



WAGENINGEN  
UNIVERSITY & RESEARCH

## Generalized AIC and chi-squared statistics for path models consistent with directed acyclic graphs

Ecology

Shipley, Bill; Douma, Jacob C.

<https://doi.org/10.1002/ecy.2960>

This publication is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this publication please contact [openscience.library@wur.nl](mailto:openscience.library@wur.nl)

## Generalized AIC and chi-squared statistics for path models consistent with directed acyclic graphs

BILL SHIPLEY <sup>1,3</sup> AND JACOB C. DOUMA<sup>2</sup>

<sup>1</sup>*Département de biologie, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1 Canada*

<sup>2</sup>*Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands*

*Citation:* Shipley, B., and J. C., Douma. 2020. Generalized AIC and chi-squared statistics for path models consistent with directed acyclic graphs. *Ecology* 101(3):e02960. 10.1002/ecy.2960

**Abstract.** We explain how to obtain a generalized maximum-likelihood chi-square statistic,  $X_{ML}^2$ , and a full-model Akaike Information Criterion (AIC) statistic for piecewise structural equation modeling (SEM); that is, structural equations without latent variables whose causal topology can be represented as a directed acyclic graph (DAG). The full piecewise SEM is decomposed into submodels as a Markov network, each of which can have different distributional assumptions or functional links and that can be modeled by any method that produces maximum-likelihood parameter estimates. The generalized  $X_{ML}^2$  is a function of the difference in the maximum likelihoods of the model and its saturated equivalent and the full-model AIC is calculated by summing the AIC statistics of each of the submodels.

**Key words:** Akaike Information Criterion; directed acyclic graph; *d*-separation; maximum likelihood; model selection; path analysis; piecewise SEM.

### INTRODUCTION

Structural equations modeling (SEM) is a multivariate statistical method that allows one to test hypotheses concerning the causal structure believed to generate the observed data. Modern SEM was developed in disciplines in which the variables of theoretical interest are primarily latent (i.e., not directly measured). By making certain assumptions concerning the data (multivariate normality, mutually independent observations, linear relationships between the variables), and by maximizing the likelihood of the data given these assumptions, one obtains an inferential test of the fit between the data and the hypothesized SEM via the maximum-likelihood chi-square statistic (Bollen 1989) and an Akaike Information statistic (AIC). Much effort has gone into relaxing the assumption of multivariate normality but available statistical SEM programs still largely require linearity of relationships, although nonlinearity due to binary or ordered categorical variables can be modeled via a link function. Furthermore, although some forms of nesting can be accommodated (e.g., Shipley 2016:Chapter 7), it

is difficult, and sometimes impossible, to accommodate more complicated forms of nonindependence of observations because of nesting, temporal, spatial or phylogenetic correlations, or cross-classification.

Some theoretical variables in ecology cannot be directly measured, or else contain substantial measurement error; in such cases, one should model such variables as latent. Examples of latent variables are soil “fertility” (Daou and Shipley 2019) or components of metal bioavailability in toxicology (Beaumelle et al. 2016). However, variables of theoretical interest in ecology can often be directly measured without substantial error. In such cases, the causal hypothesis can be represented as a directed acyclic graph (DAG) without latents, that is, a path model. Unfortunately, variables in ecological models are often linked to other variables through nonlinear functions that cannot be linearized through transformation and that follow nonnormal distributions. Furthermore, nonindependence of observations because of nesting, temporal, phylogenetic or spatial correlations, or cross-classification, are common. Shipley (Shipley 2000, Kang and Shipley 2009, Shipley 2009) developed a general inferential test (a “dsep test”) for DAGs without latent variables that is based on Pearl’s (Verma and Pearl 1988, Pearl 2009) *d*-separation claims of a DAG; a dsep test can easily include different distributional assumptions, nonindependence of

Manuscript received 20 September 2019; revised 8 November 2019; accepted 26 November 2019. Corresponding Editor: Nicholas J. Gotelli.

<sup>3</sup> E-mail: bill.shipley@usherbrooke.ca

observations, and nonlinear functional links. This allows one to test the causal implications of such path models (i.e., the topology of the DAG) more generally than with classical maximum-likelihood SEM. In classical maximum-likelihood SEM, parameter estimation and model testing are done simultaneously. Model testing via the dsep test does not estimate the free parameters of the path model; instead, parameter estimation is only done after the topology of the causal model (the DAG) has been judged consistent with the data by decomposing the full path model into a set of submodels (thus, “piecewise”) consisting of each endogenous variable and its causal parents. The parameter estimates can be based on different probability distributions and the direct links between variables can have different functional forms. A version of the dsep test whose conditional independence claims can be tested via linear, generalized linear, or linear mixed models has been implemented in the piecewiseSEM R package (Lefcheck 2016).

Although the dsep test, in combination with a significance level, allows one to decide if the causal structure of a path model should be rejected given the observed data, it does not detect errors other than in the causal topology. It is possible for models having different causal topologies to be consistent with the data based on the dsep test. When this occurs, we require a method of comparing these different models. Such a method, based on Akaike’s (1973) Information Criterion (AIC), was developed by Shipley (2013); we will call this a “dsep AIC” to distinguish it from the new AIC statistic developed in this paper. The dsep AIC statistic is based solely on the causal topology of a DAG, as captured by its d-separation claims and the resulting C-statistic, and is appropriate when comparing path models that differ in their causal topology. When path models differ in aspects besides their causal topology, Shipley’s (2013) dsep AIC statistic does not reflect such differences. Here, we generalize the maximum-likelihood chi-square statistic to allow for the complications described above and then develop a new AIC statistic for piecewise SEM. These two statistics take into account the topology of the DAG, the parameter estimates of the path model, and the distributional assumptions of the submodels of the path model. We will call this new AIC statistic a “full-model AIC.”

#### DAGS AND LOCAL MARKOV DECOMPOSITION

A DAG (Pearl 2009) represents a multivariate causal hypothesis without feedback or free covariances and links variables with their direct causes by a directed relationship (an arrow). An arrow from a cause ( $X$ ) to its direct effect ( $Y$ , i.e.,  $X \rightarrow Y$ ) means that a change in  $X$  will induce a change in  $Y$  even when all other variables in the model are held constant. A piecewise SEM consists of (1) a set of variables, (2) a DAG representing the topology of the multivariate causal hypothesis specifying the links between these variables, and (3) a statistical

model specifying the function (including its parameters) linking each effect to its direct causes. This is also called a Bayesian network (or belief network) in the field of artificial intelligence (Pearl 1988). For instance, the DAG in Fig. 1a is translated into a piecewise model  $M$  involving five submodels (Eq. 1) as follows:

$$\begin{aligned} x_1 &= p_1(x_1|\theta_1) \\ x_2 &= f_2(x_1, \Theta_2) + p_2(x_2|\theta_2) \\ x_3 &= f_3(x_2, \Theta_3) + p_3(x_3|\theta_3) \\ x_4 &= f_4(x_2, \Theta_4) + p_4(x_4|\theta_4) \\ x_5 &= f_5(x_3, x_4, \Theta_5) + p_5(x_5|\theta_5) \end{aligned} \quad (1)$$

here,  $p_i(x_i|\theta_i)$  is a probability distribution function for variable  $x_i$  given a vector of parameters  $\theta_i$  (for instance, theoretical means or variances) and  $f_i(\mathbf{pa}_i, \Theta_i)$  is a function describing the expected value of  $x_i$  given its direct causes (“parents”  $\mathbf{pa}_i$ ) having a vector of parameters  $\Theta_i$  (for example, slopes or intercepts);  $\mathbf{pa}_i$  is an empty set if  $x_i$  has no explicit causes (an exogenous variable) like  $x_1$  in Fig. 1. For notational simplicity, we will simply write  $\theta_i$  to include both the parameters of the functional link ( $\Theta_i$ ) if the variable is endogenous and of its probability density function ( $\theta_i$ ). The full set of parameters are either “fixed” (their values are specified by hypothesis) or “free” (their values are chosen based on the data by maximizing the likelihood of the model). A model  $M$  given a DAG with  $v$  variables will have  $v$  such submodels. A key property of such models is that the joint probability distribution over the variables can be decomposed through local Markov decomposition into the product of  $v$  local univariate probability distributions (Pearl 2000):

$$p_M(x_1, \dots, x_v) = \prod_{i=1}^v p_i(x_i|\mathbf{pa}_i) \quad (2)$$

#### Decomposition of the likelihood of a piecewise SEM

Our piecewise model  $M$  contains a vector of parameters  $\theta = \{\theta_1, \dots, \theta_p\}$ . This model gives the multivariate

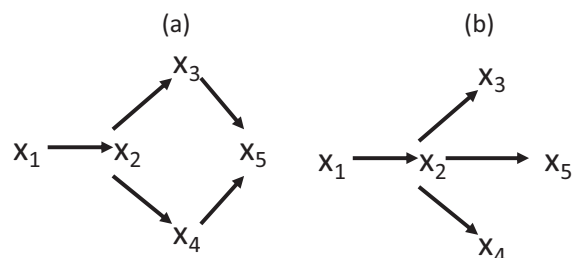


FIG. 1. Two alternative directed acyclic graphs (DAGs) representing the causal topology of a multivariate probability distribution over five variables ( $x_1$ – $x_5$ ).

probability distribution, or probability density function,  $p_M(\mathbf{X}|\boldsymbol{\theta})$ , of this set of random variables  $\mathbf{X} = \{x_1, \dots, x_v\}$ . The likelihood of this model for  $\boldsymbol{\theta}$  given the data  $(\mathbf{X})$  is  $\mathcal{L}_M(\boldsymbol{\theta}|\mathbf{X})$ , where  $\mathcal{L}_M(\boldsymbol{\theta}|\mathbf{X}) = p_M(\mathbf{X}|\boldsymbol{\theta})$ . If the values of the parameters  $\boldsymbol{\theta}$  are chosen to maximize this likelihood, then this gives the maximum likelihood  $\mathcal{L}_M(\hat{\boldsymbol{\theta}}|\mathbf{X})$ . Given Eq. 2, the log likelihood of the model  $M$  is equal to the sum of the univariate log likelihoods of each of the submodels of  $M$  (Eq. 3); in this equation the notation  $\boldsymbol{\theta}_i$  and  $\mathbf{X}_i$  refer to all of the parameters and variables involved in the  $i$ th submodel:

$$\begin{aligned} \log(\mathcal{L}_M(\boldsymbol{\theta}_M|\mathbf{X})) &= \log\left(\prod_{i=1}^v \mathcal{L}_i(\boldsymbol{\theta}_i|\mathbf{X}_i)\right) \\ &= \sum_{i=1}^v \log(\mathcal{L}_i(\boldsymbol{\theta}_i|\mathbf{X}_i)) \end{aligned} \tag{3}$$

#### GENERALIZING THE MAXIMUM-LIKELIHOOD CHI-SQUARE STATISTIC

The well-known maximum-likelihood chi-square statistic ( $X_{ML}^2$ ) of SEM (Bollen 1989) compares the observed and model-fitted covariance matrices and tests the null hypothesis that these two matrices are equal except for random sampling variation. This statistic assumes a multivariate normal distribution. However, this statistic is a special case of a more general formulation (Bollen 1989, Haughton et al. 1997) which is the well-known likelihood ratio statistic between two nested models (Eq. 4). One model ( $M_1$ ) is nested within another ( $M_2$ ) if the free parameters of the nested (more constrained) model  $M_1$  are a subset of the free parameters of the less constrained model  $M_2$  into which it is nested. When testing the hypothesized causal topology of a model, the “missing” arrows in  $M_1$  are translated into path coefficients fixed to zero, and the less constrained (“saturated”) model  $M_2$  will have these path coefficients as free parameters. Appendix S1 explains in more detail how to construct saturated models. If the additional free parameters of  $M_2$  really are equal to zero in the statistical population (i.e., their fixed values in  $M_1$ ) then  $X_{ML}^2$  will asymptotically follow a chi-squared distribution with degrees of freedom equal to the difference ( $f_2 - f_1$ ) in the number of free parameters between the less constrained  $M_2$  and the more constrained nested model  $M_1$  (Wilks 1938). Note that the minimal sample size required for obtaining a good approximation to a chi-squared distribution depends on the actual distributional forms used in the models and does not apply if some of the free parameters are estimated at their theoretical boundaries. If  $M_2$  is a fully saturated model (i.e., there are no “missing” arrows), the endogenous variables of which follow a multivariate normal distribution, then Eq. 4 reduces to the maximum-likelihood chi-square statistic of SEM (Haughton et al. 1997). Given

Eq. 3, we can obtain the maximum likelihoods of both models  $M_1$  and  $M_2$  in Eq. 4 without assuming linearity or normality, as long as we can obtain the univariate maximum likelihoods of each of the submodels. Standard solutions already exist to calculate univariate maximum likelihoods of linear or nonlinear regressions, generalized linear models, mixed models, or generalized additive models:

$$X_{ML}^2 = -2(\log(\mathcal{L}(M_1)) - \log(\mathcal{L}(M_2))) \sim \lim_{n \rightarrow \infty} \chi_{(f_2 - f_1)}^2 \tag{4}$$

#### OBTAINING A GENERALIZED AIC STATISTIC

Akaike’s Information Criterion is given in Equation 5 by Akaike (1973), where  $\mathcal{L}_M$  is the likelihood function of the model,  $\hat{\boldsymbol{\theta}}$  is a vector of free parameters (slopes, intercepts, variances, etc.) estimated at their maximum likelihoods given the data  $(\mathbf{X})$ , and  $K$  is the number of free parameters in  $\hat{\boldsymbol{\theta}}$ . The AIC statistic, as applied to classic maximum likelihood SEM, is given in Bollen (1989) and Haughton et al. (1997) and is routinely given in common SEM programs including the popular lavaan package in R (Rossee 2012). This classic maximum-likelihood AIC statistic takes into account both the causal topology of the model and the parameter estimates, but assumes multivariate normality, linearity of relationships, and complete independence of observations. The classic maximum-likelihood AIC statistic cannot be used when these assumptions are violated, which renders it useless for the path models when nonnormal distributions, nonlinear relationships, or partial nonindependence because of nesting or other complications exist.

$$AIC = -2 \log(\mathcal{L}_M(\hat{\boldsymbol{\theta}}|\mathbf{X})) + 2K \tag{5}$$

However, given Eq. 3, the AIC statistic for a piecewise model  $M$  is simply the sum of the AIC statistics obtained from each of the  $i$  submodels (Eq. 6). Note that the AIC takes the sum of the likelihood of each of submodel and not the difference in likelihood between the hypothesized and saturated model. The second-order bias correction for small samples ( $AIC_M^*$ , Eq. 7) can be used as well (Burnham and Anderson 2002):

$$AIC_M = -2 \sum_{i=1}^v \log(\mathcal{L}_i(\boldsymbol{\theta}_i|\mathbf{X}_i)) + 2 \left( \sum_{i=1}^v K_i \right) = \sum_{i=1}^v AIC_i \tag{6}$$

$$\begin{aligned} AIC_M^* &= -2 \sum_{i=1}^v \log(\mathcal{L}_i(\boldsymbol{\theta}_i|\mathbf{X}_i)) + 2 \left( \sum_{i=1}^v K_i \right) \\ &\quad + 2 \sum_{i=1}^v \frac{K_i(K_i + 1)}{n - K_i - 1} = \sum_{i=1}^v AIC_i^* \end{aligned} \tag{7}$$

NUMERICAL EXAMPLES

All analyses and simulations were done in the R statistical language (R\_Development Core\_Team 2016). The associated R code is given in DataS1: R code. We generate 100 mutually independent observations following the DAG in Fig. 1a, using different distributional assumptions. For comparison with classical SEM, we first generate these observations using linear relationships and normal distributions (Table 1). We fit the classical maximum-likelihood SEM with intercepts using the `sem()` function of the `lavaan` library of R (Rosseel 2012) to be comparable with the piecewise SEM. The four submodels were fit using the `glm()` function of R, specifying a normal error distribution. The maximum-likelihood and AIC statistics were obtained via the `logLik()` and `AIC()` functions. Table 1 gives the results. All of the resulting values are identical to the values as output using the `sem()` function of `lavaan`. If we instead fit the incorrect model that follows the DAG in Fig. 1b, the results are again identical to values as output using the `sem()` function.

Now, consider a more complicated model, whose topology is given in Fig. 1a, but having a mixture of chi-squared, Poisson, normal, and binomial distributed variables (Table 2). Furthermore, the relationship between  $x_2$  and  $x_1$  is nonlinear and cannot be linearized by transformation. It is impossible to fit such a model in classical SEM, but it is quite easy to do this using piecewise SEM. We model this nonlinear relationship with a generalized additive model via the `gam()` function with the error structure modeled as a normal distribution, but one could replace this with a nonlinear regression if desired. The remaining submodels are fit via the `glm()` function in R with appropriate distributional families and link functions. Table 2 summarizes the results. In particular, the resulting  $X^2_{ML}$  is 3.35 (5 df) giving a null probability of 0.65. Using the dsep test, we get a C-statistic of 7.68 (10 df), giving a null probability of 0.66. Note that this dsep test cannot be done via the piecewise SEM library of R because its current version does not accept generalized additive model objects.

We can now compare alternative path models via the full-model AIC statistic. Our comparison involves four

competing models that differ in either the causal topology (Fig. 1a,b) and in whether the link between  $x_2$  and  $x_1$  is linear or nonlinear. The values of this full-model AIC statistic, and its behavior relative to Shipley's (2013) dsep AIC statistic, are given in Table 3. The ranking of the four models using the full-model AIC ( $\Delta AIC$  relative to the best model) is: correct topology and nonlinear link > correct topology and linear link (+10.33) > incorrect topology and nonlinear link (+19.39) > incorrect topology and linear link (+29.73). The increase in the full-model AIC statistic identifies the three poorer models as clearly inferior given common criteria (Anderson 2008). Using the dsep AIC the ranking is correct topology and linear link > correct topology and nonlinear link (+3.126) > incorrect topology and linear link (+17.24) > incorrect topology and nonlinear link (+20.37). The increase in the dsep AIC statistic again identifies the two models with the incorrect topology as clearly inferior, but the model with both the correct topology and nonlinear link is incorrectly identified as inferior.

To explore the degree to which the  $X^2_{ML}$  statistic approximates its theoretical asymptotic chi-squared distribution, we simulated 5,000 independent data sets for sample sizes of 50, 100, 200, and 500 independent observations, and following the multivariate normal and nonlinear, nonnormal models given in Table 2. Appendix S2: Table S1 summarizes the results. The model with linear relationships and multivariate normal distributions produced correct 5% null probabilities when sample sizes were at least 100 and the full tail distribution agreed with the asymptotic chi-square distribution at a sample size of 500. The model with nonnormal distributions and a nonlinear path had conservative null probability estimates even at a sample size of 500, although the degree of bias was small with an asymptotic probability estimate of 0.05 corresponding to an empirical probability of between 0.05 and 0.08 depending on the sample size.

DISCUSSION

These new generalized  $X^2_{ML}$  and full-model AIC statistic are appropriate for any piecewise SEM, the

TABLE 1. Parameter values for each of five normally distributed random variables ( $x_1$  to  $x_5$ ) generated following the directed acyclic graph shown in Fig. 1a.

Variable	Distribution ( $\varepsilon_i$ )	Generating equation	Statistical model	LL <sub>i</sub>	K <sub>i</sub>	AIC <sub>i</sub>
$x_1$	$N(\mu = 0, \sigma^2 = 1)$	$x_i = \varepsilon_i$				
$x_2$	$N(\mu = 0, \sigma^2 = 0.75)$	$x_2 = 0.5x_1 + \varepsilon_i$	<code>glm(<math>x_2 \sim x_1, \dots</math>)</code>	-126.839	3	259.678
$x_3$	$N(\mu = 0, \sigma^2 = 0.75)$	$X_3 = 0.5x_2 + \varepsilon_i$	<code>glm(<math>x_3 \sim x_2, \dots</math>)</code>	-112.682	3	231.364
$x_4$	$N(\mu = 0, \sigma^2 = 0.75)$	$X_4 = 0.5x_2 + \varepsilon_i$	<code>glm(<math>x_4 \sim x_1, \dots</math>)</code>	-131.711	3	269.422
$x_5$	$N(\mu = 0, \sigma^2 = 0.50)$	$X_5 = 0.5x_3 + 0.4x_4 + \varepsilon_i$	<code>glm(<math>x_5 \sim x_3 + x_4, \dots</math>)</code>	-102.919	4	213.838

Notes:  $N(\mu, \sigma)$  is the Normal distribution of the random variable ( $\varepsilon_i$ ), "Generating equation" gives the population values of the variable and "Statistical model" lists the statistical model (here, generalized linear models) used to estimate the paths. LL<sub>i</sub> is the maximum log likelihood of the *i*th submodel, given simulated data containing 100 observations, and K<sub>i</sub> is the number of free parameters in that submodel, including intercepts. Note the full model  $AIC = -2(\sum_i LL_i + \sum_i K_i) = -2(-474.151 - 12) = \sum_i AIC_i = 974.302$ .

TABLE 2. A more complicated model involving five random variables ( $x_1$  to  $x_5$ ) generated following the directed acyclic graph shown in Fig. 1a.

Variable	Distribution ( $\varepsilon_i$ )	Generating equation	Statistical model	LL <sub>i</sub>	K <sub>i</sub>	AIC <sub>i</sub>
$x_1$	$\chi^2(v = 7)$	$x_i = \varepsilon_i$				
$x_2$	$N(\mu = 0, \sigma = 1)$	$x_2 = \frac{10x_1}{1+5x_1} + \varepsilon_i$	$\text{gam}(x_2 \sim s(x_1), \dots)$	-144.511	4.56	298.148
$x_3$	$P(\lambda = 0.5x_2)$	$x_3 = \varepsilon_i$	$\text{glm}(x_3 \sim x_2, \dots)$	-185.841	2	375.682
$x_4$	$P(\lambda = 0.5x_2)$	$x_4 = \varepsilon_i$	$\text{glm}(x_4 \sim x_1, \dots)$	-196.640	2	397.279
$x_5$	$B(\eta = 1, \rho = \frac{e^{(-0.5x_3+0.5x_4)}}{1+e^{(-0.5x_3+0.5x_4)}})$	$x_5 = \varepsilon_i$	$\text{glm}(x_5 \sim x_3+x_4, \dots)$	-56.820	3	119.640

Notes: The probability density function of each random variable ( $\varepsilon_i$ ):  $\chi^2(v)$  is a chi-squared distribution with degrees of freedom  $v$ .  $P(\lambda)$  is a Poisson distribution with rate  $\lambda$ ;  $B(\eta, \rho)$  is a binomial distribution of size  $\eta$  and success probability  $\rho$ ;  $N(\mu, \sigma)$  is a normal (Gaussian) distribution with mean  $\mu$  and standard deviation  $\sigma$ . “Generating equation” gives the population values of the variable and “Statistical model” lists the statistical model (here, generalized linear or additive models) used to estimate the paths. LL<sub>i</sub> is the maximum log-likelihood of the  $i$ th submodel, given simulated data containing 100 observations, and K<sub>i</sub> is the number of free parameters in that submodel, including intercepts. Note the full model AIC =  $-2(\sum_i LL_i + \sum_i K_i) = -2(-583.812 - 11.56) = \sum_i AIC_i = 1190.75$ .

TABLE 3. Four competing models that differ in either their causal topology (correct = Fig. 1a and incorrect = Fig. 1b) and in whether the link between  $x_2$  and  $x_1$  is modeled as a linear ( $x_2 \sim x_1$ ) or nonlinear function via a regression smoother ( $x_2 \sim s(x_1)$ ).

Hypothesized model with linear ( $x_2 \sim x_1$ ) or nonlinear link ( $x_2 \sim s(x_1)$ )	Statistic	Correct topology	Incorrect topology
$x_2 \sim s(x_1)$	Log-likelihood	-583.812	-594.505
$x_2 \sim s(x_1)$	Number free parameters	11.563	10.563
$x_2 \sim s(x_1)$	$X_{ML}^2$ (df, null p)	3.35 (5, 0.65)	24.73 (6, $4e^{-4}$ )
$x_2 \sim s(x_1)$	Model AIC ( $\Delta AIC$ )	1,190.75 (0)	1,210.14 (+19.39)
$x_2 \sim s(x_1)$	C (df, null p)	7.68 (10, 0.66)	26.92 (12, $3e^{-3}$ )
$x_2 \sim s(x_1)$	dsep AIC ( $\Delta AIC$ )	30.806 (+3.126)	48.046 (+20.37)
$x_2 \sim x_1$	Log-likelihood	-590.544	-601.238
$x_2 \sim x_1$	Number free parameters	10	9
$x_2 \sim x_1$	$X_{ML}^2$ (df, null p)	16.81 (6.6, 0.01)	38.20 (7.56, $5e^{-6}$ )
$x_2 \sim x_1$	Model AIC ( $\Delta AIC$ )	1201.088 (+10.33)	1220.475 (+29.725)
$x_2 \sim x_1$	C (df, null p)	7.68 (10, 0.66)	26.92 (12, $3e^{-3}$ )
$x_2 \sim x_1$	dsep AIC ( $\Delta AIC$ )	27.68 (0)	44.92 (+17.24)

Notes: C (df, null p) is the dsep C-statistic, its degrees of freedom, and its null probability.  $X_{ML}^2$  (df, null p) is the maximum-likelihood test statistic, its degrees of freedom, and its null probability. Also listed are the AIC statistics ( $\Delta AIC$  relative to the best model) of the dsep method and of this paper, maximum log-likelihoods, the number of free parameters in the model.

submodels of which are fit using maximum likelihood estimation including linear, nonlinear, generalized linear, mixed, and generalized additive models. They are based on the local Markov decomposition, which implies that they cannot be applied to models involving latent variables or having free covariances (correlated errors). The values of these generalized versions are identical to those obtained in classical SEM when using the assumptions of classical SEM. Although Eqs. 3 and 5 have already been used in the artificial intelligence literature on Bayesian networks, assigning proper credit for these equations is difficult. Koller and Friedman (2009) developed Eq. 3 from Eq. 2 but did not link it to an AIC statistic. Liu et al. (2012) used the AIC statistic given in Eq. 5 but did not develop it. Both Silander and Myllymäki (2006) and Sullivant (2010) have independently proposed this method to find optimal topologies in Bayesian networks. To our knowledge, this paper is the first time that Eqs. 2-4 have been combined to produce this generalized  $X_{ML}^2$  test.

The dsep test and the generalized  $X_{ML}^2$  test are related but not identical. The dsep test concerns only the causal topology, as fixed by the d-separation claims. Depending on how the saturated model is constructed, the generalized  $X_{ML}^2$  test can either test only the causal topology or also the distributional assumptions and functional links. For example, when fitting the correct model to the data (Table 3) the dsep test and the  $X_{ML}^2$  test give essentially the same null probability ( $P = 0.66$  vs. 0.65). However, the  $X_{ML}^2$  test rejects the model ( $X_{ML}^2=16.81$ , 6.6 df,  $P = 0.01$ ) with a linear path between  $x_2$  and  $x_1$ , even though the causal topology is correct, when comparing it to the correct saturated model that includes this nonlinear path. This occurs because the only difference between the fitted model and its saturated version is the form (linear or nonlinear) of this path. When the model with an incorrect linear path is compared to a saturated version that also contains this linear path, the  $X_{ML}^2$  test does not reject the model, because, in this comparison, the only

difference is in the causal topology, which is correct ( $X^2_{ML}=3.35$ , 5 df,  $P = 0.65$ ). The  $X^2_{ML}$  therefore has some advantages over the C-statistic whenever the structural equations of the DAG can be expressed in a maximum-likelihood context: (1) model testing and parameter estimation are done simultaneously; in the case of mixed-effect models the calculation of  $P$ -values in the dsep test can be problematic; (2) it allows one to calculate modification indices similar to what is done in classical sem (Shipley 2016). However, the dsep test is still the only test available when the conditions required for maximum likelihood are not met; for example, when using distance matrices (Frenette-Dussault et al. 2013). Additionally, the dsep test is preferred with small sample sizes and exact tests of conditional independence. Furthermore, the individual d-separation claims in the basis set of the dsep test provide guidance for model modification when the overall model is rejected. The different conclusions of the full-model and dsep AIC statistics also require proper interpretation. One interpretation of the AIC statistic is that it compares model fit (quantified by the maximum log-likelihood) to model simplicity (quantified by the number of free parameters required to produce this fit) (Burnham and Anderson 2002, Anderson 2008). The dsep AIC concerns only differences in the causal topology of the model; it is meant only for comparing alternative models that differ in their causal topologies (i.e., their d-separation claims). It is for this reason that the dsep AIC clearly identifies the two models having the incorrect causal topology as inferior but, within each group having the same causal topology, incorrectly prefers the model with fewer free parameters, that is, the linear link between  $x_2$  and  $x_1$ . However, the full-model AIC statistic incorporates both the causal topology (the d-separation claims) and any other assumptions concerning the fitted parameter values of the full model. These assumptions include different probability distributions of the endogenous variables, different functional forms of the links between a parent and its causal children, equality constraints on subsets of parameters, and any complications arising from nesting, cross-classification, or correlations arising from spatial, temporal, or phylogenetic correlations if these can be modeled in a way that provides a likelihood function. Because these other aspects of the model are included with our new full-model AIC statistic, the model in Table 3 having the correct causal topology but an incorrect (i.e., linear) function linking  $x_2$  to  $x_1$  was identified as being clearly inferior. When one wishes to compare all aspects of a piecewise SEM model based on maximum-likelihood submodels and not just the causal topology, we therefore recommend that the generalized  $X^2_{ML}$  and this new full-model AIC statistic be used instead of the dsep test AIC statistic. However, when doing so, it is important to remember that these two generalized statistics reflect all errors in model fit, not just errors in the causal topology. Depending on the

purpose of the test, this could be an advantage or a disadvantage.

#### ACKNOWLEDGMENTS

We thank John Lefcheck, Marc Bélisle, and one anonymous reviewer for helpful suggestions.

#### LITERATURE CITED

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *In* B. N. Petrov and F. Csaki, editors. Proceedings of the 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.
- Anderson, D. R. 2008. Model based inference in the life sciences. A primer on evidence. Springer, New York, New York, USA.
- Beaumelle, L., D. Vile, I. Lamy, F. Vandenbulcke, F. Gimbert, and M. Hedde. 2016. A structural equation model of soil metal bioavailability to earthworms: confronting causal theory and observations using a laboratory exposure to field-contaminated soils. *Science of the Total Environment* 569–570:961–972.
- Bollen, K. A. 1989. Structural equations with latent variables. John Wiley and Sons, New York, New York, USA.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach. Springer-Verlag, New York, New York, USA.
- Daou, L., and B. Shipley. 2019. The measurement and quantification of generalized gradients of soil fertility relevant to plant community ecology. *Ecology* 100:e02549.
- Frenette-Dussault, C., B. Shipley, and Y. Hingrat. 2013. Linking plant and insect traits to understand multitrophic community structure in arid steppes. *Functional Ecology* 27:786–792.
- Haughton, D. M. A., J. H. L. Oud, and R. A. R. G. Jansen. 1997. Information and other criteria in structural equation model selection. *Communications in Statistics—Simulation and Computation* 26:1477–1516.
- Kang, C., and B. Shipley. 2009. A correction note on “A new inferential test for path models based on directed acyclic graphs.” *Structural Equation Modeling: A Multidisciplinary Journal* 16:537–538.
- Koller, D., and N. Friedman. 2009. Probabilistic graphical models. Principles and techniques. MIT Press, Cambridge, Massachusetts, USA.
- Lefcheck, J. S. 2016. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution* 7:573–579.
- Liu, Z., B. Malone, and C. Yuan. 2012. Empirical evaluation of scoring functions for Bayesian network model selection. *BMC Bioinformatics* 13(Suppl):15.
- Pearl, J. 1988. Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann, San Mateo, California, USA.
- Pearl, J. 2000. Causality. Cambridge University Press, Cambridge, UK.
- Pearl, J. 2009. Causality: models, reasoning, and inference. Second edition. Cambridge University Press, Cambridge, UK.
- R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.r-project.org
- Rosseeel, Y. 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48:1–36.
- Shipley, B. 2000. A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling* 7:206–218.

- Shipley, B. 2009. Confirmatory path analysis in a generalized multilevel context. *Ecology* 90:363–368.
- Shipley, B. 2013. The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology* 94:560–564.
- Shipley, B. 2016. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference with R. Second edition. Cambridge University Press, Cambridge, UK.
- Silander, T., and P. Myllymäki. 2006. A simple approach for finding the globally optimal Bayesian network structure. Pages 445–452 *in* Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, UAI 2006. Cambridge, Massachusetts, USA.
- Sullivant, S. 2010. Algebraic statistics. American Mathematical Society, Providence, Rhode Island, USA.
- Verma, T., and J. Pearl. 1988. Causal networks: Semantics and expressiveness. Pages 352–359 *in* Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence, Mountain View, UAI, Mountain View, CA.
- Wilks, S. S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics* 9:60–62.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/ecy.2960/supinfo>