al. 2007, Hosack et al. 2008, Raymond et al. 2011). We argue that uncertainty in qualitative network models could be explored better and more fully in order to improve consistency in interpretations, specifically, within a Bayesian framework. As a result, qualitative network modeling will have far greater utility in addressing ecological questions.

Hosack et al. (2008) introduced an approach to exploring model structure uncertainty within the probabilistic framework of a Bayesian Belief Network (BBN). While their BBN approach holds clear advantages in terms of populating conditional probability tables, which otherwise often relies on expert opinion that may not incorporate feedback properties of ecological networks, it does not fully capture the dependencies between variables in underlying models. An alternative approach for capturing uncertainty in the structure of qualitative network models is presented by Raymond et al. (2011). This simulation method enables

uncertainty about linkages to be captured within a single model, and also allows incorporation of prior knowledge to inform model selection. However, as Raymond et al. (2011) recognize, there is greater potential for exploring the outcomes from a simulation approach to qualitative network modeling than is explored in their paper (or elsewhere).

Here we combine and extend the two approaches (i.e., Hosack et al. 2008 and Raymond et al. 2011) by: (1) recognizing that simulation results have a direct Bayesian interpretation (without needing to invoke fitted relationships from "representative" models or to embed results in belief networks), which provides useful, high-level summaries of the relationship between model structure and behavior; and (2) using multivariate approaches to summarize and explore the resulting posterior distributions that describe the behavior of the model system.

In doing so, we provide an integrated framework that goes beyond previous approaches to addressing model structure uncertainty and that facilitates model building and testing in an exploratory context. We demonstrate our framework using models for two contrasting ecological systems examined by Hosack et al. (2008) and Raymond et al. (2011).

BACKGROUND: REPRESENTING UNCERTAINTY IN NETWORK MODELS

The representation of network models as signed directed graphs (or signed digraphs), their translation to community matrix form, and the use of the community matrix to examine responses to perturbations has been described in detail elsewhere (e.g., see Dambacher et al. 2002, 2003b, Berlow et al. 2004, Hosack et al. 2008). Here, we focus on methodological approaches to representing uncertainty in qualitative network models. However, we note that the existing literature contains some minor ambiguities in descriptions of the general approach to qualitative network analysis (in particular, contrasting representations of net growth vs. per capita functions). Formal descriptions of model representation, stability assessment, and interpretation of responses to press perturbations are therefore revisited in Appendix A.

The qualitative modeling approach is based on the properties of the community matrix. Assume that the population dynamics of a system of n interacting species can be represented as a system of differential equations of the following form:

$$\frac{dN_i}{dt} = f_i(N_1, \ldots, N_n)N_i \qquad (1)$$

where N_i is the density of population i, and f_i describes the per capita growth rate of that population accounting for processes of birth, death, and migration. The behavior of the system near an equilibrium is characterized by the community matrix (sensu Berlow et al. 2004) $\mathbf{A} = [a_{ij}]$ of interaction coefficients, defined as the partial derivatives

$a_{ij} = \partial f_i / \partial N_j$ evaluated at the equilibrium. These interaction coefficients describe the pairwise interactions between the modeled populations, in that a_{ij} represents the direct effect of population j on the per capita growth rate of population i. If in addition the population densities N_i are scaled so that for each i, equilibrium occurs at $N_i = 1$, then the stability of the equilibrium is characterized by the eigenvalues of the community matrix; if all eigenvalues of $\mathbf{A}$ have negative real part, the equilibrium is stable. Most importantly, the response of the system to a press perturbation, defined as a sustained shift in the per capita growth rate of a population (Bender et al. 1984), can be determined from the inverse community matrix $\mathbf{A}^{-1}$ (Puccia and Levins 1985).

For many ecological systems, both (1) the nature (i.e., the signs [positive, negative, or zero] of the a_{ij}) and (2) the strength (i.e., the magnitudes of the $|a_{ij}|$) of the interactions between populations may be uncertain. Issue (1) relates to model structure uncertainty, while issue (2) can be thought of as a parameterization problem. Qualitative modeling focuses on the former by specifying only the signs of the interaction coefficients of the community matrix (Levins 1974), and seeking only qualitative predictions of the system's response to a perturbation. That is, given only the nature of the interactions between populations, qualitative modeling aims to predict which populations are advantaged and which are disadvantaged when other populations in the system are advantaged or disadvantaged.

As only the signs of the interaction coefficients are retained, the qualitative models can be represented by signed directed graphs (or signed digraphs). The nodes of a signed digraph represent the modeled populations, and the linkages (or directed edges) represent the nonzero interaction coefficients, annotated to indicate the sign of the interaction (e.g., see Dambacher et al. 2002). With this scheme, the sign structure of the community matrix is determined by the signed adjacency matrix of the directed graph (see Fig. 1).

Of course, qualitative predictions can be ambiguous because of the propagation of both positive and negative feedback effects (Dambacher et al. 2002). The need to address ambiguity in predictions has led to the development of "prediction weights" (derived from ratios of feedback cycles; see Dambacher et al. 2002, 2003b) and to the use of simulation, where linkages are assigned random interaction strengths, and outcomes are either used to verify prediction weightings (Dambacher et al. 2003a, Hosack et al. 2008), or responses are aggregated from multiple random assignments of interaction strengths (as in Raymond et al. 2011).

Two broad alternatives have typically been used to incorporate uncertainty about the structure of a given system in qualitative modeling:

1) The use of a representative set of models that captures alternative versions of our understanding of the interactions within the system (e.g., Dam-

bacher et al. 2002, Hosack et al. 2008). The set of alternative models is generally small (less than five models), and each model within the set is structurally unambiguous.

2) The use of a single overarching model structure that includes one or more linkages that are marked as "uncertain." This single structure can be expanded into a set of unambiguous models by including or excluding these uncertain linkages in appropriate combinations (e.g., Montaño-Moctezuma et al. 2007, Raymond et al. 2011). This differs from (1) in that the set of models can be large (hundreds or thousands of models) and can no longer be considered as a representative set, but rather as a set of candidate models that encompass all possible model structures that are consistent with the assumed understanding of the system.

These two approaches lend themselves to different applications and methods of analysis. With a small set of alternative models, one is typically interested in exploring differences in the behavior of alternative models or in identifying the most likely models from those in the set. This latter goal is essentially addressing structural uncertainty, i.e., uncertainty about which model structure best represents the true system. Hosack et al. (2008) provide one method for evaluating the likelihoods of alternative models given a set of observations, while Dambacher et al. (2002) focus on interpreting differences in responses to press perturbations for alternative models in terms of feedback cycles and stability properties.

An alternative approach to the analysis of such a set of models is through simulation, where alternative realizations of the community matrix are generated through assignment of random weights (e.g., see Dambacher et al. 2003a, Novak et al. 2011, Raymond et al. 2011). The validity of these alternative realizations can be assessed in terms of numerical stability, and according to preestablished ecological criteria (Raymond et al. 2011). Only models that are stable and that respond in a prespecified way to a previously observed perturbation scenario are used to examine responses to an unobserved press perturbation. Results are combined across alternative models to give proportional (qualitative) responses for each model variable. This approach is suited to small or large model sets (i.e., model sets generated from either alternative [1] or [2]) and provides an opportunity to investigate structural uncertainty (identifying models from within the set that best represent the true system) as well as parameter uncertainty (exploring the effects of various parameter signs and strengths on model behavior).

Approach

Simulation procedure

Given either a small number of representative models, or a single overarching model with uncertain linkages, a single community matrix is simulated as follows:
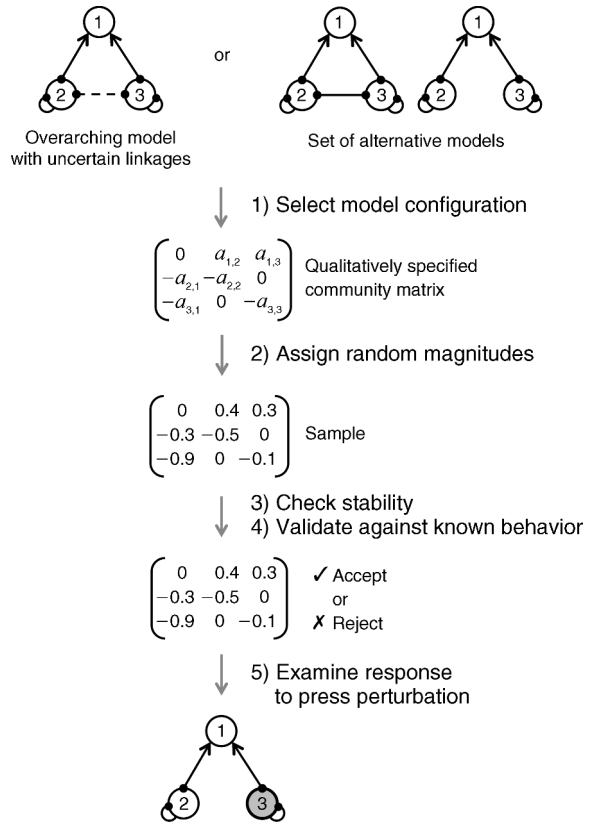

Fig. 1. Simulation approach to qualitative modeling. Network models are represented as signed digraphs, composed of model variables (or nodes) and linkages, where arrows and filled circles represent positive and negative effects, respectively, from a source node to a receiving node. The system may be represented by a single model with uncertain linkages (indicated as dashed lines), or by a set of alternative models. The simulation algorithm comprises the following steps: (1) A particular model configuration is selected and translated into a qualitatively specified community matrix. (2) The community matrix is sampled by assigning random weights to each nonzero element. (3) The stability of the sample is checked (based on eigenvalues), as is its (4) ability to reproduce any known behaviors for the system under study. (5) Accepted samples are used to examine the response of the system to perturbations of interest (here, the perturbed node is shaded gray). This procedure is repeated until a prespecified number of acceptable samples have been generated, and results are aggregated across samples.

1) First, an unambiguous model configuration is randomly selected. As appropriate, either one of the representative models is selected at random, or a random subset of uncertain linkages are deleted from the overarching model, to produce a single model configuration that specifies the sign structure of the community matrix.

2) The community matrix is then generated by assigning random magnitudes $|a_{ij}|$ from a prespecified distribution F to the nonzero interaction coefficients.

3) The eigenvalues of the generated matrix are then computed to assess the stability of the system. If the generated community matrix corresponds to an unstable equilibrium, the matrix is discarded and procedure commences again at step 1.

4) The generated matrix is validated against any known behaviors for the study system (i.e., qualitative responses for any model components under a previously observed perturbation scenario). If the predictions from the generated matrix are not consistent with known behaviors, the matrix is discarded and procedure commences again at step 1.

5) The generated matrix is used to predict the response of the system to perturbations of interest.

This procedure (which is represented schematically in Fig. 1) is repeated until a prespecified number of acceptable samples have been generated, and the predictions from the final step together with simulated magnitudes a_{ij} and model indicators M that uniquely identify the selected model configuration are all tabulated. This process is equivalent to the methods of Dambacher et al. (2003a), Novak et al. (2011), and Raymond et al. (2011), except for the additional model selection step 1.

An alternative means of conceptualizing our simulation procedure is as "searches" over different spaces, i.e., the space which defines alternative model configurations, and the parameter space of $|a_{ij}|$ magnitudes. The former search is exhaustive (all model configurations are examined), while the second is determined by the sampling distribution F. Only dynamically stable combinations of configuration and parameterization (that meet some prespecified validation criterion if available) are used to predict system responses to press perturbations.

Bayesian framework for interpretation

To take a Bayesian interpretation of the simulation procedure presented in the previous section, we view the attribution of signs to the interaction coefficients a_{ij} as the model, and the magnitudes $|a_{ij}|$ of the interaction coefficients as model parameters θ. The total parameter set consists of these model parameters together with the model indicators M that record which model configuration has been selected. The prior model probabilities $p(M)$ are determined by the probabilities with which models or uncertain linkages are selected in step 1, while the prior distribution $p(\theta)$ of the model parameters is the sampling distribution F in step 2. The observed data y consists of both the observation that the equilibrium is stable (step 3) together with any data used to validate the model (step 4). In keeping with the spirit of qualitative analysis, the likelihood $p(y|\theta, M)$ is simply 1 if the matrix corresponds a stable equilibrium and is consistent with the validation data, and zero otherwise. Finally, any results $\hat{y}$ calculated from the sampled matrix (step 5) can be viewed as posterior model predictions.

With these definitions, the simulation procedure of the previous section can be viewed as sampling from the joint posterior $p(\theta, M | y)$ of the model parameters and model indicators by importance sampling using the prior and the sampling density (Geweke 1989, Kass and Raftery 1995). We note that this Bayesian framework deals analytically with both of the search spaces described in *Approach: Simulation procedure*, i.e., the parameter space of $|a_{ij}|$ magnitudes (θ) and the alternative model configurations defined by the model indicators M. It bears some similarity to the uncertainty framework introduced by Spear and Hornberger (1980), although the focus of this earlier framework is on parameter sensitivity rather than model structure uncertainty.

The model parameters θ and predictions $\hat{y}$ generated in this way are samples from the model averaged posterior $p(\theta | y)$ and posterior predictive distribution $p(\hat{y} | y)$, respectively. To obtain results for a particular model M, that is, to obtain samples from the model conditional posterior $p(\theta | M, y)$ and posterior predictive distribution $p(\hat{y} | M, y)$, it suffices to consider only those samples generated from the required model M.

The posterior model probabilities $p(M | y)$ are estimated by the proportion of samples in which the particular model/uncertain linkage appears, allowing the candidate models to be directly compared. The Bayes factor for the comparison of two alternative models M_1 and M_2 (Kass and Raftery 1995) can then be calculated as the odds ratio:

$$B_{12} = \frac{p(M_1 | y)p(M_2)}{p(M_2 | y)p(M_1)}. \qquad (2)$$

Alternately, if each model is simulated separately, the marginal likelihood $p(y | M_i)$ for that model is estimated by the proportion of matrices not rejected in steps 3 and 4, and the Bayes factor can be calculated as a ratio of marginal likelihoods:

$$B_{12} = \frac{p(y | M_1)}{p(y | M_2)}. \qquad (3)$$

Example models

In the interests of consistency and comparability, we use models that have previously been introduced and analyzed by Hosack et al. (2008) and Raymond et al. (2011) to test our Bayesian framework (*Approach: Testing ...*), and to apply it in comparisons of alternative model formulations (*Approach: Applying ...*). We also use these models to demonstrate a set of complementary techniques for exploring prediction space generated through simulation (*Approach: Exploring ...*). The three model sets analyzed here are: (1) a set of five-variable models used by Hosack et al. (2008) that represent a range of real-world systems (note that these models were also analyzed by Dambacher et al. 2003a); (2) alternative

models for the lake mesocosm system used by Hosack et al. (2008, originally described in Hulot et al. 2000); and (3) the subantarctic Macquarie Island model developed by Raymond et al. (2011). These latter two systems provide a useful contrast in terms of complexity and model formulation. The lake mesocosm models (Fig. 2) are for a simple, well-studied system, with seven- and eight-variable versions capturing differences in responses with and without the presence of carnivorous fish. Both Hosack et al. (2008) and Hulot et al. (2000) indicate that alternative model structures to those explored in these two papers are conceivable, and may better capture the properties of the mesocosm system. We explore structural uncertainty for seven- and eight-variable models by adding and deleting key linkages, and comparing prediction capabilities of alternative formulations using Bayes factors.

By contrast, the Macquarie Island model (Fig. 3) is for a complex system with multiple unknown interactions (which are captured within a single representation). The pest eradication scenarios examined by Raymond et al. (2011), i.e., removal of introduced rabbits, rats, and mice, have recently been carried out in the real system; hence data are likely to become available in the near future to confirm or reject model predictions. We note that the version of the Macquarie Island model used here does not include variables for cats and myxoma because this better reflects the recent state of the system (cats were eradicated from the island in 2000 and myxoma releases ceased in 2006; Copson 2002, Dowding et al. 2009). The validation scenario used in analysis of this model (step 4 of the simulation procedure outlined in *Approach: Simulation procedure*) is a decrease in tall tussock vegetation and an increase in rabbits under a press perturbation representing an increase for rabbits; these changes were observed following the release of rabbits from cat predation and myxoma mortality (Scott and Kirkpatrick 2008, Bergstrom et al. 2009, Dowding et al. 2009). While Raymond et al. (2011) modeled this predation and disease release explicitly, our modifications do not have a significant effect on model predictions. All analyses of these models, and all algorithms described in the following sections, are implemented in R (version 2.13.0; R Development Core Team 2011; example code is provided in the Supplement).

Testing the Bayesian framework

Bayesian model selection techniques can perform poorly in the presence of non-informative priors (the so-called Bartlett-Lindley paradox; see Kass and Raftery 1995). Yet a primary aim of qualitative modeling is to compare models when little is known about the model parameters—the interaction coefficients. We therefore used simulations to test the capacity of the Bayesian approach to distinguish models with only weakly informative priors for the interaction coefficients. In each iteration of these simulations, a model was selected
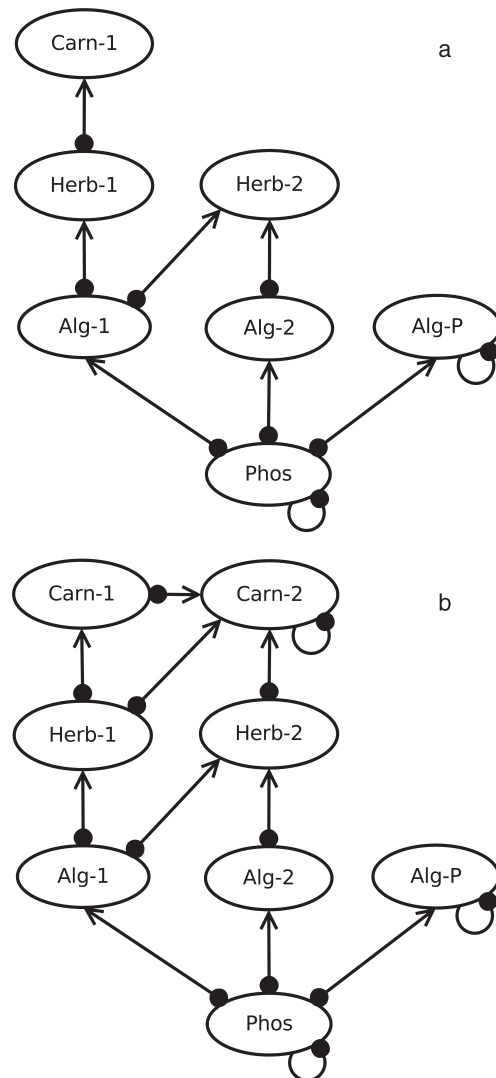

FIG. 2. (a) Seven- and (b) eight-variable lake mesocosm models analyzed by Hosack et al. (2008; originally presented by Hulot et al. 2000). Abbreviations are: Alg-1, edible algae; Alg-2, protected algae; Alg-P, periphyton; Carn-1, invertebrate carnivores; Carn-2, fish; Herb-1, small herbivores; Herb-2, large herbivores; Phos, phosphorus.

at random and nominated to be "true." Data were generated from this true model by perturbing a randomly selected node and monitoring two other randomly selected nodes, and the posterior probabilities $p(M \mid y)$ of model correctness were calculated for each model. Outputs were examined in terms of comparative probabilities for model correctness across 100 iterations, as well as the frequency with which the true model was assigned the highest probability of correctness. A schematic for the algorithm underlying this process is presented in Fig. 4.

The process was applied to two sets of models. The first was a subset of the five-variable models used by
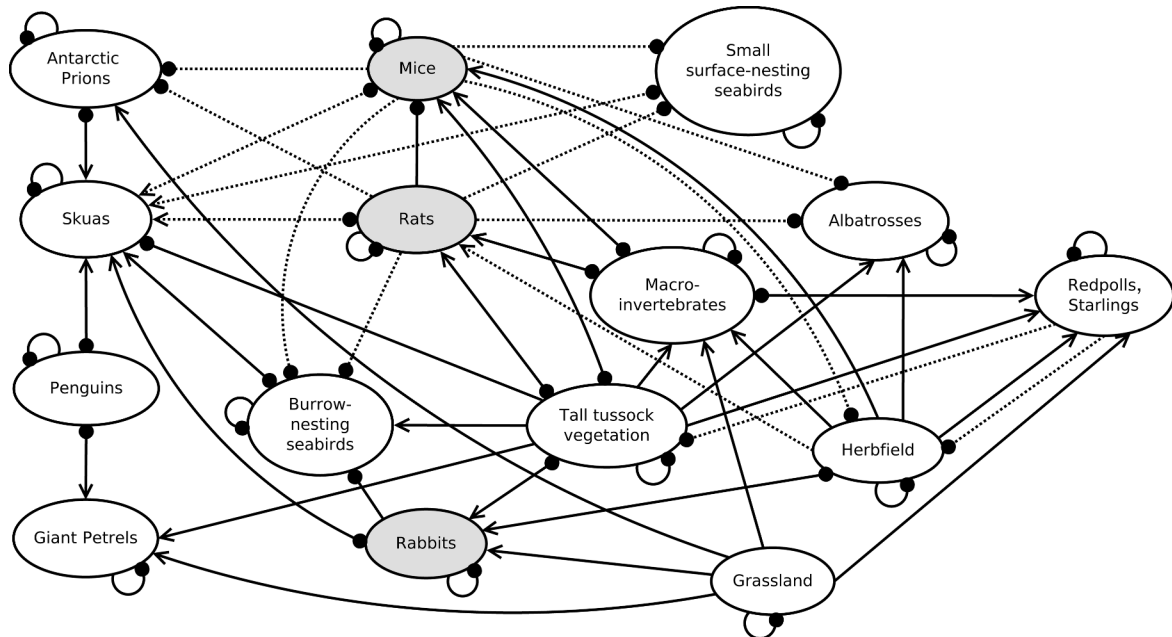
Fig. 3. Subantarctic Macquarie Island model (modified from Raymond et al. [2011]). Solid lines represent certain linkages, while dashed lines are uncertain linkages. Pest species are shaded gray.

Hosack et al. (2008) for deriving fitted relationships to calculate conditional probability values. We chose five of the 10 models analyzed by Hosack et al. (2008), selecting models that represent a range of structures observed in real systems (Fig. 5a). To test our ability to distinguish between a set of more closely related models (i.e., alternative representations of a particular system), we also performed simulations using three versions of
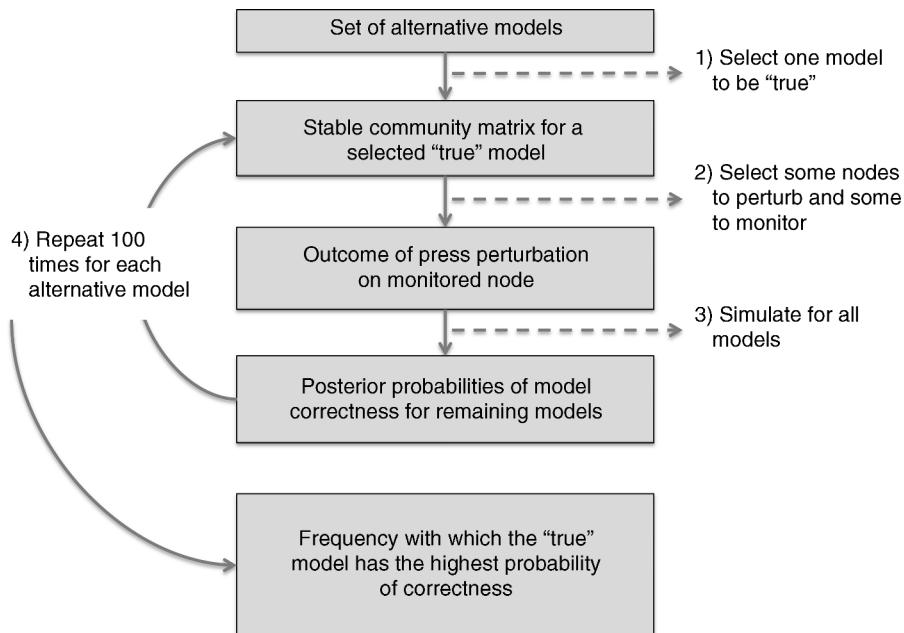

Fig. 4. A Bayesian approach to comparison of qualitative network models. (1) One model is selected to be "true" from a set of alternative models, and is translated to a stable community matrix form. (2) A subset of model nodes are perturbed, and outcomes of a press perturbation are determined for some subset of monitored nodes. (3) Simulations are performed for all models to compute posterior probabilities of model correctness. (4) This process is repeated 100 times for each of the possible models, and the frequency with which the correct model is given the highest probability of correctness is recorded.
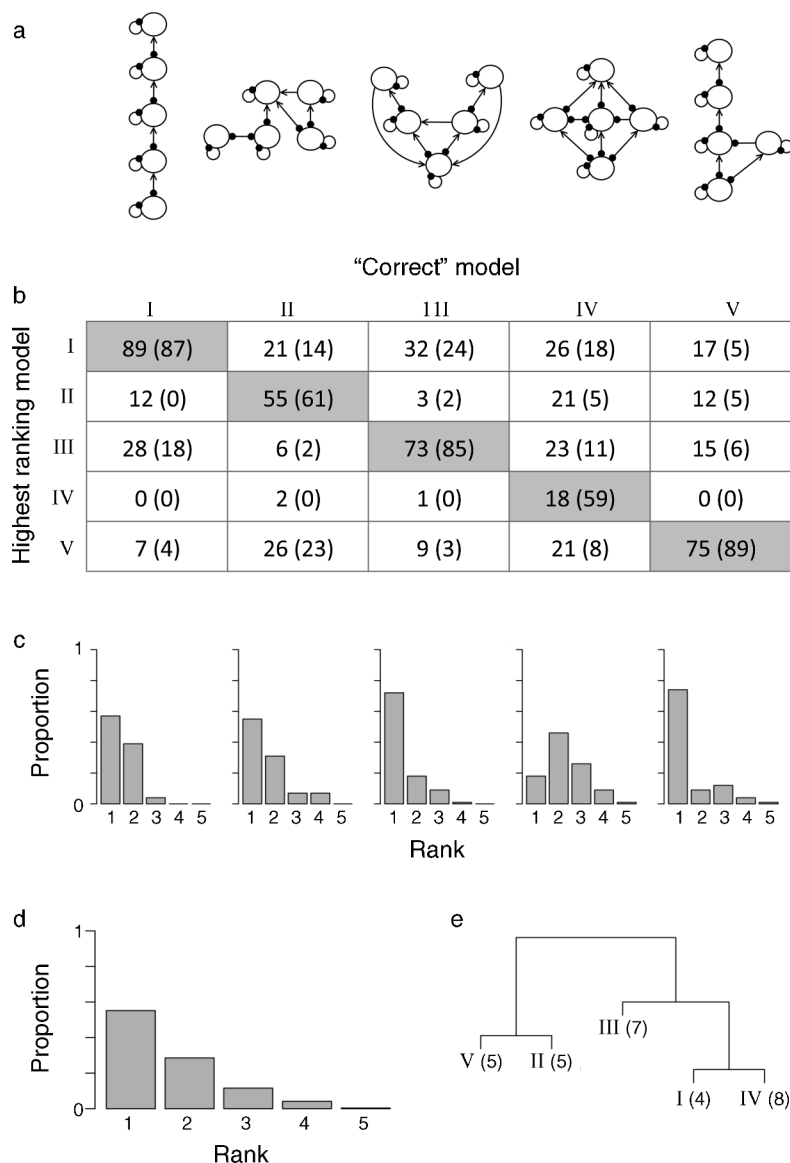
Fig. 5. Bayes model correctness test results for: (a) a set of five-variable models; (b) the confusion matrix for test results. Here, the frequency with which the assumed true model was given the highest probability of correctness (shaded cells) is compared with the number of cases where an alternative model ranked highest in terms of correctness probability. Values given in parentheses are for the case where three (rather than two) model components were monitored. Columns of the confusion matrix may sum to >100 (the number of repetitions) because it is possible to have equal (highest) probabilities for more than one model. Panel (c) shows the proportion of repetitions in which the true model ranked first, second, third, fourth, or fifth in terms of correctness probability. With the exception of model IV, the true model was ranked first for the largest proportion of repetitions. Aggregated rankings across all models are shown in panel (d), while panel (e) shows a classification of models (using correctness probabilities as a distance measure and complete linkage hierarchical clustering); the algorithm is best able to distinguish model III from other models and least able to distinguish models I and IV from each other.

the Raymond et al. (2011) Macquarie Island model. These alternative versions contained no uncertain linkages (i.e., each of the dashed lines in Fig. 3 was unambiguously either present or absent in each model) and were derived based on expert selection of three distinct, realistic representations of the Macquarie Island system (see Appendix B).

Applying the framework to compare models

We introduce the use of marginal likelihoods ($p(y \mid M)$, derived from simulation) and Bayes factors to make direct comparisons of alternative models that represent uncertainty about the structure of a system. Our approach is an alternative to the BBN approach introduced by Hosack et al. (2008) in which one or two

models are compared with a null model (which allocates equal probabilities of observing an increase, decrease, or no response across every possible prediction given a press perturbation). Specifically, we use simulations to compute marginal likelihoods for a given set of observations under a given press perturbation, i.e., $p(y \mid M)$ is the proportion of replications in a simulation (with random linkage weights) that generates perturbation responses that match with observations. We then compute Bayes factors as the ratio of these probabilities for alternative models (Eq. 3).

To demonstrate this Bayesian approach to model comparison, we use alternative versions of Hulot et al.'s (2000) functional group models (Fig. 2). For the eight-variable model we focus on uncertainty regarding the role of fish predators in the lake mesocosm system. Hulot et al. (2000) state that "according to the qualitative analysis of our model, the positive effects of nutrient enrichment on mineral phosphorus, periphyton, large herbivores and fish in mesocosms with fish cannot be explained without incorporating fish predation on invertebrate carnivores." We tested this assertion by calculating marginal likelihoods and the Bayes factor for two alternative versions of the eight-variable model; the original model as presented by Hulot et al. (2000; see Fig. 2b) and a version of the model with (1) the predation link between invertebrate carnivores (Carn-1) and fish (Carn-2) deleted, and (2) self-limitation for invertebrate carnivores (which is realistic because cannibalism is common in copepods; Bonnet et al. 2004; and copepods are included in the Carn-1 functional group). We calculated probabilities from 5 $\times$ 10^5 simulations (this large number was chosen to achieve convergence in probability values across replicate simulations) of a positive perturbation to phosphorus (Phos), and observed increases to each of phosphorus, periphyton (Alg-P), large herbivores (Herb-2), and fish (Carn-2). We did not simulate observations of zero change to any model variables for reasons discussed in our results.

We used the same approach to test the effect of adding self-limitation to invertebrate carnivores in the seven-variable lake mesocosm model (Fig. 2a). Here, we calculated probabilities and the Bayes factor from simulations of a positive perturbation to phosphorus and an observed decrease in invertebrate carnivores (as reported in Hulot et al. 2000). Finally, we use our Bayesian interpretation of the lake mesocosm models to highlight potential issues with the null model comparison approach used by Hosack et al. (2008).

Exploring prediction space

Simulation approaches to qualitative network analysis generate large data sets that capture model behavior under different combinations of parameter values and potential model structures. A range of analyses can be applied to explore these relationships, which do not necessarily require formal Bayesian interpretation. We

present four complementary examples of analysis tools for exploring prediction space from simulations. These tools address (1) the effects of observations on marginal likelihoods, (2) the relative power of model variables to distinguish between alternative model structures, (3) the influence of model linkages on perturbation outcomes, and (4) identifying groupings of model variables that respond similarly to perturbation.

Interactive tool.—We present a simple interactive tool for (1) exploring prediction space for a given model and (2) visualizing changes in marginal likelihoods under different perturbation and observation scenarios. The algorithm underlying this interface identifies and stores stable realizations of simulated community matrices, and displays changes in proportional outcomes across simulations depending on user queries. Probability values can be extracted from the interface to compute Bayes factors for model comparison. The example provided is for Hulot et al.'s (2000) eight-variable lake mesocosm model under differing assumptions about observed responses to phosphorous enrichment.

Mutual information.—Mutual information can be used for evaluating the combined prediction space from two or more alternative models, where predictions are weighted based on posterior model probabilities. This is similar to Hosack et al.'s (2008) use of mutual information for sensitivity analyses, wherein model variables are identified that, if their response to a particular perturbation scenario was known, would provide the most power to distinguish between alternative models.

The mutual information of two discrete random variables X and Y is defined as

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \qquad (4)$$

where $p(x, y)$ is the joint probability of X and Y, and $p(x)$ and $p(y)$ are the marginal probabilities of X and Y, respectively. We calculate mutual information for each variable in a set of alternative models (which are required to have identical nodes, but can have different linkages) by defining the set of accepted models (i.e., those that are stable) as X and the set of observed perturbation responses for a particular variable as Y. This effectively computes the mutual information for model variables and the model indicators M described in the *Approach: Simulation procedure*. Our algorithm simulates over the full set of alternative models so that the predictions used to calculate $I(X; Y)$ are weighted by posterior model probabilities. We demonstrate our algorithm using the two alternative versions of Hulot et al.'s (2000) eight-variable lake mesocosm model described in the *Approach: Applying the framework to compare models* and show how this approach could potentially be used for targeting observational studies to better inform understanding of ecological network structure.

Linkage influence.—Several authors have explored the effects of drawing random interactions weights from different distributions in simulation approaches to qualitative modeling (e.g., Hosack et al. 2008, Novak et al. 2011). Novak et al. (2011) also assess the effects of inaccuracy in interaction weight estimates for predictions of community responses to perturbation. However, the relative importance of various interactions in determining simulation outcomes for model variables of particular interest has not been assessed. We demonstrate the use of boosted regression trees (BRTs; Elith et al. 2008) for visualizing the relative influence of network linkages in determining perturbation responses in qualitative models. BRTs combine the strengths of regression trees (which relate a response to predictor variable through recursive binary splits) and boosting (an adaptive method for combining many simple models to give improved predictive performance). BRT models can be understood as additive regression models in which individual terms are simple trees, fitted in a stagewise fashion. As an example analysis, we examine the influence of linkages in the eight-variable lake mesocosm model on the response of edible algae to phosphorous enrichment.

Distance measures for perturbation responses.—Our final example of an approach to exploring prediction space focuses on the characterization of differences in the response of model components to perturbation scenarios. The method presented here has applications as an alternative means to visualize prediction space [i.e., the posterior predictive distribution $p(\hat{y}|M)$] under a given scenario or set of scenarios, and as a means for examining whether model variables might sensibly be aggregated. This latter application can also inform the representation of taxa in quantitative simulation models.

We constructed distance matrices for model components, based on qualitative responses to press perturbation scenarios, which can be visualized using cluster analysis (the method used here) or ordination analysis (e.g., nonmetric multidimensional scaling). As in any multivariate analysis, the choice of distance metric elicits particular information about the relationships between variables. Here, we implemented three alternative measures for the distance between nodes in terms of the sign of their response to press perturbations. If x, y are the responses ($+$, 0, $-$) of two nodes, and d the distance, the alternatives are

$$d(x, y) = |x - y| \qquad (d1)$$

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \text{ and } x, y \neq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (d2)$$

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases} \qquad (d3)$$

Under alternative ($d1$), the maximum distance value of $d = 2$ occurs when the response signs of two nodes are opposite, and $d = 1$ occurs for a response mismatch where either prediction is for zero change. By comparison alternative ($d2$) assumes $d = 0$ for a mismatch where either change is zero. Alternative ($d3$) gives $d = 1$ for any mismatch in response prediction (and $d = 0$ otherwise). Of these three alternatives, ($d1$) gives high distance scores for any mismatch in responses (although mismatches involving zero change receive a lower weighting), ($d2$) down-weighs mismatches for cases where one variable in a pair does not change (as compared with ($d1$)), while ($d3$) weighs all mismatches equally. Distance values d are summed across all models generated within a simulation, and across responses to either of (1) a particular perturbation scenario of interest, or (2) the perturbation of every node (in a node-by-node fashion).

Matrices of these distance measures, and their visualization using cluster or ordination analyses, are potentially useful for simplifying or classifying response types in complex models within many nodes and linkages. We therefore use this approach to make comparisons within the Macquarie Island model, both for the case where rabbits, rats, and mice are suppressed (the eradication scenario explored by Raymond et al. 2011), and for perturbations of every node in the model. By identifying groupings of variables that arise consistently across all three distance measures ($d1$), ($d2$), and ($d3$), we make suggestions about modeled species that might be either combined, or modeled using similar functional representations, in future modeling efforts for this system.

RESULTS AND DISCUSSION

Testing the Bayesian framework

Results for our Bayes correctness test, applied to a subset of the five-variable models used by Hosack et al. (2008), indicate that the algorithm performs well at distinguishing between alternative models (Fig. 5). In all cases but one, the correct model was assigned the highest probability of correctness in the greatest number of simulations (Fig. 5b, c). The model with the simplest structure (linear food chain model; model I in Fig. 5) was most readily distinguished from other models, while the model with the most complex structure (model IV in Fig. 5) was the only model that did not rank first in terms of the proportion of simulations in which it was identified as correct. We note, however, that model IV ranked first in ~20% of iterations, and that it ranked second in ~50% of cases.

Results presented in Fig. 5 are for the case where one randomly selected node was perturbed in each iteration, and two other randomly selected nodes were monitored. This scenario is conservative but realistic, given that logistical constraints often mean that responses to perturbation can only be observed for a subset of species or groups in real ecological systems. Sensitivity testing indicated that the algorithm's ability to distinguish between alternative models improved as the number of monitored nodes was increased. All five-
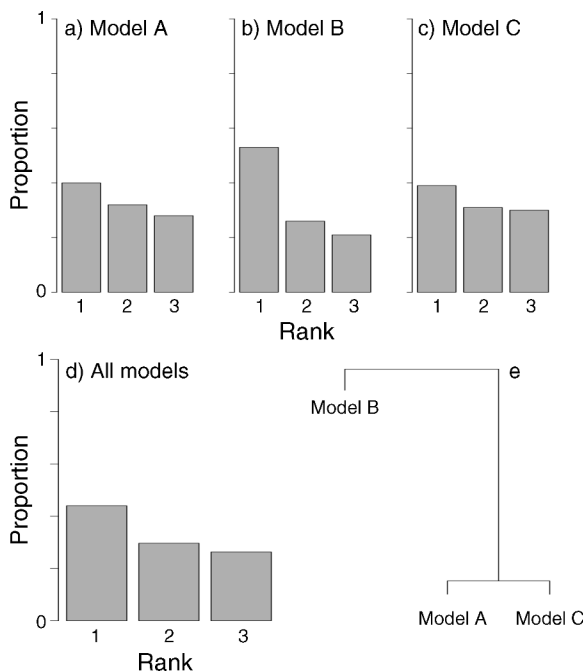
FIG. 6. Bayes model correctness test results for three alternative versions of the Macquarie Island model (models A, B, and C). Panels (a), (b), and (c) show the proportion of repetitions in which the true model ranked first, second, or third in terms of correctness probability. Aggregated rankings across all models are shown in panel (d), while panel (e) shows a classification of models (using correctness probabilities as a distance measure and complete linkage hierarchical clustering). While the algorithm was able to identify the selected true model as correct most frequently, the proportion of iterations in which the true model ranked first was ≤0.5 in all cases.

variable models were correctly identified in the majority of cases when the number of monitored nodes was increased from two to three (see values given in parentheses in Fig. 5b). We also applied the algorithm for a set of five 10-variable models (which were also a subset of those analyzed by Hosack et al. 2008), and found that the algorithm performed equally well at distinguishing models as for the five-variable case presented here. Results for both sensitivity testing and analysis of 10-variable models are provided in Appendix C.

Applying our correctness test to three alternative versions of the Macquarie Island model indicated that, while the algorithm was able to identify the selected true model as correct most frequently, the proportion of iterations in which the true model ranked first was ≤0.5 in all cases (Fig. 6a–d). There was therefore a relatively high number of cases in which models were confused by the algorithm. This is despite that the models were selected based on expert identification of three distinct (but realistic) realizations of the system shown in Fig. 3. This result suggests that, in some applications, alternative models may be only weakly distinguished. The use of such models (e.g., for decision making or conserva-

tion planning) in such cases should obviously take into account the full model suite, not just a single model.

The ability of our Bayesian algorithm to distinguish alternative models in the majority of cases (both for the set of loosely related five-variable models, and for the more closely related versions of the Macquarie Island model) suggests that we need not be overly concerned about our Bayesian approach performing poorly due to weakly informative priors for interaction coefficients. Nevertheless, it is possible to confuse closely related sets of models based on their overall behavior when nodes are randomly perturbed and monitored.

Applying the framework to compare models

We used simulation results to calculate marginal likelihoods $p(y \mid M)$ of observed changes for two versions of Hulot et al.'s (2000) eight-variable lake mesocosm model: (A) the original model as presented in their paper (and used by Hosack et al. 2008, in their BBN approach), and (B) a version of the model with self-limitation for invertebrate carnivores (Carn-1 in Fig. 2b) and with the predation link between invertebrate carnivores and fish (Carn-2) deleted. The marginal likelihoods for a positive perturbation to phosphorus (Phos), and observed increases for each of phosphorus, periphyton (Alg-P), large herbivores (Herb-2), and fish were 0.55 and 0.83 for the original (A) and modified (B) versions of the model, respectively. These probabilities give a Bayes factor for the original model (A) compared to our alternative model (B) of 0.66. Because this is the posterior odds in favor of model A compared with model B (Kass and Raftery 1995), and the value is less than one, we argue that model B provides a better description of the system given the observed perturbation responses.

Importantly, neither of the two eight-variable models that we examined were able to reproduce zero change for any of edible algae (Alg-1), protected algae (Alg-2), small herbivores (Herb-1), or invertebrate carnivores (Carn-1) under a positive press perturbation for phosphorus, despite the fact that Hulot et al. (2000) reported nonsignificant changes for these groups in response to increased phosphorus. We argue that the 90% support for model A indicated in Hosack et al.'s BBN analysis (which uses all of Hulot et al.'s observations as assumed inputs) is somewhat misleading; in particular the high level of support may have arisen because this model was only compared with a null case that assumed equal probabilities of positive, zero, and negative change for all variables. We also recommend caution in assuming that observations of nonsignificant (quantitative) change under perturbations in real systems equate to zero change in qualitative network analyses.

We performed similar calculations of marginal likelihoods and the corresponding Bayes factor for two versions of Hulot et al.'s (2000) seven-variable model for the mesocosm system without fish (Fig. 2a). The first

version was the model presented by Hulot et al. (2000), while the second included a self-limitation effect for invertebrate carnivores (as in our modified eight-variable model). The marginal likelihoods for an observed decrease in invertebrate carnivores under phosphorus enrichment were 0.56 and 0.51 for the original and modified versions of the model, respectively. In this case the posterior odds (Bayes factor) in favor of the original model are 1.1; this is weak evidence that the original model better captures the system dynamic. These combined findings suggest that additional information is needed about potential self-limiting effects for invertebrate carnivores in the experimental mesocosms examined by Hulot et al. (2000). It may also be the case that a more complex model with additional functional groups is needed to adequately describe the system.

Exploring prediction space

Interactive tool.—We complement our analysis of likelihoods and Bayes factors through the use of an interactive tool for visualizing changes in outcome probabilities under various combinations for perturbation and observation of model variables. Fig. 7 provides a demonstration of how our "node selector" interface can be used to explore prediction space, and to assess marginal likelihoods for particular sets of observations, for a given model. The example provided is for Hulot et al.'s (2008) original eight-variable mesocosm model under a positive perturbation to phosphorus. Marginal likelihoods decrease as additional observations are added, with the probability value shown in Fig. 7b corresponding with the value of 0.55 given previously (*Results and discussion: Applying the framework*) in our calculation of the Bayes factor for eight-variable models. Notably, adding an observed increase for protected algae (Alg-2) under phosphorous enrichment (Fig. 7c) to mimic the marginally significant increase reported by Hulot et al. (2000) for this group, gives a marginal likelihood of only 0.36.

Mutual information.—We used mutual information to rank variables in Hulot et al.'s (2000) lake mesocosm system in terms of their ability to distinguish between alternative models (Table 1). High values of mutual information (in relative terms) for edible algae (Alg-1), fish (Carn-2), and invertebrate carnivores (Carn-1) imply that knowledge of the response of these variables under increased phosphorus would be most useful for distinguishing between the original eight-variable model described by Hulot et al. (2000) and the variant we introduced in the *Approach: Applying the framework to compare models* (with self-limitation instead of predation for invertebrate carnivores; see "two model comparison" in Table 1).

Both our method for calculating mutual information and our findings for the lake mesocosm system are distinct from those presented in Hosack et al. (2008). While the method presented here calculates mutual information from aggregate simulations across all

alternative models (and in doing so, applies weightings for posterior model probabilities), Hosack et al. (2008) use an algorithm that is part of the BBN software Netica, and which calculates mutual information from conditional probabilities embedded in a Bayesian network. A further difference is that Hosack et al. (2008) compare Hulot et al.'s (2000) lake mesocosm model with a "null" model that predicts equal probabilities for qualitative responses (increase, decrease, or no change) for all model variables under a given perturbation scenario. In order to directly compare results from our mutual information algorithm with those reported by Hosack et al. (2008) we mimicked the behavior of their null model and recalculated mutual information values, again under a press perturbation scenario of increased phosphorus. While the magnitudes of mutual information values differ between the two algorithms, the ranking of variables is the same (except for a reversal of the order for Alg-2 and Alg-1; see "null model comparison" in Table 1).

Marked differences in the ranking of variables between the two sets of model comparisons provided in Table 1 emphasize the need for careful thought in choosing pairs of models to compare (both in terms of mutual information, and when applying the Bayesian framework outlined in *Approach: Bayesian framework for interpretation*). In particular, if some form of null model is to be used as the basis for comparison, then consideration should be given to what form of null model is most informative. We argue that the null model used by Hosack et al. (2008) is the least informative of a broad range of possibilities (e.g., a null model that retains the configuration of assumed linkages but not their direction, or one that conserves linkages between but not within trophic levels), and that choosing variables to distinguish against this null based on mutual information is somewhat trivial. Importantly, it is unclear that there is a form of null model that is equally applicable across all cases. We therefore recommend that a null model, if required, should be configured on a case-by-case basis.

Linkage influence.—We used simulation results (based on uniform random linkages) to examine the relative influence of network linkages in Hulot et al.'s (2000) eight-variable lake mesocosm model. Specifically, we assessed linkage influence for the response of edible algae to phosphorus enrichment. Relative influence outcomes from boosted regression tree analysis indicate that the positive effects of phosphorus on edible algae (Alg-1) and protected algae (Alg-2) are most important in affecting outcomes for edible algae, followed by the negative effects of consumption by large herbivores (Herb-2) on edible and protected algae (Fig. 8). This is consistent with Hulot et al.'s (2000) original analysis of the seven-variable mesocosm system where they indicate that the qualitative effects of nutrient enrichment on edible algae depend on the ratio of the direct effects of edible and protected algae on large herbivores, and on
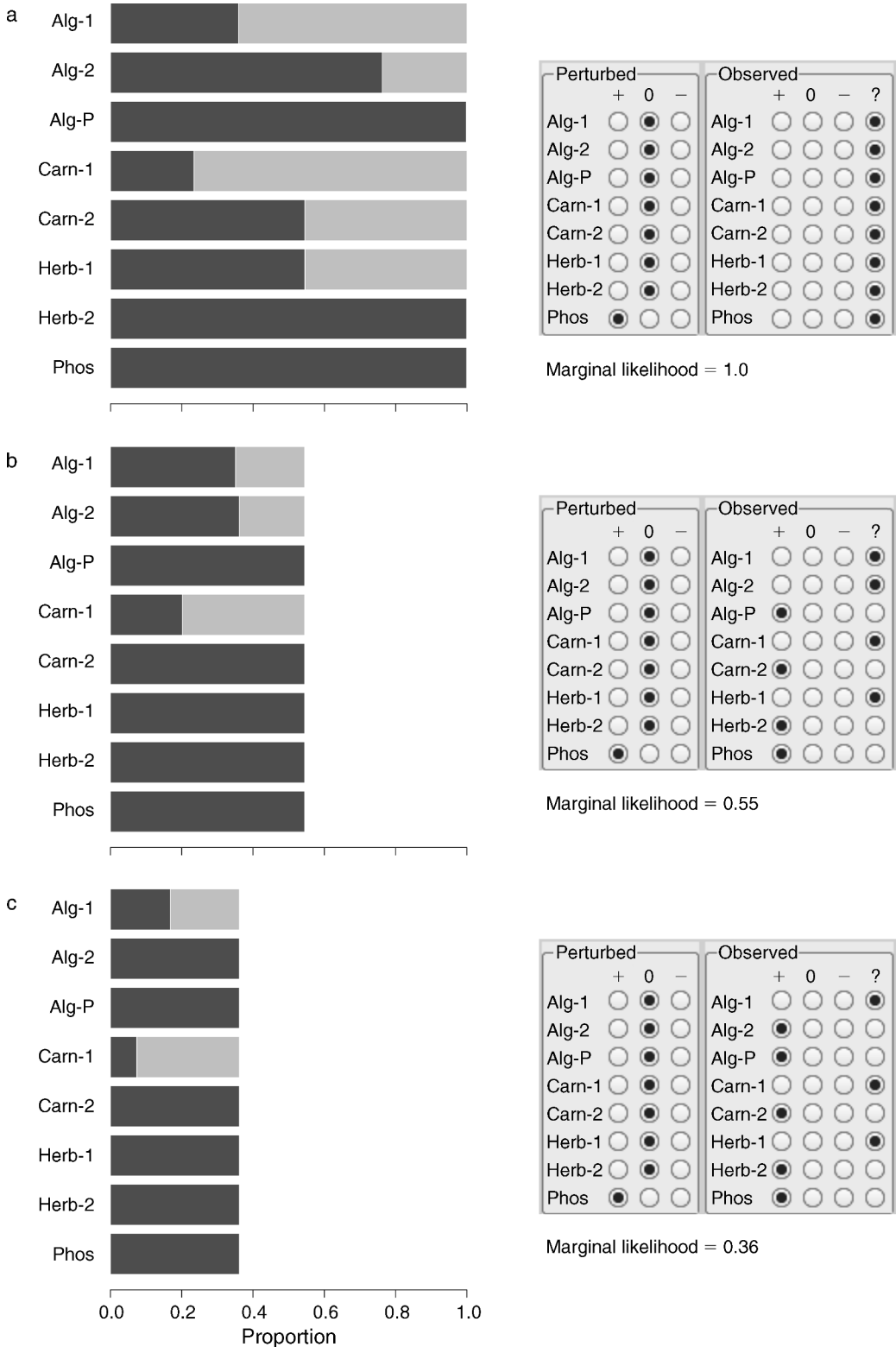
FIG. 7.   Example usage of a "node selector" interface to explore outcome probabilities for Hulot et al.'s (2000) eight-variable lake mesocosm model under a positive perturbation to phosphorus. Proportional outcomes are shown on the left (from 5×10^5 simulations), where dark gray represents a positive change, and light gray is a negative change. Corresponding selections for the node selector are shown on the right. In panel (a), no assumptions are made regarding observations for model variables under an increase to phosphorus (hence all variables are assigned a question mark for "observed"), and the marginal likelihood $p(y \mid M) = 1$.

TABLE 1. Mutual information for variables (nodes) in alternative versions of Hulot et al.'s (2000) eight-variable lake mesocosm model under a perturbation scenario of increased phosphorus.

Node	Mutual information†
Two model comparison	
Alg-1	0.078
Carn-2	0.064
Carn-1	0.048
Alg-2	0.016
Herb-1	0.002
Herb-2	0.000
Phos	0.000
Alg-P	0.000
Null model comparison	
Alg-P	0.263 (0.358)
Phos	0.263 (0.358)
Herb-2	0.263 (0.358)
Carn-1	0.050 (0.039)
Alg-2	0.048 (0.014)
Alg-1	0.016 (0.022)
Herb-1	0.002 (0.000)
Carn-2	0.002 (0.000)

Notes: Higher values of mutual information imply that knowledge of the response for a given variable would improve power to distinguish between models. The "two model comparison" is for Hulot et al.'s original model and the variant introduced in *Approach: Applying the framework to compare models*, while the "null model comparison" parallels the analysis conducted by Hosack et al. (2008) (mutual information values calculated by Hosack et al. are provided in parentheses).

† Results presented here are from 10^5 simulated samples. While values vary within ±0.002 between simulations, the ranking of nodes is consistent.

the ratio of the nutrient uptakes of edible and protected algae (although the authors were unable to resolve the importance of interaction effects for the eight-variable model).

Distance measures for perturbation responses.—In our final example of an approach for exploring prediction space we apply classification analysis to distance matrices calculated for the Macquarie Island model under a perturbation scenario for suppression of mice, rats, and rabbits. This example demonstrates how our distance-based approach can be used as an alternative means for visualizing and grouping similar sets of outcomes for model variables (Fig. 9a, b). For simplicity, we only present results for distance measure (d1), and identify patterns and groupings that arise consistently for all three measures, (d1), (d2), and (d3). Clusters in Fig. 9b that are consistent between classifications for each of the three distance measures are (1) rats, penguins, and rabbits (all of which show a high proportion of negative changes in Fig. 9a), and (2) giant petrels, albatrosses, and redpolls and starlings (redpolls

and starlings were treated as a variable in the model), burrow-nesting seabirds, and tall tussock vegetation (all of which show a high proportion of positive changes in Fig. 9a).

For the case where distance measures were calculated across a broad range of perturbations (i.e., for every node perturbed in a node-by-node fashion), there was also some consistency in classification clusters from the three distance measures (shown by dashed boxes in Fig. 9c). This information could inform the representation of species in quantitative simulation models (i.e., species that consistently respond in a similar manner might be combined or modeled in a similar fashion), or might be used to identify groups of species that are particularly vulnerable, or particularly robust, to change within the
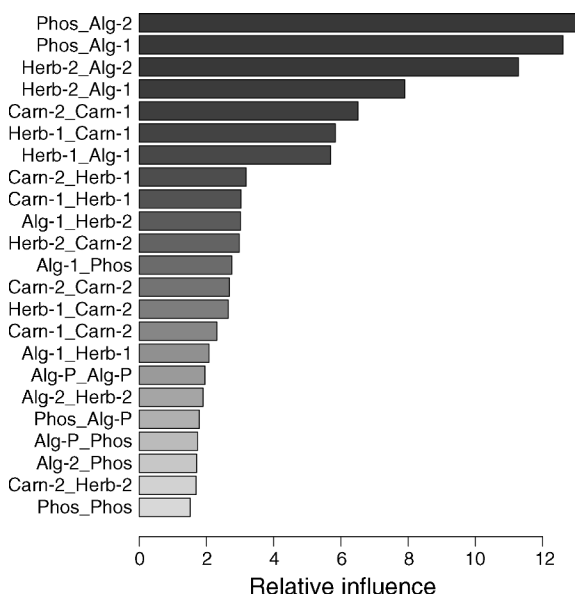

FIG. 8. Relative influence of network linkages in Hulot et al.'s (2000) eight-variable lake mesocosm model on the response of edible algae (Alg-1) to phosphorus enrichment. Results are from boosted regression tree analysis based on 1000 simulations and with 1300 fitted trees. Relative influence is determined from the number of times a variable is selected for splitting in regression trees, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Elith et al. 2008). The *y*-axis notation refers to directed edges "from" one model variable (the first in each pair) "to" another (the second in each pair). The positive effects of phosphorus on protected algae and edible algae (i.e., Phos_Alg-2 and Phos_Alg-1) have the greatest relative influence on the response of edible algae to nutrient enrichment, followed by the negative effects of grazing by large herbivores on protected algae and edible algae (Herb-2_Alg-2 and Herb-2_Alg-1). See Fig. 2 for explanation of abbreviations.

←
Panel (b) corresponds with the observations made by Hulot et al. (2000) in their mesocosm experiment; in this case the marginal likelihood drops to 0.55. Finally, panel (c) shows a hypothetical case where an increase is observed for protected algae (Alg-2; note that Hulot et al. observed a marginally significant increase for this group), for which the marginal likelihood decreases further to 0.36.
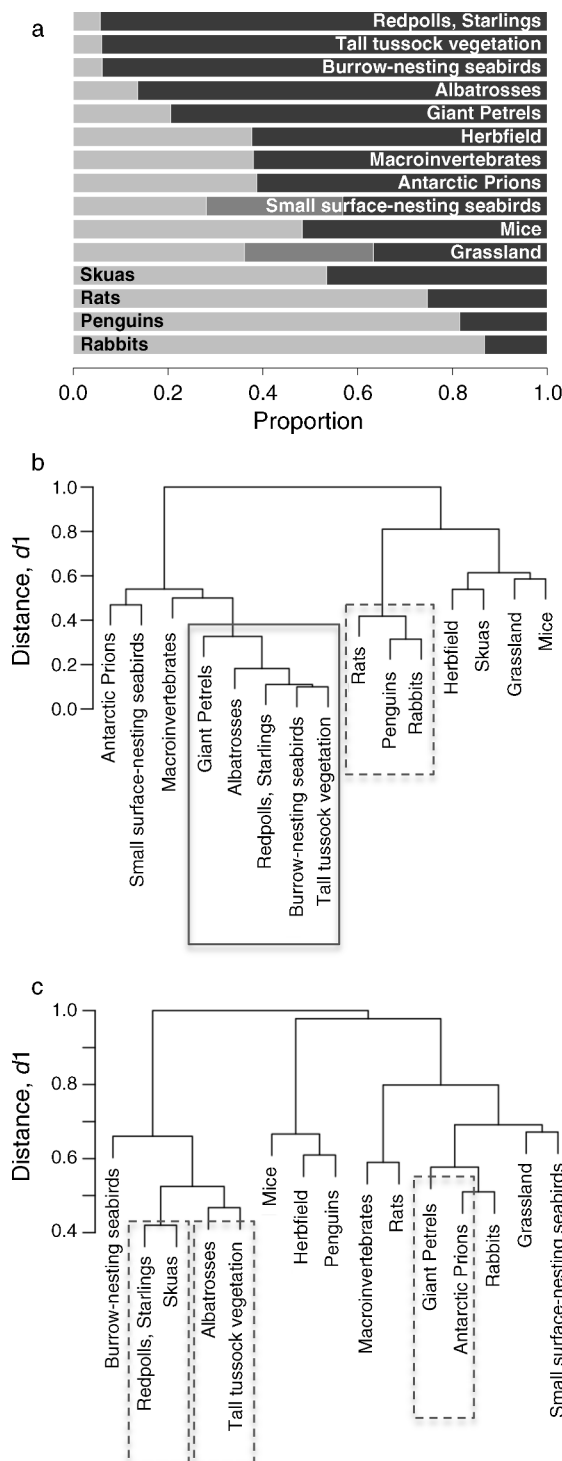
system. For the Macquarie Island system, our results suggest that a quantitative simulation model might usefully represent starlings and skuas, or petrels and prions, in a similar fashion, while albatrosses and tall tussock vegetation might be expected to respond in the same qualitative fashion to future changes in the system.

CONCLUSIONS

We have presented an integrated framework for interpreting outcomes from simulations of alternative qualitative network models. While simulation approaches to analyzing qualitative network models have been explored previously (e.g., Dambacher et al. 2003a, Hosack et al. 2008, Novak et al. 2011, Raymond et al. 2011), we are the first to provide methods for consistent Bayesian interpretation of simulation outcomes, and to relate this to model structure uncertainty. Importantly, our framework provides a means to assess and compare alternative model formulations through the use of Bayes factors, and to explore how posterior probabilities change when observational information is considered.

Our example approaches for visualizing prediction space from simulations, i.e., our node selector interface, our use of mutual information and regression trees to examine node and linkage influence, respectively, and our distance measures for perturbation responses, complement our Bayesian framework in providing a complete toolbox for addressing model uncertainty in qualitative network analysis. The need for greater depth in analyses of simulation results for qualitative network models has been acknowledged previously; our integrated framework is the first to fill this gap. While the examples we provide use models of relatively well understood ecological systems that have been examined by other authors, the potential ecological applications of approaches presented here are clearly much broader. Given the wide-ranging advantages to using qualitative modeling approaches to formulate and test fundamental understanding of system structure, we see the application of our framework as an important step toward developing better quantitative understanding of ecological systems.

FIG. 9. (a) Proportional outcomes from 1000 simulations for the Macquarie Island model under a perturbation scenario for suppression of mice, rats, and rabbits. Light gray represents a negative change, medium gray is no change, and dark gray is a positive change. The corresponding classification of node responses using distance measure (d1), as defined in *Approach: Distance measures for perturbation responses*, for this perturbation scenario is shown in panel (b); clusters generally correspond with sets of nodes that show similar responses in panel (a). Panel (c) shows a classification, again using distance measure (d1), for node responses across all possible node-wise perturbations. In panels (b) and (c), groupings that arise consistently for all three distance measures, (d1), (d2), and (d3), and that are of potential interest, are highlighted by solid and dashed boxes. Classifications were generated using hierarchical cluster analysis with complete linkage.

Literature Cited

Bender, E. A., T. J. Case, and M. E. Gilpin. 1984. Perturbation experiments in community ecology: theory and practice. Ecology 65:1–13.

Bergstrom, D. M., A. Lucieer, K. Kiefer, J. Wasley, L. Belbin, T. K. Pedersen, and S. L. Chown. 2009. Indirect effects of invasive species removal devastate World Heritage Island. Journal of Applied Ecology 46:73–81.

Berlow, E., et al. 2004. Interaction strengths in food webs: issues and opportunities. Journal of Animal Ecology 73:585–589.

Bonnet, D., J. Titelman, and R. Harris. 2004. *Calanus* the cannibal. Journal of Plankton Research 26:937–948.

Copson, G. R. 2002. Integrated vertebrate pest management on subantarctic Macquarie Island 1997–2002, Final Report. Department of Primary Industries, Water and Environment, Hobart, Tasmania, Australia.

Dambacher, J. M., D. J. Gaughan, M.-J. Rochet, P. A. Rossignol, and V. M. Trenkel. 2009. Qualitative modelling and indicators of exploited ecosystems. Fish and Fisheries 10:305–322.

Dambacher, J., R. Levins, and P. Rossignol. 2005. Life expectancy change in perturbed communities: derivation and qualitative analysis. Mathematical Biosciences 197:1–14.

Dambacher, J. M., H. W. Li, and P. A. Rossignol. 2002. Relevance of community structure in assessing indeterminacy of ecological predictions. Ecology 83:1372–1385.

Dambacher, J., H. Li, and P. Rossignol. 2003a. Qualitative predictions in model ecosystems. Ecological Modelling 161:79–93.

Dambacher, J., H. Li, J. Wolff, and P. Rossignol. 1999. Parsimonious interpretation of the impact of vegetation, food, and predation on snowshoe hare. Oikos 84:530–532.

Dambacher, J. M., H. K. Luh, H. W. Li, and P. A. Rossignol. 2003b. Qualitative stability and ambiguity in model ecosystems. American Naturalist 161:876–888.

Dowding, J. E., E. C. Murphy, K. Springer, A. J. Peacock, and C. J. Krebs. 2009. Cats, rabbits, Myxoma virus, and vegetation on Macquarie Island: a comment on Bergstrom et al. (2009). Journal of Applied Ecology 46:1129–1132.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77:802–813.

Geweke, J. 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57:1317–1339.

Hosack, G. R., K. R. Hayes, and J. M. Dambacher. 2008. Assessing model structure uncertainty through an analysis of system feedback and Bayesian networks. Ecological Applications 18:1070–1082.

Hulot, F. D., G. Lacroix, F. Lescher-Moutoué, and M. Loreau. 2000. Functional diversity governs ecosystem response to nutrient enrichment. Nature 405:340–344.

Kass, R., and A. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90:773–795.

Levins, R. 1974. The qualitative analysis of partially specified systems. Annals of the New York Academy of Sciences 231:123–138.

Marzloff, M. P., J. M. Dambacher, C. R. Johnson, L. R. Little, and S. D. Frusher. 2011. Exploring alternative states in ecological systems with a qualitative analysis of community feedback. Ecological Modelling 222:2651–2662.

Metcalf, S. J., M. B. Pember, and L. M. Bellchambers. 2011. Identifying indicators of the effects of fishing using alternative models, uncertainty, and aggregation error. ICES Journal of Marine Science 68:1417–1425.

Montaño-Moctezuma, G., H. W. Li, and P. A. Rossignol. 2007. Alternative community structures in a kelp–urchin community: a qualitative modeling approach. Ecological Modelling 205:343–354.

Novak, M., J. T. Wootton, D. F. Doak, M. Emmerson, J. A. Estes, and M. T. Tinker. 2011. Predicting community responses to perturbations in the face of imperfect knowledge and network complexity. Ecology 92:836–846.

Puccia, C. J., and R. Levins. 1985. Qualitative modeling of complex systems: an introduction to loop analysis and time averaging. Harvard University Press, Cambridge, Massachusetts, USA.

R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raymond, B., J. McInnes, J. M. Dambacher, S. Way, and D. M. Bergstrom. 2011. Qualitative modelling of invasive species eradication on subantarctic Macquarie Island. Journal of Applied Ecology 48:181–191.

Scott, J., and J. Kirkpatrick. 2008. Rabbits, landslips and vegetation change on the coastal slopes of subantarctic Macquarie Island, 1980–2007: implications for management. Polar Biology 31:409–419.

Spear, R., and G. Hornberger. 1980. Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research 14:43–49.

Supplemental Material

Appendix A

Qualitative modeling derivations (*Ecological Archives* M082-019-A1).

Appendix B

Alternative model representations of the Macquarie Island system (*Ecological Archives* M082-019-A2).

Appendix C

Bayes correctness test: additional results (*Ecological Archives* M082-019-A3).

Supplement

Example R code and models (*Ecological Archives* M082-019-S1).