Received: 4 June 2021 | Accepted: 1 October 2021

DOI: 10.1111/2041-210X.13742

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

REVIEW ARTICLE

A model selection approach to structural equation modelling: A critical evaluation and a road map for ecologists

Mario Garrido[1] | Scott K. Hansen[2] | Rami Yaari[3] | Hadas Hawlena[4]

[1]Jacob Blaustein Center for Scientific Cooperation, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

[2]The Department of Environmental Hydrology and Microbiology, The Zuckerberg Institute for Water Research (ZIWR), The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

[3]The Biostatistics and Biomathematics Unit, Sheba Medical Center, Tel Hashomer, Israel

[4]Mitrani Department of Desert Ecology, The Swiss Institute for Dryland Environmental and Energy Research, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

Correspondence
Hadas Hawlena
Email: hadashaw@bgu.ac.il

Funding information
Israel Science Foundation, Grant/Award Number: 1391/15; Kreitman School of Advanced Graduate Studies; Blaustein Center for Scientific Cooperation

Handling Editor: Claudia Coleine

Abstract

1. Structural equation modelling (SEM) can illuminate complex interaction networks of the sort found in ecology. However, selecting optimally complex, data-supported SEM models and quantifying their uncertainty are difficult processes. To this end, we recommend a formal model selection approach (MSA) that uses information criteria.
2. Using a suite of numerical simulations, we compare MSA-SEM against two traditional methods.
3. We find that MSA-SEM exhibits superior, unbiased results under the suboptimal realistic conditions characteristic of ecological studies. We then provide a road map for MSA-SEM and demonstrate its use via a case study.
4. We illustrate the unique abilities of SEM to confirm a network structure within the realm of known causal pathways and delineate the boundaries within which MSA-SEM should be applied.

KEYWORDS

causal analyses, information criteria, model selection, multimodal inference, network structure, path analysis, review, structural equation modelling

1 | INTRODUCTION

1.1 | Inferring complex causal networks from data

Ecologists commonly explore relationships within and between variables featuring complex, imperfectly known interactions. Structural equation modelling (SEM) offers an ideal approach for this task (Box 1), as it aims at creating a single causal network encoding the interactions of multiple variables (Kline, 2016; Shipley, 2016). The SEM family (see Glossary) includes many distinct approaches, differing in their relative focus on observed and unobserved variables, fitting approach, and abilities to deal with longitudinal and hierarchical data. Of all the family members, path analysis constitutes the most popular and accessible technique used by ecologists (e.g. Figure 1).

There are two properties that distinguish SEM from traditional univariate generalized linear model (GLM) analyses and alternative multivariate techniques. First, SEM can be used to test hypotheses about network structure, decomposing total effects into direct and indirect and dealing with reciprocal effects, multiple observable and unobservable variables (see Glossary), multiple outcomes, and the relationships within and between them (Barringer et al., 2013; Hoyle, 2012). Such networks are useful in quantifying cascading effects and in exploring potential inter-dependency within and between intrinsic and extrinsic factors (e.g. Chen et al., 2013; Duffy et al., 2015). Second, in SEM, relationships between variables are interpreted as causative rather than associative (Box 1). In addition, SEM techniques can handle a wide range of conditions—data with missing cells, a diverse array of residual distributions and different

GLOSSARY

Akaike information criterion index (AIC)

The most common information theory index used for model selection procedures. It estimates the relative amount of information lost by a given model based on the likelihood function (L) and the number of estimated parameters (k). The formula is $AIC = 2k - 2\ln(L)$. The lower this value, the higher the quality of the model.

Akaike information criterion index corrected for the sample size (AICc)

AIC corrected for small sample sizes (n). The formula is $AICc = AIC + \frac{2k^2+2k}{n-k-1}$.

Bayesian information criterion index (BIC)

An information theory index that is similar to *AIC* but uses another penalty for the number of estimated parameters. The formula is $BIC = \ln(n)k - 2\ln(L)$.

Sample-size-adjusted Bayesian information criterion index (aBIC)

BIC index with a modified (smoother) penalty function for sample size effect. The formula is $aBIC = \ln(\frac{n+2}{24})k - 2\ln(L)$.

Composite variables

A type of unobservable variables, which are made up of the total combined influence of indicator variables based on given weights. In a SEM diagram, these variables are denoted by a hexagon, and the arrows from the indicator variables lead into them. For example, in Eldridge et al. (2017), grazing was depicted as a composite variable composed of the combined effects of current and historic livestock grazing, together with those of other herbivores. Note that in the case of composite variables, the indicators (e.g. relative nutrient concentration) do not necessarily share a common cause.

Latent variables (hidden variables, hypothetical variables or hypothetical constructs)

A type of unobservable variables, which are not directly observed but are inferred from the indicator variables. In a SEM diagram, these variables are denoted by an oval, and the path arrows lead from them to the indicator variables. For example, Liu et al. (2016) used the functional traits of trees as a latent variable inferred from four indicator variables (wood and crown diameter, height and hydraulic conductivity). They tested the hypothesis that basic differences between functional traits are the common cause for the differences in the measured traits. The error associated with the indicator variables implies that they constitute an imperfect approximation of the latent variable, and presumably, other factors are manifested in the latent variable (e.g. root complexity).

Model relative weight (w_i; Akaike weights)

The difference in the information theory index scores between each model and the best model (i.e. the model with the lowest information theory index score), normalized across the set of candidate models to sum to one (or 100%) and thus interpreted as a probability.

Multicollinearity

The inclusion of highly correlated independent variables in regression-type models. Our rule of thumb in the case study was that correlation coefficients ($|r|$) that are equal to or greater than 0.6 denote multicollinearity.

Multivariate normality

A requirement for using the maximum likelihood (ML) or general least-squares (GLS) calibration approaches (Box 1), according to which the observed variables should be jointly multi-Gaussian. This condition occurs when all independent sources of randomness are themselves normally distributed.

Nested models

A collection of models is nested if for each pair of models in the collection, one's free parameter set is a proper subset of the other's.

Observed variables

Variables, also termed measured variables or indicators, have a measured quantity that was estimated during data collection. In a SEM diagram, these variables are denoted by rectangles.

Path

A direct relationship between two variables in SEM. The strength of the relationship is defined by the path coefficient.

Saturated SEM model

A model, also termed a meta-model or SEMM, that includes the maximum number of paths tested in a given set of competing models. This model should reflect the study design and attributes of the system studied.

SEM family

The SEM family includes a diverse array of approaches, differing in their relative focus on observed and unobserved variables, the fitting approaches, and their abilities to deal with longitudinal and hierarchical data. Confirmatory and exploratory factor analyses

(CFA and EFA; Kline, 2016), latent variable SEM (Loehlin & Beaujean, 2016) and confirmatory composite analysis (CCA; Schuberth et al., 2018) focus on, or at least include links between observable and unobservable variables, whereas path analysis include solely observable variables. Longitudinal SEM (Duncan et al., 2013; Newsom, 2015) models are used for exploring longitudinal data, and hierarchical Bayesian SEM models are used for exploring data that include variables at both the individual and the group levels, with both within- and between-group variation and covariation (Hox et al., 2017). Most SEM models are calibrated based on the covariance matrix (Box 1), but partial least squares path modelling (PLS; Dijkstra & Henseler, 2015) and piecewise SEM (Lefcheck, 2016) are variance-based, which make them more resistant to violation of the parametric tests.

Unobservable variables

Abstract variables that are part of the theory but not in the dataset, and thus are only inferred from the observed variables. They include latent and composite variables.

Variance–covariance matrix

A matrix with the different variables simultaneously appearing both vertically and horizontally, where the diagonals are the variances of each variable and the off-diagonals are the covariances between each pair.

data designs (e.g. hierarchical models and repeated measure design; Newsom, 2015; Shipley, 2016; Vinzi et al., 2010). These properties make it a useful framework for ecologists (Fan et al., 2016).

Once SEM has been chosen, two challenges are the selection of an optimally complex, data-supported model and the quantification of its uncertainty. Often there are multiple plausible models, each representing a hypothesis or a set of hypotheses, and one wishes to select the model that best accounts for the observed data. Traditionally in SEM, selection of the best model depends on the concept of dichotomous hypothesis testing. In the basic strictly confirmatory approach (SCA), the saturated SEM model (see Glossary) is tested, and only the paths (see Glossary) with significant coefficients are considered. The backward elimination approach to SEM broadens the parameter space dimension, by competing between nested models (see Glossary), in a process analogous to backward elimination in univariate models (BEA, also termed stepwise SEM; e.g. Bentler & Bonett, 1980; Tarka, 2018).

However, while these traditional approaches are straightforward, they often suffer from the dependence of the results' interpretations on an artificial probability threshold (usually $p < 0.05$), and they are expected to depend on the level of model complexity (i.e. the number of parameters and the number and complexity of interactions) and on the sample size (Amrhein et al., 2019). Adding parameters to the SEM model cannot weaken, and will generally improve, the model fit. However, the more degrees of freedom in the model, the less it is constrained by a given dataset, and the less reliable the predictions made from it (i.e. the problem of overfitting). Furthermore, the outcome of both approaches is a single model that depends on the sequence of parameter or path deletions (Bentler & Bonett, 1980; Tarka, 2018) and ignores model uncertainty (Mundry & Nunn, 2009; Whittingham et al., 2006).

1.2 | Using information criteria to choose between alternative SEM models

Using a formal model selection approach or MSA (i.e. information theoretical approach to model selection and multimodel interference) can resolve many of these calibration difficulties, providing an unbiased model inference, independent of test power and parameter space dimension, as well as uncertainty quantification (Burnham & Anderson, 2002; Wu et al., 2020). The information criteria are designed to address this matter of optimal model selection.

In this approach, the researcher develops an a priori set of hypotheses (represented as alternative models) and quantifies the databased evidence for ranking each hypothesis. The models that minimize the loss of information about the full reality, as reflected in a lower Akaike information criterion (AIC; see Glossary) index or any other information theory index, receive the highest ranks. The information theory indices rank the models by balancing between fit and precision, considering model complexity and avoiding the use of artificial thresholds. In some cases, the data are best explained by a single model, and results can be inferred solely from it. In other cases, information should be inferred from the set of models that best fit the data, and the results are interpreted considering the relative likelihood of each model, given the data and the set of models (e.g. Messika et al., 2017). These advantages suggest that a combined MSA-SEM approach can be a powerful tool to evaluate multiple causal hypotheses regarding complex interactions between variables.

Model selection approach allows the comparison of non-nested models and can compare models that do not have any particular structural relationship to one another (Burnham & Anderson, 2002). Furthermore, MSA does not require any of the models compared to be the 'true' SEM model (Lin et al., 2017). It is possible to compare SEM models with different structures, so long as each of them is calibrated against the same empirical covariance matrices representing the relationships between variables (see 'an MSA-SEM road map').

1.3 | Goals for this work

The formal MSA has become a popular paradigm in ecology, but the use of MSA with SEM is comparatively rare. In addition, the current

BOX 1 SEM overview

SEM basics. The major goal of ecologists describing a system with SEM is to uncover key relationships and mechanisms explaining the behaviour of natural systems. Each SEM model is represented by matrix equations, which can be pictured as a directed graph consisting of variables that may depend on one another, constructed without loops (i.e. it is impossible to start from and end with the same variable; Figure 1). When the causal analysis step is correctly conducted (Figure 2), the arrows of the graph represent causality. Some variables are independent (only pointed towards other variables), others are dependent (only pointed at by other variables), and the rest are mediators (both pointed to and pointed at by other variables). The independent variables may be correlated (indicated by bidirectional curved arrows), but the causal nature of this correlation is considered as external to the SEM and is not explained by the graph.

***SEM operation**. Once the SEM equations and diagram are ready, one should determine the numerical quantities defining the model, namely the path coefficients and the covariance structure of the independent variables. The goal of this procedure is to choose θ (the vector containing all the parameters that define the SEM model), which would minimize the distance between S (the empirical covariance matrices representing the relationships between variables) and $\sum(\theta)$ (the estimated variance–covariance matrix [see Glossary] of all random variables in a given SEM model). Consequently, for each SEM model, the calibration is global, and equations are solved simultaneously, yielding path coefficients that best correspond to the matrix S. There are multiple calibration approaches, most belonging to the maximum likelihood (ML) and the least squares (LS) families. Traditionally, the commons are ML and GLS (general least-squares), which relies on multivariate normality (see Glossary; Hoyle, 2012; Kline, 2016). Today, different versions that do not embed this assumption prevail (Lee, 2007; Olsson et al., 2000; Sarstedt et al., 2017). Different SEM formulations include different sources of uncertainty. The Bentler–Weeks and LISREL formulations assume that all independent, mediator and dependent variables participate in $\sum(\theta)$ (Lee, 2007), whereas in Asparouhov and Muthén (2009) formulation, the independent variables are not included there.

SEM boundaries. SEM calibration is possible only if $\sum(\theta)$ and S are defined in terms of the same variables. A common misperception about SEM is that it can defer to the data and will automatically choose the best data-supported model. Bollen (1989) demonstrated that this is wrong by presenting 10 SEM models with different independent variables and causal structures, all supported by the same S matrix. Considering that correlations between independent variables are permitted, it is reasonable that a degenerate SEM model, with all variables treated as independent and mutually correlated, can be made to fit any number of variables. Accordingly, given a plausible set of candidate SEM models, MSA can help us select the most data-supported models, but it cannot detect the optimal model *ex nihilo*. Furthermore, since the calibration relies solely on covariance data, it is not possible to infer the direction of causation; one can only specify it. Taken together, the causal modelling and model selection processes are essential for the correct use of SEM (Figure 2).

*There are multiple alternative model fitting approaches, and MSA-SEM can work with any metric of model fit, but here for simplicity, we focus on the most widely implemented one in SEM software.

lack of consistency in MSA-SEM use increases the likelihood of encountering pitfalls, restricts comparability between studies and makes the conclusions and interpretations of each study more subjective. Here, we present a critical evaluation of MSA-SEM, which both presents new simulation results and introduces the subject at a level accessible to ecologists. The work's two major contributions are the following:

1. A large-scale simulation study, which demonstrates that the performance of MSA-SEM under realistic conditions is superior to those of the two traditional approaches to SEM selection, highlighting its importance for ecologists.
2. A road map for the use of MSA-SEM by ecologists, which emphasizes both the strengths and the limitations of the approach, outlining best practices and presenting a detailed case study.

2 | MODEL SELECTION VERSUS LEGACY ALTERNATIVES: A SIMULATION STUDY

2.1 | Study outline

We conducted a simulation study to compare the performances of MSA-SEM, SCA-SEM and BEA-SEM under different scenarios. We analysed the data that were generated based on a known SEM model (designated as a true SEM model; see Section 2.2.1 and Figure S4A,B) and determined the final SEM model selected by each of the three approaches (see the Section 2.2.3). The approaches' explanatory and predictability properties were then assessed through comparisons between each selected SEM model and the true SEM model (see the Section 2.2.4). The scenarios simulated a gradient that ranged from ideal to realistic scenarios by manipulating the structure of the true or saturated SEM models, the dataset

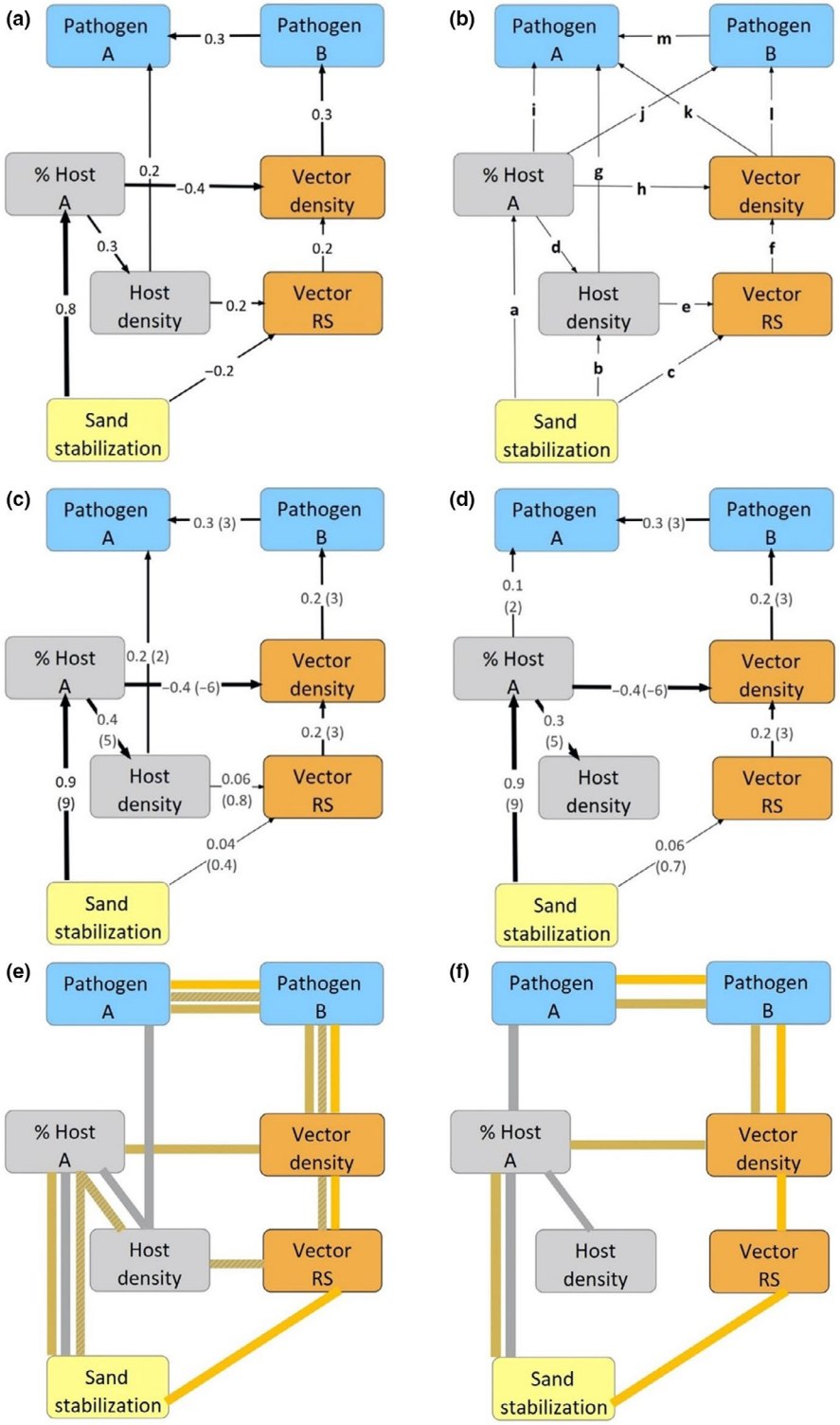

FIGURE 1 SEM* diagrams in the case study. The diagrams represent the true SEM model that was used to generate the data (a), the saturated SEM model (b) that was used in the SEM specification process (Figure 2) and was one of the competing models in the set (model 18 in Table S2), and the two 'best' SEM models (model 17 in C and model 13 in D; Table S2) selected during the model selection process (Figure 2). (e)–(f) are the same as (c)–(d) but instead of the coefficient values, the routes for possible cascading effects are highlighted. The dependent variable is the probability of any host to be infected with pathogen A, the independent variable is the level of sand stabilization (yellow), and the others are host-related (grey) and vector-related (orange) mediators, as well as the host probability to be infected with pathogen B (blue). Numbers are path coefficients (β) and β/SE in parentheses, and arrows indicate hypothesized causality, where broader arrows represent stronger effects. The grey, orange and mixed (grey and orange) colours of the possible routes highlight potential host-, vector- and host–vector cascading effects, respectively. *Note that these are all path analysis models, but for generality, we term them SEM models

size or the variance value of the dependent variable (see the Section 2.2.2). We then ran sensitivity analyses to quantify the dependence of the simulation results on (a) the true SEM model's structure and parameter values and (b) the criteria used in the MSA-SEM to specify the 'best' SEM model set (see Section 2.3). Considering the mentioned advantages of MSA-SEM, we hypothesized that it would best handle the limited information, as well as the high variability and complexity of ecological systems, and thus predicted that its performance would outweigh those of the others when the sample size is small, and the variability in the dependent variable, the network complexity and the parameter uncertainty are high.

The same simulation's rationale was used to compare the performances of the MSA-SEM approaches based on different information criteria, and to compare those differing in the criterion used for specifying the 'best' SEM model set under suboptimal conditions (see the Section 2.2.2).

2.2 | Methods

The R code used for the simulations can be found in https://doi.org/10.6084/m9.figshare.14627289.

2.2.1 | Generating the data

A random 10,000 row dataset was generated per simulation run, based on a true SEM model (Figure S4A,B, for the scenarios of full knowledge and uncertainty, respectively). Since the MSA-SEM approach relies on the development of well-thought-out a priori explanatory hypotheses, the structures of the true SEM models were inspired by the model system of the rodent communities in the northwestern Negev Desert's sands in Israel. In particular, the six independent and mediator variables represented the environment-related, host-related, vector-related and pathogen-related factors (marked in yellow, grey, orange and blue, respectively) that are expected to affect the probability of any host to be infected with pathogen A (dependent variable, marked in blue, hereafter, host probability to be infected with pathogen A; Figure S4A,B). In addition, the paths (i.e. links between variables irrespective of their signs) conformed to the theoretical foundation of a host–vector–pathogen model system (Table S1) and reflected the host–vector hypothesis (e.g. the effect of sand stabilization on the host probability to be infected with pathogen A is mediated by both host and vector mediators). However, for simplicity, we initially assumed that all paths are positive, all parameter estimates' means are zero and all variances are one.

2.2.2 | Simulating 16 scenarios

We evaluated the relative capabilities of the three approaches under scenarios representing common constraints in ecological research. Specifically, four factors—each with two levels—were manipulated in all possible combinations (2^4 unique scenarios). To simulate a researcher constraint in population sampling, we manipulated the sample size by analysing either the full generated dataset or a random subset of 100 rows in each replicate run. To simulate the high variability in the dependent variable that characterizes data collected in nature, we generated the data by using either low or high variance values (one or five units, respectively). To simulate the scenario of high network complexity that also characterizes natural environments, we manipulated the presence of extra paths in the saturated SEM model (paths k and n − r that are present in Figure S4D but not in Figure S4C). Finally, to simulate a common scenario in which a researcher is not aware of all the important parameters that underlie the observed pattern, we manipulated the presence of extra mediator variables in the true SEM model that were not used during the data analysis stage (parameter uncertainty; Figure S4B).

In the MSA-SEM refinement analyses, we compared the SEM models (a) selected as 'best' through either AIC, AICc, BIC or aBIC ranking (see Glossary), and (b) those specified as belonging to the 'best' SEM model set, based on either a summation of model weights ($\sum w_i \geq 55\%$ or $\sum w_i \geq 90\%$; see definition for w_i in the Glossary), ΔAICc ≤ 2 between each SEM model and the best SEM model (i.e. the one with the lowest information criterion values), or based on the inclusion of all SEM models with $w_i \geq 10\%$. These simulations were all done under suboptimal conditions (a small sample size and high variability in the dependent variable, complexity and uncertainty). Moreover, the comparison between the four information criterion indices was made under the $\sum w_i \geq 55\%$ criterion.

2.2.3 | Determining the final SEM model selected by each approach

In each replication, the first stage was common to all approaches and included the definition of the saturated SEM model (Figure S4C,D, for the low and high network complexity scenarios, respectively). The saturated SEM model was used as a starting point for the SCA and BEA approaches and represented one of the competing models in the 20-model set of the MSA. In the second stage, we determined the final selected SEM model, using each approach. In the SCA approach, we achieved this goal by creating a SEM model that included all the significant ($p < 0.05$) paths of the saturated SEM model. In the BEA approach, we ran a stepwise elimination procedure in which, at each step, the path with the highest associated p-value was dropped. Then, both nested models (with and without this specific path) were compared, and the one with the lower AICc was kept for the next elimination step. Through all elimination steps, we maintained the structure of the mediators and dependent variables necessary for model selection comparison. To define the final SEM model selected by the MSA, we followed the MSA-SEM analysis process (Figure 2 and case study). In addition to the saturated and true SEM models, the competing model set included 18 models. These were logical varieties of models nested in the saturated SEM model and representing either the host-mediated, vector-mediated or host–vector hypotheses, postulating that the effect of sand stabilization on the host probability to be infected with pathogen A is either mediated by host-related, vector-related or both types of factors, respectively. During the model selection process, we then ranked the 20 models using AICc and specified the best models based on the $\sum w_i \geq 55\%$ criterion (see Figure S5 as a proof of concept). The final model was created using a weighted average of parameter estimates over the 'best' SEM model set.

2.2.4 | Comparing the approaches' performances

The approaches' performances were assessed based on a comparison between each selected final SEM model and the true SEM model (Figure S4A) and was quantified by three indices. The first two indices assessed the explanatory properties of the selected SEM models, without considering their path coefficient values. Specifically,

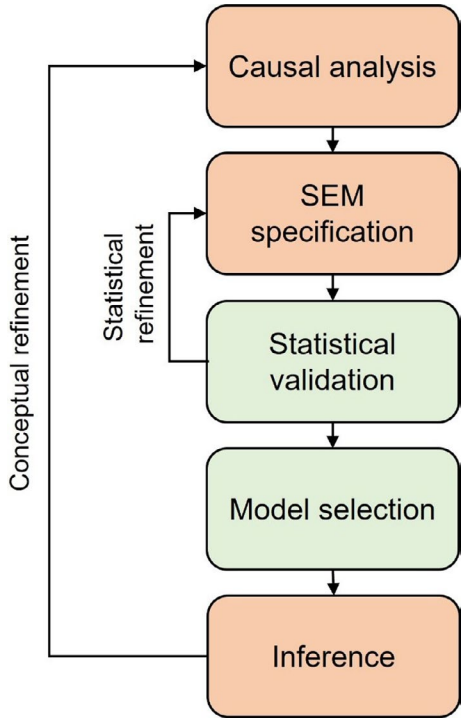

FIGURE 2 A road map for an MSA-SEM analysis process. Steps highlighted in green are those in which the candidate models are confronted with specific data

the 'true model' index quantified the probability of each approach to select a model with an identical structure to that of the true SEM model over the 500 replications per scenario, receiving, in each run, the value one when successful and zero when not. The 'true paths' index quantified the probability of each approach to include any and only true paths and was calculated as one minus the summation of the false-positive and the false-negative rates. The false-positive rate was computed as the ratio between the number of paths included in the selected SEM model and not in the true SEM model and the total extra (false) paths in the saturated SEM model. The false-negative rate was computed as one minus the ratio between the number of shared paths between the selected SEM model and the true SEM model and all paths in the true SEM model. The third index, entitled the 'predictability' index, assessed the predictability of the selected SEM model by calculating the similarity between the path coefficient values of the selected SEM model and the true SEM model. It was computed as one minus the sum of the absolute differences between the values of each path in the selected SEM model and the corresponding path in the true SEM model, divided by the number of paths in the saturated model. The coefficients of paths that were not included in the models of interest were assumed to be zero.

Preliminary analyses suggest that the performances of the final SEM models selected by BEA are never superior to the other two approaches; these latter two reverse their priority in response to the simulated scenarios (Figure S6), and thus to simplify the interpretation, we adjusted the indices to test the relative advantage of MSA over the SCA. Accordingly, the 'true model' index received the value one when only the MSA selected the true SEM model and zero for all other outcomes. The modified 'true paths' index summed the differences between the false-positive rates and the false-negative rates of two approaches. The modified version of the 'predictability' index calculated first the deviations of the path coefficients in the selected models compared to those of the true SEM model, according to each approach, and then calculated the differences between the MSA and the SCA deviations. In all cases, higher values indicated an MSA advantage.

When the MSA showed an advantage over SCA, further analyses were conducted to confirm that the MSA advantage was the result of the special MSA properties rather than just a product of the greater number of paths included by the multimodal selection procedure. Accordingly, the performance of the MSA approach was also compared to two null versions of MSA, using the above three indices. The first null version was used to test the importance of the SEM specification process (Figure 2) and included, in each run, a random selection of 20 competing SEM models (MSA0a). The second null version was used to test the importance of the model selection step (Figure 2) and included, in each run, a random selection of the 'best' SEM models from the competing model set (MSA0b). The number of SEM models allowed to be selected in this version was equal to the number of SEM models selected by the MSA-SEM of each run.

2.2.5 | Statistical analysis

For each scenario, we ran 500 replications, and in each replication, the data were independently generated based on the true SEM models (Figure S4A,B). The saturated SEM model for each run was then fitted to the generated data by the maximum likelihood (ML) method. When the model did not converge to a dataset or the saturated SEM model significantly deviated from the data, its dataset was ignored and replaced with additionally generated data until 500 datasets yielded proper solutions.

Then, we used GLMs to test the effect of the four factors, namely, the sample size, variability in the dependent variable, network complexity, and parameter uncertainty and their two-way interactions (independent variable) on the performance difference between the MSA-SEM and the SCA-SEM (each of the three performance indices, dependent variables), indicating the MSA-SEM advantage by positive values. The 'best' models explaining the approaches' performances were selected based on the relative weights of the AICc values (w_i). When an MSA-SEM advantage was revealed, we used the same approach to compare its performance with those of the two null MSA versions.

2.3 | Results

All final SEM models selected either by the three SEM approaches or by the MSA0a approach fitted the data well (the comparative fit index; CFI > 0.90), while almost half of the SEM models selected through the null version of the model selection process (MSA0b) did not.

Our analyses suggest that the relative advantage of MSA-SEM over the SCA-SEM depends on the performance index, the tested factors and the interaction between them (Table S3). The interaction between the sample size and the network complexity was the best predictor of the variability in the relative advantage of MSA-SEM considering the 'true model' index (Table S3). This is because under low network complexity, the MSA-SEM had a greater advantage at small sample sizes than at large samples sizes, whereas under high network complexity, its advantage was consistent at both sample sizes (Figure 3a). The interaction between the sample size and the parameter uncertainty was the best predictor of the variability in the relative advantage of MSA-SEM considering both the 'true paths' and the 'predictability' indices (Table S3). This is because when the sample sizes were large, the SCA-SEM had a greater advantage under high than under low parameter uncertainty, whereas at small sample sizes, the MSA-SEM had a consistent advantage, independent of the parameter uncertainty level (Figure 3b,c). In all cases of MSA-SEM superiority, it was also superior to both null MSA versions (Figure S7). Neither the level of variability in the dependent variable nor its interaction with any of the other three factors had an important effect on the comparison between the MSA-SEM and SCA-SEM performances (Table S3; Figure S6).

The inclusion of negative path coefficients in the true SEM models or the replacement of their zero parameter estimate means with the actual values observed in the field did not affect the results presented in Table S3 and Figure 3. This consistency, along with analyses showing that the variance of the dependent variable is not an important predictor of the variability in the relative advantage of MSA-SEM (Figure S6), suggests that our simulation results are not affected by the properties of the true SEM model.

We found that under suboptimal conditions, the AICc and BIC outperformed all other explored information criteria, in terms of both explanatory and predictability properties (Figure S5A,C,E). This AICc and BIC superiority was consistent under the various 'best' SEM selection criteria (data not shown). Among the criteria specifying the 'best' SEM set, we found that while the $\sum w_i \geq 55\%$ and the $\Delta AICc \leq 2$ criteria showed similar performances to the others in terms of model predictability (Figure S5F), they outperformed them in terms of the model's explanatory properties (Figure S5B,D). Consistent with these results, our second sensitivity analysis showed that using either one of these two superior criteria in the MSA-SEM increased the overall advantage of this approach over the SCA-SEM and the BEA-SEM approaches (Figure S8). However, the performances of the final SEM models selected by BEA were still never superior to the other two approaches (due to the high similarity to Figure S6, the results are not shown), and the effects of the four factors and their interactions on the relative advantage of the MSA over the SCA remained robust (Figure S8).

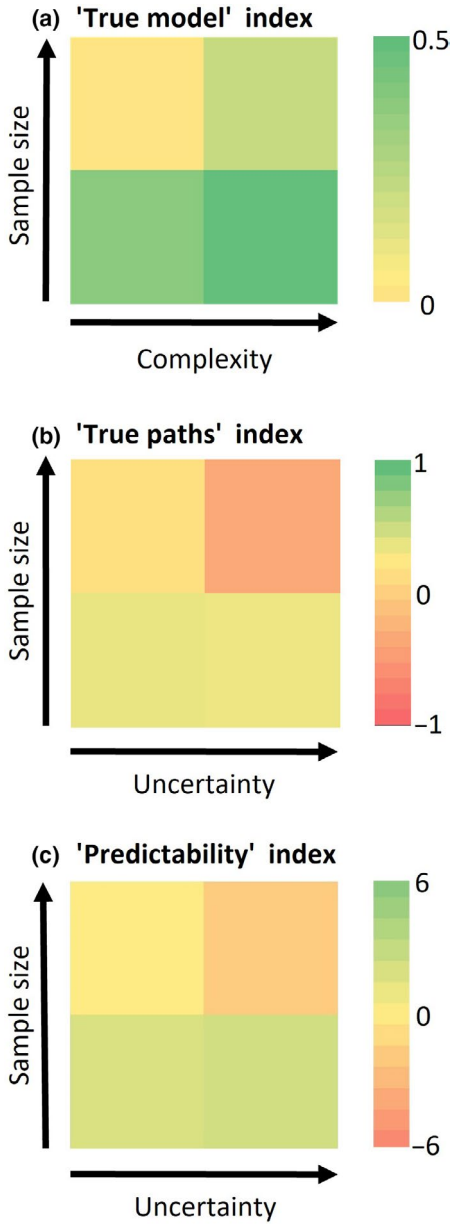

FIGURE 3 Relative advantages of the model selection approach (MSA) over the strictly confirmatory approach (SCA) to SEM. Heat maps of the probability that only MSA will select the true SEM model (Figure S4A) ('true model' index; a), and of the differences between MSA and SCA in the probability to include any and only true paths ('true paths' index; b) and in the predictability of the path coefficients ('predictability' index; c) as a function of the sample size ($N = 100$ or 10,000), network complexity (12 or 18 paths in the saturated SEM model; Figure S4C,D) and parameter uncertainty (zero or four extra mediators used to generate the data but left unspecified in the analyses; Figure S4A,B). The green hues become darker with increasing MSA-SEM advantage, and the red hues become darker with increasing SCA-SEM advantage. The heat maps focus on the most important predictors determined by the model selection analysis (Table S3)

2.4 | Conclusions

The simulation results suggest that SCA-SEM is preferable under the ideal conditions of extremely large sample sizes (~10,000), especially when there is uncertainty regarding the relevant variables, probably owing to its lower tendency to produce false-positive errors (i.e. including an unimportant path in the final SEM model; Figure S6E–H). In contrast, MSA-SEM is seen as a superior approach when addressing

the study constraints faced by ecologists: limited sample population (~100 subjects), complex networks (~18 paths in the saturated SEM model) or both. Its superiority over the two null MSA versions further suggests that this advantage can be attributed to the MSA-SEM analysis processes (Figure 2), which may decrease the likelihood of incurring false-negative errors (i.e. omitting an important path from the selected SEM model; Figure S6I–L). Under these suboptimal scenarios, the use of AICc or BIC information criteria and the specification of the 'best' SEM set, based on the $\sum w_i \geq 55\%$ or $\Delta\text{AICc} \leq 2$ criteria, achieved the best performance, probably since they select fewer models, lowering the tendency to produce false-positive errors.

3 | AN MSA-SEM ROAD MAP

To encourage consistent and correct use of MSA-SEM, we propose a five-step road map (Figure 2). The road map highlights unique MSA-SEM considerations that will be important to most analyses. Informed by Grace and Irvine (2020), it begins with a causal analysis, which leads to the development of hypothetical explanatory models for evaluation. Considering data availability and refined by statistical validation, a set of competing models encoding the background knowledge are then specified, through the process of SEM specification. The third process of statistical validation sets the conditions for the model selection process by determining the SEM models' relative fit to the data, the estimation method and the information criterion. During the fourth process of model selection, the likelihood of all SEM models is quantified, and the 'best' SEM models are specified and characterized. Based on this information, during the process of inference, ecologists interpret their findings, considering the study hypotheses, and refine their conceptual view to simulate further studies. Note that only the third and fourth processes involve confronting the SEM models with the data. Below, we discuss each process in turn, and in the Supporting Information, we demonstrate them via a case study, supported by a tutorial R code (https://doi.org/10.6084/m9.figshare.14627343).

3.1 | Causal analysis

SEM requires prior knowledge of plausible causal pathways; there is not enough information in the correlations of field observations, absent a priori hypotheses, to infer the correct causal pathways (there are too many possible bivariate associations). Prior explanatory hypotheses are required for SEM to be valid for causal inference (Box 1 and Grace & Irvine, 2020). The assembly of prior knowledge includes compiling a list of biological assumptions and predictions relevant to all possible relationships between the measured variables (Table S1).

3.2 | SEM specification

Based on the study questions and the data's nature and availability, the general structure of candidate SEM models should be determined. Today, it is possible to integrate into SEM unordinary data types, representing basic ecosystem processes and dynamics, including hierarchical linear modelling, time series, repeated measure analyses and feedback loops (Bentler & Bonett, 1980; Newsom, 2015). In addition, meta-parameter partitioning is now made possible through multigroup methods (e.g. Flatau et al., 2018).

The simplification of all candidate SEM models is required to set the space of possible hypotheses, avoiding overfitting and spurious findings due to unnecessary complexity (Box 1). Limiting the number of parameters is recommended, allowing at least five observations per free parameter (Bentler & Chou, 1987; Kline, 2016). Some paths may be removed due to violations of conditional independence (Kline, 2016; Shipley, 2016). To reduce the parameter count, one may remove those that produce multicollinearity (see Glossary), formulate an equation that incorporates multiple dependent variables (e.g. the integrated index of female flea reproductive success in the case study) or posit unobservable variables.

Candidate SEM models can be generated via a combination of knowledge of overall network structure, as employed by Mitchell (1992), Cayenne-Engel and Irwin (2003), Cardon et al. (2011), Lootvoet et al. (2013), and von Hardenberg and Gonzalez-Voyer (2013), and/or inclusion and exclusion of individual hypothesized relationships (e.g. Johnson, 2002; Palomares et al., 1998).

Regardless of how candidates are generated, understanding whether models are comparable, which relies on their being defined over shared random variables, is crucial; confusion about this matter is a common pitfall leading to incorrect interpretations among MSA-SEM users. SEM software programs use various calibration formulations that may affect which models are comparable (Lee, 2007; Narayanan, 2012). For example, in some formulations, the ensembles of independent, mediator and dependent variables are shared between all candidates, but the assignment of variables to each type may vary between candidates. In other cases, only the mediator and dependent variables are required to be shared between candidates, whereas the independent variables may be unshared. Thus, based on one's SEM calibration software, it is important to determine in advance which models are comparable.

It is not possible to recommend a 'correct' number of candidate models independent of a specific ecological problem. However, it is generally recommended that their number should be small, relative to the size of the dataset, since an excessive number of models in the competing set is expected to increase the risk of including spurious relationships and to reduce the ability to distinguish one or a few 'best' models (Burnham & Anderson, 2002).

3.3 | Statistical validation

The candidate SEM models are selected by confronting them with data, which necessitates (a) a test of model fit to the data and (b) the selection of an estimation method and the (c) information criterion. In most GLM analyses, it is recommended to take into consideration various statistical constraints, such as missing data, outliers,

multicollinearity, non-normal distribution (or in the case of SEM, no multivariate normality) and restricted sample size. These constraints are more common in SEM due to its multidimensional nature (e.g. violations of multicollinearity are more common in these models since they include both independent and mediator variables that can potentially correlate) and are crucial to consider in the MSA-SEM approach since these statistical properties determine the methods used for the above three tasks (Fan et al., 2016; Hu & Bentler, 1999). In many cases, fitting the SEM models to non-normal residual distributions or choosing a proper estimator would resolve these constraints (Hoyle, 2012; Kline, 2016). Alternative solutions are addressed in previous SEM reviews (Fan et al., 2016; Farrell et al., 2007; Hoyle, 2012; Kline, 2016). In cases of high overdispersion, the inference can be improved by applying Quasi-AIC as the information theory index (Johnson & Omland, 2004).

3.4 | Model selection

The information criterion is used to rank the competing models and to weigh the relative support for each one (Johnson & Omland, 2004). The two main information criteria classes are those based on AIC and BIC information criteria. Despite major theoretical differences, both apply to models that are represented by multivariate probability distributions and aim to minimize the distance between the unknown true likelihood function of the data and the fitted likelihood function of the model. Their main difference is that the BIC penalizes model complexity to a greater extent (Altman & Bland, 1994; Dziak et al., 2020). In our simulation study, which featured a suboptimal scenario, we found that the models ranked based on AICc and BIC showed the best performance in terms of both explanatory and predictability properties (Figure S5). Accordingly, we recommend using them under similar conditions.

The likelihood of each model is quantified by its relative weight w_i. Often there is no single model overwhelmingly supported by the data ($w_i < 90\%$), and thus, the inference is based on a group of the 'best' models. Various rules of thumb were offered to specify the set of the 'best' SEM models, considering w_i (models with $w_i \geq 10\%$ are regarded as 'best'; Cohen et al., 2015; Messika et al., 2017), $\sum w_i$ (e.g. $\sum w_i \geq 90\%$; Kedem et al., 2014) or ΔAIC between each of the SEM models and the best SEM model (e.g. models with ΔAIC ≤ 2 are regarded as best; Johnson, 2002). Based on our computational exploration, we recommend adopting the $\sum w_i \geq 55\%$ criterion to specify the 'best' SEM model set under suboptimal conditions (Figure S5).

3.5 | Inference

Once the 'best' models emerge from the previous process, an inference about the study predictions and hypotheses can take place, considering the model structure and absence/presence of specific paths. It may be quantitatively valuable to compare the likelihood of each by means of w_i, the evidence ratio (i.e. the ratio of model weights, ER), or ML values (Burnham et al., 2011). To assess the strength and importance of specific paths of interest, researchers often examine whether their coefficients are significantly different from zero. However, since it is recommended not to mix dichotomous hypothesis testing with MSA (Burnham et al., 2011), we propose to use here any MSA as well (e.g. comparison of models including or excluding a given path or calculation of the relative importance of specific paths by model averaging).

In parallel and regardless of the method of quantifying the path importance, we recommend reporting standardized or unstandardized path coefficients, respectively, depending on whether one is evaluating paths or predicting relationships (Kline, 2016; Shipley, 2016). These could be separately done for each 'best' SEM model or for a weighted average SEM model of the 'best' models (Johnson & Omland, 2004; Palomares et al., 1998).

Finally, the influence of one variable on another can be decomposed into direct, indirect (calculated as the product of the effect of the independent variable on the mediator(s) and of the mediator(s) on the dependent variable), and total (the sum of the direct and indirect effects) effects (Hoyle, 2012; Kline, 2016). The decomposition into direct and indirect effects may be especially important in quantifying cascading effects (Byrnes et al., 2011). Altogether, this information on the models' likelihood and the relative importance of specific direct and indirect effects should then be used for hypothesis validation.

This approach can be embedded in a broader weight-of-evidence framework, in which other types of evidence, such as modification indices, residual relationships and d-separation tests, can be reported (Grace, 2020). Regardless, the results should be interpreted with caution, depending on the researcher's confidence that the paths indeed represent directional cause–effect relationships, and that parameter estimates provide satisfying approximations of true values and represent causal effects (see case study).

Finally, and most importantly, we recommend exploiting the best SEM models for feeding back into the ***causal analysis*** of new research avenues. Here it is possible to use the exploratory mode of SEM to generate new hypotheses that can be further tested through experimentation or against new data. This could be done by comparing models in the absence or presence of specific paths of interest and by comparing alternative indirect effects. Provided that we constrain the use of this process for raising new questions and ideas, these *post-hoc* 'games' can be done even with no solid background or theoretical foundations.

4 | SUMMARY AND CONCLUSION

This paper encompasses the pros and cons of MSA-SEM relative to other SEM approaches, highlights its relevance to the study of complex communities and provides a unified road map for its proper uses and boundaries. It emphasizes the important consideration of both practical and philosophical principles to realize the unique features of this combined approach. We predict that our unified approach, made accessible to ecologists, will guide the investigation of questions that were not previously tractable in this field.

ACKNOWLEDGEMENTS

This work was supported by the Israel Science Foundation (# 1391/15 to HH). M.G. was sponsored by the Kreitman School of Advanced Graduate Studies (BGU) and the Blaustein Center for Scientific Cooperation (Jacob Blaustein Institutes for Desert Research, BGU).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

H.H. conceived the idea and jointly, with M.G., compiled the literature review and contributed to the conception, design, statistical analysis and interpretation of data. They also wrote the first draft of the manuscript; R.Y. wrote the R code and contributed to the design of the simulation study; S.K.H. contributed to the conceptual parts of the manuscript, including the road map, and contributed substantially to the revisions.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13742.

DATA AVAILABILITY STATEMENT

All data and codes for the simulation and case studies were deposited in an open repository https://doi.org/10.6084/m9.figshare.14627289 and https://doi.org/10.6084/m9.figshare.14627343, respectively.

ORCID

Mario Garrido https://orcid.org/0000-0001-7899-8913
Scott K. Hansen https://orcid.org/0000-0001-8022-0123
Rami Yaari https://orcid.org/0000-0002-8808-8937
Hadas Hawlena https://orcid.org/0000-0002-0634-2920

REFERENCES

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, *308*, 1552.

Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. *Nature*, *567*, 305–307.

Asparouhov, T., & Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

Barringer, S. N., Eliason, S. R., & Leahey, E. (2013). A history of causal analysis in the social sciences. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research*. Springer.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*, 78–117. https://doi.org/10.1177/0049124187016001004

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley & Sons.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*, 23–35. https://doi.org/10.1007/s00265-010-1029-6

Byrnes, J. E., Reed, D. C., Cardinale, B. J., Cavanaugh, K. C., Holbrook, S. J., & Schmitt, R. J. (2011). Climate-driven increases in storm frequency simplify kelp forest food webs. *Global Change Biology*, *17*, 2513–2524. https://doi.org/10.1111/j.1365-2486.2011.02409.x

Cardon, M., Loot, G., Grenouillet, G., & Blanchet, S. (2011). Host characteristics and environmental factors differentially drive the burden and pathogenicity of an ectoparasite: A multilevel causal analysis. *Journal of Animal Ecology*, *80*, 657–667. https://doi.org/10.1111/j.1365-2656.2011.01804.x

Cayenne-Engel, E., & Irwin, R. E. (2003). Linking pollinator visitation rate and pollen receipt. *American Journal of Botany*, *90*, 1612–1618. https://doi.org/10.3732/ajb.90.11.1612

Chen, D., Lan, Z., Bai, X., Grace, J. B., & Bai, Y. (2013). Evidence that acidification-induced declines in plant diversity and productivity are mediated by changes in below-ground communities and soil properties in a semi-arid steppe. *Journal of Ecology*, *101*, 1322–1334. https://doi.org/10.1111/1365-2745.12119

Cohen, C., Einav, M., & Hawlena, H. (2015). Path analyses of cross-sectional and longitudinal data suggest that variability in natural communities of blood-associated parasites is derived from host characteristics and not interspecific interactions. *Parasites & Vectors*, *8*, 429. https://doi.org/10.1186/s13071-015-1029-5

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, *39*, 297–316. https://doi.org/10.25300/MISQ/2015/39.2.02

Duffy, J. E., Reynolds, P. L., Boström, C., Coyer, J. A., Cusson, M., Donadi, S., Douglass, J. G., Eklöf, J. S., Engelen, A. H., Eriksson, B. K., Fredriksen, S., Gamfeldt, L., Gustafsson, C., Hoarau, G., Hori, M., Hovel, K., Iken, K., Lefcheck, J. S., Moksnes, P.-O., … Stachowicz, J. J. (2015). Biodiversity mediates top-down control in eelgrass ecosystems: A global comparative-experimental approach. *Ecology Letters*, *18*, 696–705. https://doi.org/10.1111/ele.12448

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2013). *An introduction to latent variable growth curve modeling: Concepts, issues, and application*. Routledge.

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*, 553–565. https://doi.org/10.1093/bib/bbz016

Eldridge, D. J., Delgado-Baquerizo, M., Travers, S. K., Val, J., Oliver, I., Hamonts, K., & Singh, B. K. (2017). Competition drives the response of soil microbial diversity to increased grazing by vertebrate herbivores. *Ecology*, *98*, 1922–1931. https://doi.org/10.1002/ecy.1879

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes*, *5*, 19–30. https://doi.org/10.1186/s13717-016-0063-3

Farrell, P. J., Salibian-Barrera, M., & Naczk, K. (2007). On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation*, *77*, 1065–1080. https://doi.org/10.1080/10629360600878449

Flatau, R., Segoli, M., Khokhlova, I., & Hawlena, H. (2018). *Wolbachia*'s role in mediating its flea's reproductive success differs according to flea origin. *FEMS Microbiology Ecology*, *94*, fiy157. https://doi.org/10.1093/femsec/fiy157

Grace, J. B. (2020). A 'Weight of Evidence' approach to evaluating structural equation models. *One Ecosystem*, *5*, e50452. https://doi.org/10.3897/oneeco.5.e50452

Grace, J. B., & Irvine, K. M. (2020). Scientist's guide to developing explanatory statistical models using causal analysis principles. *Ecology*, *101*(4), e02962. https://doi.org/10.1002/ecy.2962

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis. Techniques and applications*. Routledge.

Hoyle, R. H. (2012). *Handbook of structural equation modeling*. Guilford Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.

Structural Equation Modeling: A Multidisciplinary Journal, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Johnson, J. B. (2002). Divergent life histories among populations of the fish *Brachyrhaphis rhabdophora*: Detecting putative agents of selection by candidate model analysis. *Oikos*, *96*, 82–91.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*, 101–108. https://doi.org/10.1016/j.tree.2003.10.013

Kedem, H., Cohen, C., Messika, I., Einav, M., Pilosof, S., & Hawlena, H. (2014). Multiple effects of host-species diversity on coexisting host-specific and host-opportunistic microbes. *Ecology*, *95*, 1173–1183. https://doi.org/10.1890/13-0678.1

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Wiley & Sons.

Lefcheck, J. S. (2016). PIECEWISESEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*, 573–579.

Lin, L., Huang, P., & Weng, L. (2017). Selecting path models in SEM: A comparison of model selection criteria. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 855–869. https://doi.org/10.1080/10705511.2017.1363652

Liu, X., Swenson, N. G., Lin, D., Mi, X., Umaña, M. N., Schmid, B., & Ma, K. (2016). Linking individual-level functional traits to tree growth in a subtropical forest. *Ecology*, *97*, 2396–2405. https://doi.org/10.1002/ecy.1445

Loehlin, J. C., & Beaujean, A. A. (2016). *Latent variable models: An introduction to factor, path, and structural equation analysis*. Taylor & Francis.

Lootvoet, A., Blanchet, S., Gevrey, M., Buisson, L., Tudesque, L., Loot, G., & Mayhew, P. (2013). Patterns and processes of alternative host use in a generalist parasite: Insights from a natural host-parasite interaction. *Functional Ecology*, *27*, 1403–1414. https://doi.org/10.1111/1365-2435.12140

Messika, I., Garrido, M., Kedem, H., China, V., Gavish, Y., Dong, Q. F., Fuqua, C., Clay, K., & Hawlena, H. (2017). From endosymbionts to host communities: Factors determining the reproductive success of arthropod vectors. *Oecologia*, *184*, 859–871. https://doi.org/10.1007/s00442-017-3906-4

Mitchell, R. J. (1992). Testing evolutionary and ecological hypotheses using path analysis and structural equation modelling. *Functional Ecology*, *6*, 123–129. https://doi.org/10.2307/2389745

Mundry, R., & Nunn, C. L. (2009). Stepwise model fitting and statistical inference: Turning noise into signal pollution. *The American Naturalist*, *173*, 119–123. https://doi.org/10.1086/593303

Narayanan, A. (2012). A review of eight software packages for structural equation modeling. *The American Statistician*, *66*, 129–138. https://doi.org/10.1080/00031305.2012.708641

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*, 557–595. https://doi.org/10.1207/S15328007SEM0704_3

Palomares, F., Ferraras, P., Travaini, A., & Delibes, M. (1998). Coexistence between Iberian lynx and Egyptian mongooses: Estimating interaction strength by structural equation modelling and testing by an observational study. *Journal of Animal Ecology*, *67*, 967–978. https://doi.org/10.1046/j.1365-2656.1998.6760967.x

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. *Handbook of Market Research*, *26*, 1–40.

Schuberth, F., Henseler, J., & Dijkstra, T. K. (2018). Confirmatory composite analysis. *Frontiers in Psychology*, *9*, 2541. https://doi.org/10.3389/fpsyg.2018.02541

Shipley, B. (2016). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference with R* (2nd ed.). Cambridge University Press.

Tarka, P. (2018). An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity*, *52*, 313–354. https://doi.org/10.1007/s11135-017-0469-8

Vinzi, V. E., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares*. Springer-Verlag.

von Hardenberg, A., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution*, *67*, 378–387. https://doi.org/10.1111/j.1558-5646.2012.01790.x

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, *75*, 1182–1189. https://doi.org/10.1111/j.1365-2656.2006.01141.x

Wu, H., Fai Cheung, S., & On Leung, S. (2020). Simple use of BIC to assess model selection uncertainty: An illustration using mediation and moderation models. *Multivariate Behavioral Research*, *55*, 1–16. https://doi.org/10.1080/00273171.2019.1574546

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Garrido, M., Hansen, S. K., Yaari, R., & Hawlena, H. (2022). A model selection approach to structural equation modelling: A critical evaluation and a road map for ecologists. *Methods in Ecology and Evolution*, 13, 42–53. https://doi.org/10.1111/2041-210X.13742