Ecology, 95(3), 2014, pp. 631–636

Model selection for ecologists: the worldviews of AIC and BIC

Ken Aho,[1,4] DeWayne Derryberry,[2] and Teri Peterson[3]

[1]*Department of Biological Sciences, Idaho State University, Pocatello, Idaho 83209 USA*
[2]*Department of Mathematics, Idaho State University, Pocatello, Idaho 83209 USA*
[3]*Division of Health Sciences, Idaho State University, Pocatello, Idaho 83209 USA*

Introduction

Ecologists frequently ask questions that are best addressed with a model comparison approach. Under this system, the merit of several models is considered without necessarily requiring that (1) models are nested, (2) one of the models is true, and (3) only current data be used. This is in marked contrast to the pragmatic blend of Neyman-Pearson and Fisherian significance testing conventionally emphasized in biometric texts (Christensen 2005), in which (1) just two hypotheses are under consideration, representing a pairwise comparison of models, (2) one of the models, H_0, is assumed to be true, and (3) a single data set is used to quantify evidence concerning H_0.

As Murtaugh (2014) noted, null hypothesis testing can be extended to certain highly structured multi-model situations (nested with a clear sequence of tests), such as extra sums of squares approaches in general linear models, and drop in deviance tests in generalized linear models. This is especially true when there is the expectation that higher order interactions are not significant or nonexistent, and the testing of main effects does not depend on the order of the tests (as with completely balanced designs). There are, however, three scientific frameworks that are poorly handled by traditional hypothesis testing.

First, in questions requiring model comparison and selection, the null hypothesis testing paradigm becomes strained. Candidate models may be non-nested, a wide number of plausible models may exist, and all of the models may be approximations to reality. In this context, we are not assessing which model is correct (since none are correct), but which model has the best predictive accuracy, in particular, which model is expected to fit future observations well. Extensive ecological examples can be found in Johnson and Omland (2004), Burnham and Anderson (2002), and Anderson (2008).

Second, the null hypothesis testing paradigm is often inadequate for making inferences concerning the falsification or confirmation of scientific claims because it does not explicitly consider prior information. Scientists often do not consider a single data set to be adequate for research hypothesis rejection (Quinn and Keough 2002:35), particularly for complex hypotheses with a low degree of falsifiability (i.e., Popper 1959:266). Similarly, the support of hypotheses in the generation of scientific theories requires repeated corroboration (Ayala et al. 2008).

Third, ecologists and other scientists are frequently concerned with the plausibility of existing or default models, what statistician would consider null hypotheses (e.g., the ideal free distribution, classic insular biogeography, mathematic models for species interactions, archetypes for community succession and assembly, etc.). However, null hypothesis testing is structured in such a way that the null hypothesis cannot be directly *supported* by evidence. Introductory statistical and biometric textbooks go to great lengths to make this conceptual point (e.g., DeVeaux et al. 2013:511, 618, Moore 2010:376, Devore and Peck 1997:300–303).

Parsimony: Fit vs. Complexity

In deciding which model is the best, criteria are necessary that allow model comparisons. While some scientists feel that more complex models are always more desirable (cf. Gelman 2009), others prefer those that balance uncertainty, caused by excessively complex models, and bias, resulting from overly simplistic models. The latter approach emphasizes parsimony. A parsimonious model should (Aho 2013), "be based on (be subset from) a set of parameters identified by the investigator as ecologically important, including, if necessary, covariates, interactions, and higher order terms, and have as few parameters as possible (be as simple as possible, but no simpler)."

Consider the examination of species population descriptor (e.g., number of individuals) as a function of an environmental factor in which the true relationship between Y and X is $Y_i = e^{(X_i - 0.5)} - 1 + \varepsilon_i$, where $\varepsilon_i \sim N(0, 0.01)$ (black lines in Fig. 1). We randomly sample for the conditional values of Y_i 10 times and apply two models, a simple linear regression (Fig. 1a), and a fifth-order polynomial (Fig. 1b). The simpler model underfits the data and misses the nonlinear association of Y and X

Manuscript received 26 July 2013; revised 7 August 2013; accepted 12 August 2013. Corresponding Editor: A. M. Ellison. For reprints of this Forum, see footnote 1, p. 609.

[4] E-mail: ahoken@isu.edu

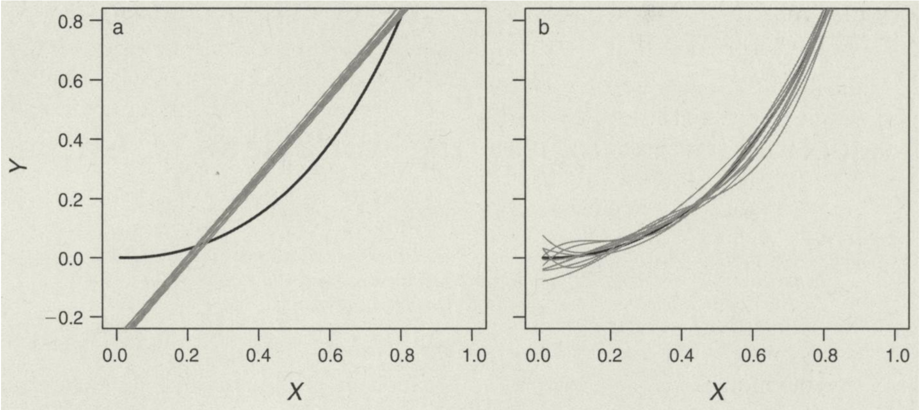

FIG. 1. Two sorts of models fit to a random process: (a) two parameter (simple linear regression) and (b) six parameter (fifth-order polynomial) (c.f., Sakamoto et al. 1986). The heavy black line indicates the true relationship between *Y* and *X*, while the gray lines are fits from linear models based on random data sets, each with 100 paired observations. Despite its complexity, the polynomial model is more parsimonious (average Akaike information criterion [AIC] = −167 vs. −42) because it captures the curvilinear nature of the association.

(Fig. 1a). The polynomial model, however, introduces erratic variability, and prevents statements of generality. Thus, both simplistic and overly complex models prevent valid inferences. The usefulness of a criterion that establishes the line between underfit and overfit models is obvious.

Two Parsimony Estimators: AIC and BIC

A Web of Science search conducted for this paper revealed that for ecological publications from 1993–2013, the two most popular measures of parsimony were the Akaike information criterion (AIC; Akaike 1973) and the Bayesian information criterion (BIC), also called the Schwarz or SIC criterion (Schwarz 1978). Specifically, for publications that implemented formal methods for multi-model inference, 84% used AIC, 14% used BIC, while only 2% used some other approach (Table 1). Murtaugh (2013) discusses AIC extensively in his defense of *P* values, but ignores BIC, prompting its consideration and comparison here. We posit that *P* values are at odds with BIC in the same way Bayesian hypothesis testing is at odds with *P* values (cf. Kass and Raftery 1995). Indeed, when substituting BIC for AIC in Murtaugh's derivation of *P* values from ΔAIC, fixed *P* values do not equate to fixed differences in BIC, unless *n* is fixed. This is consistent with the fact that *P* values must decrease (holding other factors constant) to favor the alternative hypothesis as sample size increases. AIC and BIC are defined as

$$\mathrm{AIC} = -2\ln\mathcal{L}(\hat{\theta}) + 2p$$

$$\mathrm{BIC} = -2\ln\mathcal{L}(\hat{\theta}) + p\ln n$$

where $\mathcal{L}(\hat{\theta})$ is the likelihood of the estimated model (in the context of general linear models, e.g., regression and ANOVA, this is the likelihood of the parameters in $N(0, \hat{\sigma}^2)$ given the model residuals, where $\hat{\sigma}^2$ is the maximum likelihood estimate for the variance of the error term distribution), p is the total number of parameters that are estimated in the model (including σ^2 for general linear models), and n is the sample size. For both indices, smaller values indicate better models.

AIC and BIC are generally introduced in textbooks (often together) as alternative measures for parsimony (cf. Kutner et al. 2005). Perhaps as a consequence, ecologists often use these measures interchangeably (or even simultaneously) without consideration of their differing qualities and merits. This view of exchangeability has, perhaps, been further entrenched by a recent ecological comparison of these methods that found no difference in efficacy among AIC, BIC, and several other criteria (Murtaugh 2009), and by articles that present these summaries side by side. However, we will show that this view is misplaced. In the remainder of this paper we explore and contrast BIC and AIC and make recommendations for their respective use in multi-model inference by ecologists.

TABLE 1. Results from a Web of Science search of publications on 13 August 2013 using the Science Citation Index Expanded (SCI-EXPANDED) database for the years 1993–2013.

Search terms for topic	No. citations	Proportion
Model selection AND (AIC* OR Akaike) AND ecol*	139	0.84
Model selection AND (BIC OR Bayes factor OR Schwarz) AND ecol*	23	0.14
Model selection AND (mallow OR FPE OR KIC OR Hannan-Quinn, Geweke-Meese) AND ecol*	4	0.02

AIC and BIC: Mathematical Motivations and Objectives

When simply examining the formula for AIC and BIC it is easy to misunderstand AIC and BIC as competing criteria intended to achieve the same goal. Both criteria balance simplicity (measured by, p, the dimension of the fitted model parameter space) and goodness of fit (measured by maximized likelihood). However the initial question, which curve "best" fits the data, can be paraphrased in a number of ways, and AIC and BIC are each answers to different questions, once the question is stated more precisely.

The most obvious reason likelihood alone cannot be used to pick between models is that models with more free parameters (when models are nested) will always have higher maximum likelihood. Akaike (1973) wanted to estimate the likelihood of a model while adjusting for the bias introduced by maximum likelihood. Using the Kullback-Liebler (KL) distance, he was able to formulate the log-likelihood maximization problem in such a way that the bias associated with likelihood maximization could be estimated and corrected for (see Burnham and Anderson 2002). Given discrete probability models, KL information is

$$I(f,g) = \sum_{x} f(x) \ln \left[\frac{f(x)}{g(x)} \right]$$

where $f(x)$, defines the probabilistic densities of the error distribution associated with the true model, while $g(x)$ defines the error density of an approximating model with known parameters. The term $I(f, g)$ represents the information lost when the candidate model is used to represent truth. Because the log of a quotient is the difference of logs, KL information can be separated into the difference of two summations. The first is equivalent to Shannon-Weiner diversity (information per individual) from community ecology (Pielou 1966). The second represents the log of the probability of the union of observed disjoint events.

Akaike's approach achieves an important objective: asymptotic efficiency (Shibata 1976). Asymptotic efficiency is essentially minimized prediction error. Criteria like AIC maximize predictive accuracy.

The approach taken by Schwarz (1978) is the asymptotic approximation, for the regular exponential family, of a Bayesian hypothesis testing procedure (Kass and Raftery 1995, Robert 2007). The BIC procedure derived by Schwarz is consistent. Thus, when the sample size increases, the correct model, from any group of models, is selected.

Schwarz and Akaike appear to have thought their approaches were in conflict. Schwarz (1978) wrote: "For large numbers of observations, the procedures differ markedly from each other. If the assumptions of Section 2 are accepted [for the formulation of the problem as a Bayesian hypothesis test see Kass and Raftery (1995)], Akaike's criterion cannot be asymptotically optimal." Akaike felt compelled to write a paper in response (Akaike 1978), which in our view does not clarify much, but does seem to indicate Akaike would like to address an apparent paradox. In fact the conflict is easily resolved once it is acknowledged that "asymptotically optimal" can have several meanings. Asymptotic efficiency and (asymptotic) consistency are different kinds of optimality.

McQuarrie and Tsai (1998) compare a large number of model selection procedures, and immediately divide them into two classes: consistent estimators, namely BIC, Hannan and Quinn information (Hannan and Quinn 1979), and GM (Geweke and Meese 1981), and efficient estimators, namely AIC, Mallows' C_p (Mallows 1973), predicted residual sum of squares (PRESS; Allen 1974), Akaike's FPE (Akaike 1969), and cross validation. A close link between leave-one-out cross validation and AIC can be found in Stone (1977).

It is now known that there is a class of model selection tools that provide the best predictive accuracy, and that class is headed by AIC. There is also a class of confirmation/falsification tools that are consistent, and that class is headed by BIC. So when would each be used?

Two World Views

Two different approaches to simulation

Consider two different simulations, A and B. In simulation A, a very complex model produces the data, and a number of models are candidates to fit the data. Because the process producing the data is very complex, we never expect the sample size of our data sets to approach d, the parameter space of the model (or process) producing the data (i.e., $d \gg n$), nor do we necessarily expect our candidate models to match the exact functional form of the true model. Thus, d, the number of parameters in the true model need not equal p, the number of parameters in a candidate statistical model, and the parameters for an optimal model may not include the complete pool of true parameters, and/or may include extraneous parameters.

In simulation B, a relatively simple process produces the data. The sample size of the data sets can be expected to greatly exceed d, the parameter space of the model generating the data (i.e., $d \ll n$). One of the candidate models being fitted to the data is actually equivalent to the actual model that produced the data

In these two contexts, the model that best fits the data must be interpreted differently. In simulation A, we can never find the true model, we can only find the model that maximizes predictive accuracy (model selection). In simulation B, we actually expect to find the correct model, as sample size increases (confirmation/falsification).

It will become clear that AIC is appropriate for real-world situations analogous to simulation A, and BIC is appropriate for real-world situations similar to simulation B. AIC will almost always outperform BIC in simulations designed like simulation A, and BIC will almost always outperform AIC in simulations similar to simulation B.

The BIC world

In an effort to make Bayesian inference more objective and more closely tied to Jeffreys' (1935) notion of evidence for a hypothesis, a number of statisticians (e.g., Casella et al. 2009), biometrists (Goodman 1999, Suchard et al. 2005), and ecologists (Link and Barker 2006, Ellison 1996) have adopted the notion of the Bayes factor (or posterior P values, see Ramsey and Schafer 2012) for hypothesis or model comparison. Suppose that two hypotheses, H_1 and H_2, are to be compared, then $\Pr(H_1\,|\,\text{data})/\Pr(H_2\,|\,\text{data})$ = posterior odds = Bayes factor × prior odds.

Kass and Raftery (1995), in their definitive paper, provide a motivation for Bayes factors and a number of applications where Bayes factors seem especially useful. The BIC formulation is an asymptotic approximation to the Bayes factor (Schwarz 1978, Robert 2007). Kass and Raftery routinely treat BIC as an approximation to Bayes factors. Thus, applications in this paper provide excellent examples where BIC would also be appropriate.

Kass and Raftery provide five such examples, including two from biology/environmental management. Each of these has two characteristics: (1) only a few potential hypotheses are considered and (2) one of the hypotheses is (essentially) correct. Although the second characteristic is not always overtly stated, they often cite consistency as a desirable asymptotic property of Bayes factors and/or BIC.

It is clear from their discussion of "Bayes factors vs. the AIC" (which is primarily a comparison of AIC and BIC) that they value BIC over AIC because it is consistent. That is, when the sample size is sufficiently large, BIC picks the correct model, while AIC picks a model more complex than the true model. This reflects a "worldview" in which hypotheses are being compared, and one of the hypotheses is correct.

The AIC world

A few scientists have a very different "world view." Breiman (2001) writes: "There are two cultures in the use of statistical modeling to reach conclusions about data. One assumes the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown." Breiman does not have an opinion on the question "AIC or BIC?" but he nonetheless seems to live in the world of simulation type A: he emphasizes the importance of cross-validation predictive accuracy as the measure of success, and models that grow in complexity as sample size increases. Similarly, Hurvich and Tsai (1989), with reference to autoregressive moving average (ARMA) time series modeling, write: "If the true model is infinite dimensional, a case that seems most realistic in practice, AIC provides an asymptotically efficient selection of a finite dimensional approximating model."

The prevalence of type A thinking is obvious throughout the popular biometric text on model selection by Burnham and Anderson (2002) and in other works by these authors (e.g., Anderson and Burnham 2002). This is because this worldview corresponds more closely to the reality of many biological investigations, particularly in ecology: extremely complex systems with an unknown (and perhaps unknowable) underlying structure (cf. Johnson and Omland 2004, Burnham et al. 2011).

Fig. 1 is typical of a type A simulation. The correct model is not one of the candidate models, so consistency is irrelevant. For those who are interested in AIC (often in forecasting or open-ended model selection) the common characteristics are (1) numerous hypotheses and (2) the conviction that all of them are to differing degrees wrong.

In the type A world, efficiency (predictive accuracy) is important. Overfitting means a model that will have a lot of random noise if used for future prediction, while underfitting means a model that will have a bias when used for future prediction. In the type A world, as sample size increases, more small (tapering) effects are picked up, and the size of the selected model increases.

In the type B world, consistency is important. Overfitting is picking a model more complex than the true model, and underfitting is picking a model simpler than the true model. As sample size increases, the true model rises to the top (cf. Anderson 2008: Appendix E).

It is easy to confuse these worlds

The quest for a procedure that is both consistent and efficient seems impossible, when looked at in this way. Specifically, efficient methods must pick larger models with increased sample size, whereas consistent methods must settle on a fixed complexity with increased sample size. One approach to model selection cannot do both. This view is supported mathematically by Yang (2005) who showed that while BIC is consistent in optimal model selection, it cannot be optimal for regression function estimation in the sense of multi-model inference, and that while AIC represents minimax-rate optimal rules for estimating the regression function, it is not consistent for optimal model selection.

One of the paradoxes of model selection is that almost all research is based on type B simulations (Breiman 2001), but most statisticians, and even ecologists (e.g., Bolker 2008, Scheiner and Willig 2011) love to quote George Box: "All models are wrong, but some are useful." It should be noted that Box, and his most important work, time series forecasting, is fundamentally type A. At least one well-known statistics textbook suggests data splitting as the optimal way to find the best model, but if this is impossible one should use BIC, as opposed to AIC, because it is consistent—an odd

TABLE 2. The worlds of AIC and BIC contrasted.

Factor	AIC	BIC
Mathematical characteristics		
Derivation	Estimated information loss.	Approximate Bayes factor.
Optimality criterion	Asymptotic efficiency.	Consistency.
Close cousins	Data splitting, Mallows' C_p, PRESS.	Hannan-Quinn, Geweke and Meese, Bayes factors, and Bayesian hypothesis testing.
World View		
Problem statement	Multiple incompletely specified or infinite parameter models.	A small number of completely specified models/hypotheses.
Perspective	"All models are wrong, but some are useful."	"Which model is correct?"
Simulation structure	$d \gg n$	$d \ll n$
With increased n ...	Best model grows more complex.	Procedure focuses in on one best model.
Applications		
Context	Exploratory analysis; model selection to address which model will best predict the next sample; imprecise modeling; tapering effects.	Confirmatory analysis; hypothesis testing; model selection to address which model generated the data; Low dimension, precisely specified models.
Ecological examples	Complex model selection applications, e.g., predictive models for community, landscape, and ecosystem ecology; time series applications including forecasting.	Controlled experiments, for instance in physiology/enzymatics/genetics with a limited number of important, well-understood, biological predictors; models including expected or default (null) frameworks, e.g., enzyme kinetics models, Hardy-Weinberg equilibrium, or RAD curves, one of which is expected to be correct.

Notes: The number of parameters in the true model is d; sample size is n. Abbreviations are: PRESS, predicted residual sum of squares; and RAD, ranked abundance distribution.

mixture of type A and type B reasoning (Montgomery et al. 2008:60).

It is interesting to consider the performance of AIC and BIC in the context of increasingly large data sets. With respect to BIC, it is clear that, given type B simulation, the larger the sample size, the larger the probability BIC selects the true model. The relationship is less well defined with AIC (since n is not specified in its formula). However, one would expect that, in a type A simulation, as sample size increases (and consequently larger models are selected), that predictive power would also increase. Thus, as n grows larger both criteria will work better, but with different goals in mind.

There doesn't seem to be any basis for always preferring one world view over the other, both have a place in ecological model selection (cf. Murtaugh 2009). However, there are reasons to be aware that there are two world views, and to remain consistently within a given world view on a given modeling problem.

Model selection and confirmation/falsification contrasted

Table 2 can be seen as a framework for asking questions to pin down whether AIC (or related tools) or BIC (or related tools) are appropriate for a given application. Some questions, motivated by this table are: Is your analysis exploratory (AIC) or confirmatory (BIC)? Is the analysis open-ended (AIC), or are a few specific models representing a well understood process being compared (BIC)? As the data set gets larger, do you expect your model to grow in complexity (AIC), or stabilize (BIC)? Do you believe you have chosen the correct functional form of the relationship as well as the correct variables (yes, BIC; no, AIC)? Is your goal accurate prediction (AIC) or finding the correct model (BIC)?

CONCLUSION

Murtaugh (2014) revealed an important mathematical connection between ΔAIC and P values for a comparison of two models (one nested in the other). Such an application, however, constitutes a very narrow use of an information-theoretic criterion. We agree with Murtaugh that null hypothesis testing has an important role in ecology, and that conceptual problems with this paradigm are often due to misapplication and misunderstanding by users. Nonetheless, many ecological endeavors pose questions that are not easily answered by null hypothesis tests. For instance, models may not be nested, and the ecologist may want to treat the null and alternative hypothesis as having the same status with regard to support based on the evidence. There are tools for this situation, but the proper tool depends on a further distinction. What has often been designated as model selection has been here further parsed into complex (infinite) model selection, for which AIC and related tools are the appropriate; and confirmation/falsification, for which BIC and related tools are appropriate.

Literature Cited

Aho, K. 2013. Foundational and applied statistics for biologists using R. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Akaike, H. 1969. Statistical predictor identification. Annals of the Institute of Statistical Mathematics 22:203–217.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 *in* B. N. Petrov and S. Caski, editors. Proceedings of the Second International Symposium on Information Theory. Akademiai Kaido, Budapest, Hungary.

Akaike, H. 1978. A Bayesian analysis of the minimum AIC procedure. Annals of the Institute of Statistical Mathematics 30:9–14.

Allen, D. M. 1974. The relationship between variable selection and data augmentation and a method of prediction. Technometrics 16:125–127.

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

Anderson, D. R., and K. P. Burnham. 2002. Avoiding pitfalls when using information-theoretic methods. Journal of Wildlife Management 66(3):912–916.

Ayala, F. J., et al. 2008. Science, evolution, and creationism. National Academies Press, Washington, D.C., USA.

Bolker, B. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey, USA.

Breiman, L. 2001. Statistical modeling: the two cultures. Statistical Science 16(3):199–215.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference, a practical information-theoretic approach. Second edition. Springer, New York, New York, USA.

Burnham, K. P., D. R. Anderson, and K. P. Huyvaert. 2011. AIC model selection and multi-model inference in behavioral ecology: some background, observations and comparisons. Behavioral Ecology and Sociobiology 65:23–35.

Casella, G., F. J. Giron, M. L. Martinez, and E. Moreno. 2009. Consistency of Bayesian procedures for variable selection. Annals of Statistics 37(3):1207–1228.

Christensen, R. 2005. Testing Fisher, Neyman, Pearson, and Bayes. American Statistician 59(2):121–126.

DeVeaux, R. D., P. F. Velleman, and D. E. Bock. 2013. Intro stats. Fourth edition. Pearson, Upper Saddle River, New Jersey, USA.

Devore, J. L., and R. Peck. 1997. Statistics: the explorations and analysis of data. Duxbury, Pacific Grove, California, USA.

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications 6:1036–1046.

Gelman, A. 2009. Bayes, Jeffreys' prior distributions and the philosophy of statistics. Statistical Science 24(2):176–178.

Geweke, J., and R. Meese. 1981. Estimating regression models of finite but unknown order. International Economic Review 22:55–70.

Goodman, S. N. 1999. Toward evidence based medical statistics 2: The Bayes factor. Annals of Internal Medicine 130(12):1005–1013.

Hannan, E. J., and B. G. Quinn. 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society B 41:190–195.

Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76(2):297–307.

Jeffreys, H. 1935. Some tests of significance, treated by the theory of probability. Proceedings of the Cambridge Philosophical Society 31:203–222.

Johnson, J., and K. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19(2):101–108.

Kass, R. E., and E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90(430):773–795.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. Applied linear statistical models. Fifth edition. McGraw-Hill, Boston, Massachusetts, USA.

Link, W. A., and R. J. Barker. 2006. Model weights and the foundations of multimodel Inference. Ecological Applications 87:2626–2635.

Mallows, C. L. 1973. Some comments on C_P. Technometrics 15 (4):661–675.

McQuarrie, A. D. R., and C.-L. Tsai. 1998. Regression and time series model selection. World Scientific, Singapore.

Montgomery, D. C., C. L. Jennings, and M. Kulahci. 2008. Introduction to time series analysis and forecasting. Wiley series in probability and statistics. Wiley, Hoboken, New Jersey, USA.

Moore, D. S. 2010. The basic practice of statistics. Fifth edition. Freeman, New York, New York, USA.

Murtaugh, P. A. 2009. Performance of several variable-selection methods applied to real ecological data. Ecology Letters 12(10):1061–1068.

Murtaugh, P. A. 2014. In defense of *P* values. Ecology 95:611–617.

Pielou, E. C. 1966. Shannon's formula as a measure of specific diversity: its use and misuse. American Naturalist 100(914):463–465.

Popper, K. 1959. The logic of scientific discovery. Routledge, London, UK.

Quinn, G. P., and M. J. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, UK.

Ramsey, F. L., and D. W. Schafer. 2012. The statistical sleuth: a course in the methods of data analysis. Third edition. Brooks/Cole, Belmont, California, USA.

Robert, C. P. 2007. The Bayesian choice: from decision-theoretic foundations to computational implementation. Second edition. Springer, New York, New York, USA.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1986. Akaike information criterion statistics. KTK Scientific Publishers, Tokyo, Japan.

Scheiner, S. M., and M. R. Willig. 2011. The theory of ecology. University of Chicago Press, Chicago, Illinois, USA.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.

Shibata, R. 1976. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 63:117–126.

Stone, M. 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. Journal of the Royal Statistical Society B 39(1):44–47.

Suchard, M. A., R. E. Weiss, and J. S. Sinsheimer. 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophylogeny. Biometrics 61:665–673.

Yang, Y. 2005. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. Biometrika 92(4):937–950.