AVOIDING PITFALLS WHEN USING INFORMATION–THEORETIC METHODS

DAVID R. ANDERSON,[1,2] Colorado Cooperative Fish and Wildlife Research Unit, Room 201 Wagar Building, Colorado State University, Fort Collins, CO 80523, USA

KENNETH P. BURNHAM,[1] Colorado Cooperative Fish and Wildlife Research Unit, Room 201 Wagar Building, Colorado State University, Fort Collins, CO 80523, USA

Abstract: We offer suggestions to avoid misuse of information–theoretic methods in wildlife laboratory and field studies. Our suggestions relate to basic science issues and the need to ask deeper questions (4 problems are noted), errors in the way that analytical methods are used (7 problems), and outright mistakes seen commonly in the published literature (5 problems). We assume that readers are familiar with the information–theoretic approaches and provide several examples of misuse. Any method can be misused—our purpose here is to suggest constructive ways to avoid misuse.

Key words: Akaike's Information Criterion, analysis guidelines, information–theoretic methods, Kullback-Leibler information, model selection uncertainty, modeling, *P*-values.

Support for null hypothesis testing and associated *P*-values has declined among statisticians over the past several decades (Anderson et al. 2000). Papers by Yoccoz (1991), Cherry (1998), Johnson (1999), and Anderson et al. (2000) outline the overuse, misuse, and limitations of null hypothesis testing in wildlife research. Guthery et al. (2001:379) note the "collapse of null hypothesis significance testing as a statistical paradigm...." Many wildlife biologists and ecologists have changed their perspectives regarding data analysis as a result of the limitations of null hypothesis testing. Some investigators have merely refocused attention on estimating effect sizes and measuring their precision, without the undo emphasis on a null hypothesis, test statistics, *P*-values, and arbitrary notions of significance (Yoccoz 1991, Anderson et al. 2001*b*). This simple approach is effective for many biological questions. We support the estimation of effect size and use it in our own research work. Some investigators have begun to explore an assortment of Bayesian methods (Gelman et al. 1995, Ellison 1996), often in close collaboration with a statistician. Other researchers have begun to use the relatively new information–theoretic methods, especially for science problems involving some substantial complexity. Our review of several of the leading theoretical and applied ecological journals shows that use of the information–theoretic methods is increasing.

Any methodology can be misused. In the sciences, this misuse usually is unintended; the investigator misunderstood some aspect of an approach and failed to use it appropriately. The information–theoretic approach can be misused, and we have observed that some misuse has occurred already. Our objective is to provide suggestions that we hope will allow investigators to reduce the inappropriate use of information–theoretic approaches in wildlife studies that involve analysis of empirical data.

We must assume that the reader is familiar with Kullback-Leibler (K-L) information and the related information criteria for model selection (i.e., AIC [Akaike's Information Criterion], AIC_c [AIC corrected for small-sample bias], and $QAIC_c$ [AIC_c for overdispersed data]; these are defined and explained in Burnham and Anderson 2002). Furthermore, the reader must have an understanding of information criterion differences (Δ_i); likelihood of model i, given the data [$\mathcal{L}(g_i \mid data)$]; Akaike weights ($w_i$); evidence ratios (e.g., w_i/w_j); and simple approaches to incorporating model selection uncertainty into estimates of precision (e.g., unconditional standard errors, model averaging). Assuming the reader has knowledge of the information–theoretic approaches, we provide several suggestions to avoid misuse.

SCIENCE ISSUES

We outline 4 problems that deal with the underlying science of the issue under the information–theoretic approach. These issues precede formal data analysis; they are the very reason for analysis and therefore of fundamental importance.

---

[1] Employed by U.S. Geological Survey, Biological Resources Division.

[2] E-mail: anderson@cnr.colostate.edu

Poor Science Question

The information–theoretic approach, as developed by Akaike (1973, 1974), represents an integrated package; much of this package rests on the quality of the underlying science question and the specific alternative hypotheses. Good science questions should be addressed, and this is a continued challenge to us all (O'Connor 2000, Morrison 2001). We must move beyond null hypotheses (which usually are trivial) to more exciting, relevant questions at the edge of current knowledge. Instead of a (silly) null and a single alternative hypothesis, science is better served by multiple working hypotheses, a concept advocated by Chamberlin (1965). Science hypotheses and the specific data must be represented carefully by well-chosen statistical models. Failure to achieve several good alternative science hypotheses and models to represent them leads to the problem that none of the models are useful—a recent concern of Leopold (2001). We urge additional thought and consideration of the quality of the science question posed.

The failure to properly address these basic issues often leads to a poor or even counter-intuitive result. This undesirable endpoint is not the fault of the information–theoretic approaches; rather, one should not expect a good model, as an inference, if they start with only poor models, representing poor science hypotheses. The challenge is to begin with better science questions.

Too Many Models

A common problem, related closely to the issue above, is that too often little thinking precedes data analysis; thus, the computer is asked to analyze and compute results for many models. Often dozens, hundreds, or even thousands of models are analyzed using a statistical software package. It is not unusual to see papers in the published literature in which the number of models exceeds the number of data points (sample size, n), and this is certainly problematic (Freedman 1983, Miller 1990). Fleishman et al. (2001) used over 8 million models to study butterflies as a function of 23 environmental variables. Results from such an analysis strategy will almost surely be spurious and are an example of Freedman's paradox (Freedman 1983).

We recommend careful consideration of the science of the issue to reduce the number of models (Burnham and Anderson 1998:176–199). Use of all subsets regression is a flag that an unthinking approach has been employed. While this approach might be useful in exploratory work, or maybe if prediction is the only goal, it would seem better to first try to reduce the model set to a few where some a priori support exists.

The True Model Is Not in the Set

No models exist that exactly represent full truth, and researchers should not expect such a true model to be in the a priori set of models. This error in thinking lead to model selection criteria such as BIC (Bayesian Information Criterion, sometimes called SIC [Schwarz's Information Criterion]; Schwarz 1978) that attempt to estimate the dimension of the true model. Such criteria are not estimates of Kullback-Leibler information and generally are less useful since they are only asymptotic criteria and hinge on assumptions that we believe are unreasonable. (Even then, BIC often will perform better than many standard approaches such as stepwise regression [McQuarrie and Tsai 1998].) The information–theoretic approach does not assume that the true model is in the set and, in fact, does not assume any such full reality is even parameterized.

The information–theoretic approach estimates the relative closeness of each fitted model to conceptual truth or full reality. The goal is to find the best fitted-model in the set; this is quite different than finding truth—finding full truth is unattainable with a finite sample size. Bayesians are reaching this same view of models as good approximations to truth and not full truth (Spiegelhalter 2002).

Information–theoretic Methods Are Not a "Test"

Null hypothesis tests and associated *P*-values should not be mixed with various evidentiary results from an information–theoretic approach (e.g., AIC$_c$, Δ_i, w_i); this mixing error has been common in recent issues of *The Journal of Wildlife Management* (e.g., Budnik et al. 2002). For example, once AIC$_c$ has been used to rank models, no theory exists to suggest that a test statistic associated with a test between any 2 models has some known distribution (i.e., *F* or *t* or *z* or χ^2) as a basis for computing *P*-values. Hypothesis testing usually involves nested models, and this is another limitation. There are other problems (e.g., *P*-values are not a valid measure of evidence; Sellke et al. 2001) when mixing these 2 analysis paradigms, and we recommend strongly against this practice.

The largely arbitrary notions of significance and *P*-values are not part of the information–theoretic paradigm. There is no focus on a null

hypothesis that is nearly always false on a priori grounds (Cherry 1998, Johnson 1999, Anderson et al. 2000); instead, there is a focus on a small set of science hypotheses, all of which are plausible. There is no arbitrary, automatic decision to be reached (e.g., $P < \alpha$); rather, quantitative evidence is provided to allow a careful interpretation (e.g., Akaike weights and evidence ratios). Major advances have been made in the theory and application of data analysis methods in the past 10 years. It should not be surprising that null hypothesis testing is no longer very useful, considering that it was developed 70–80 years ago.

The information–theoretic approach allows a more extended analysis than that allowed by more traditional, testing-based approaches. For example, science hypotheses, represented by models, can be ranked from best to worst, and scaled to allow an understanding as to which hypotheses might be close, while others are separated widely, in terms of empirical support for them. The evidence, rather than the usual significant versus not significant, is quantified by the Akaike weights and evidence ratios. Model selection uncertainty can be incorporated easily into estimates of precision; this feature is neither easy nor well-founded under the testing-based methods. Finally, formal statistical inference can be based routinely on all the models in the set (multimodel inference), rather than on merely the model estimated to be best. This inference feature is particularly useful in prediction, in which often 2–4 models may be nearly tied in their level of empirical support, but can often give quite different predictions. These additional levels of inference are not supported by the testing-based methods.

METHODOLOGICAL ISSUES

We provide 7 suggestions related to data analysis methodologies under an information–theoretic paradigm. In several instances, we cite an example to aid in understanding the various points.

Poor Modeling of Hypotheses

Once a set of alternative hypotheses have been defined, one must carefully represent these by mathematical models relevant to the data at hand. Such modeling may not always be easy, since the data and various sampling issues must enter consideration. The need for modeling expertise is an excellent reason to seek the help of a statistician. There is a great deal of sophistication to allow good models of various hypotheses, but few biologists and ecologists are aware of

these approaches. For example, one might first think of a set of multiple linear regression models to represent hypotheses concerning some response variable (Y) as a function of wildlife habitat variables (X_i). While this is often taught, one should be leery of relationships that are hypothesized to be linear. Further thinking should lead one to consider asymptotes, thresholds, and other nonlinearities in modeling biological systems. For example, rather than modeling sage-grouse (*Centrocercus* spp.) abundance as a linear function of percent canopy cover of sagebrush (*Artemisia* spp.), it might be more relevant to hypothesize a quadratic relationship. This approach might suggest, for example, that sage-grouse abundance is at a maximum at 40% canopy cover, whereas abundance is less at both lesser and greater values than 40%. Such insights would not come from a linear model. Too little thought often is given to predictor variables measured or constructed, and this represents an area for improvement.

Binomial random variables are very common in wildlife research, but relatively few people realize the advantages of parameterizations based on the logit transform $[\mathrm{logit}(p) = \log_e \frac{p}{1-p}]$, where p is a binomial parameter such as a survival, emigration, capture, or transfer probability (McCullagh and Nelder 1989). Such models often allow meaningful nonlinearities without introducing additional parameters to be estimated. A set of only linear models is another flag that modeling may be relatively poor. We urge that more expertise be brought to bear in modeling the science hypotheses of interest on relationship to the sampled data.

Failure to Consider Various Aspects of Model Selection Uncertainty

Estimates of precision should include a variance component for model selection uncertainty; Breiman (1992) noted that failure to do so constituted a quiet scandal. Procedures such as stepwise regression often mislead the investigator concerning the large uncertainty in model selection. This misimpression stems from the fact that so little output is given and the investigator is tempted to believe that a clearly best model has been found and is a solid basis for inference.

One might have 10 models, each representing 1 of 10 science hypotheses, and finds that 3 models have similar levels of empirical support but provide substantially different predictions; the remaining 7 models have little support. In this example, uncertainty exists concerning which

model is best, and this uncertainty should be part of the estimate of precision; simple methods exist to allow this addition.

Model selection uncertainty is a large, important issue. We have found that many advantages are gained by making formal statistical inference from all models in the set. Multimodel inference (Burnham and Anderson 2002) is useful in numerous science contexts, including the ranking of the relative importance of explanatory variables and parameter estimators that tend to decrease model selection bias. Arnold et al. (2002) provide an interesting example in which model averaging might have provided valuable insights.

Failure to Consider Overdispersion in Count Data

Count data often are overdispersed, that is, the estimated theoretical sampling variance is too small, and this is often caused by a lack of independence. While a model might assume independence, the data may be somewhat dependent; thus, the data are said to be overdispersed and the sampling variances are underestimated if based only on theory. Numerous ways have been suggested to deal with this common case, and most methods fall under the name quasi-likelihood (Wedderburn 1974). A simple approach that has been shown to be effective is to estimate a variance inflation factor (c) based on a standard goodness-of-fit statistic (Burnham et al. 1987:243–246) from a high-dimensional model. Given an estimate of the variance inflation factor ($\hat{c}$) one should use QAIC$_c$ for model selection and inference and inflate the estimates of the variance–covariance matrix by $\hat{c}$ (i.e., $\hat{c} \times \hat{\Sigma}$, where Σ is the theoretical variance–covariance matrix). Franklin et al. (2002) examined long-term data sets for a variety of avian species and found overdispersion to be common, but not large. Only 17% of the 107 cases were found in which $\hat{c} > 1.8$, and estimates of overdispersion ranged from 1.13 to 2.37.

Heyde (1997) provides more advanced methods for analysis of overdispersed data. Shefferson et al. (2001) provide an excellent example of the use of QAIC$_c$ and variance inflation in estimating dormancy in a rare orchid species.

Post hoc Exploration of Data Not Admitted

After some post hoc examination of the data has been done, the publication of results should separate clearly those inferences arising from a priori considerations, which then tend to be more confirmatory, from the hunches developed after examining the data in detail (Anderson et al. 2001*a*). While we fully support some post hoc, exploratory activities to better understand the data (or generate hypotheses), the results from such activities should be admitted in publication. One can always conduct analyses based on a priori thinking, followed by some post hoc considerations; the reverse is never an option.

Statistical Significance versus Quantitative Evidence

Many of us were taught to think in terms of α-levels and *P*-values, and these numerical values made it seemingly easy to judge importance automatically (even if it was statistical rather than biological importance; Yoccoz 1991). The new challenge asks one to evaluate the quantitative evidence directly in terms of the biological science. For example, if the evidence ratio between models of compensatory and additive mortality was 150 (corresponding to a Δ_i value of about 10), nearly everyone would interpret this evidence as strong, whereas an evidence ratio of only 2.7 (corresponding to a Δ_i value of 2) would not be strong at all. Some evidence is equivocal (Royall 1997), such as an evidence ratio of 7.4 (corresponding to a Δ_i value of 4); here, there is no attempt at an automatic dichotomy (which is really a decision). Instead, a strength of evidence is given and open to biological interpretation; we should not expect everyone to agree on the same, exact interpretation of the evidence (just as juries often are not unanimous in judicial proceedings).

Science is about evidence and conclusions, not so much about decisions (see e.g., Tukey 1960). The traditional Neyman-Pearson system of testing null hypotheses essentially was embedded in a decision-making context, not a weight of evidence context. Some people still misuse *P*-values as if they were a proper weight of evidence (Sellke et al. 2001). The usefulness of *P*-values is quite limited, and we continue to suggest that these procedures be euthanized.

GOF Should Be Assessed Using the Global Model

While computer software often computes a goodness-of-fit test for each model, this assessment should be done primarily for the global model or the most highly parameterized model. If the global model fits, then criteria based on K-L information will not select a parsimonious

model that does not fit (BIC does not have this property).

Failure to Provide All the Needed Information

We suggest presenting a table to show the value of the maximized log-likelihood [$\log(\mathcal{L})$], number of estimable parameters (K), the appropriate information criterion used (e.g., $QAIC_c$), the differences (Δ_i), and the Akaike weights (w_i) for each model in the Results section. Failure to provide this information limits certain kinds of interpretation and understanding. For example, consider model A with K parameters and model B with $K + 1$ parameters where Δ_2 is approximately 2. There are cases where the fit, as measured by the $\log(\mathcal{L})$ value, remains nearly unchanged for model B; thus, the only reason model B seems competitive is that it is within 2 units of the best model. Lacking knowledge of the maximized log-likelihood value one might conclude incorrectly that model B was a close competitor. None of the 6 papers using information–theoretic approaches in the January 2002 issue of *The Journal of Wildlife Management* (vol. 66, no. 1) provided all the information needed for careful review and understanding.

Relatively few published papers seem to recognize that inference is based on Kullback-Leibler information (Kullback and Leibler 1951) and that various information criteria (i.e., AIC, AIC_c, and $QAIC_c$) are estimators of relative K-L information (Burnham and Anderson 2002). While not a central point, it seems preferable to keep in mind that relative K-L information is the fundamental issue and that there are simple ways to compute estimates of this quantity for each model in the set. A single AIC_c value is useless; it is the comparison of AIC_c values across the approximating models in the set that is relevant.

OUTRIGHT MISTAKES

Several misuses of the information–theoretic methods are merely mistakes; we offer 5 suggestions here.

The Incorrect Number of Estimable Parameters, *K*

In least squares regression it is easy to forget to include the estimate of σ^2 in K. In these cases, K should reflect the intercept (β_0), the slope parameters (β_i), and the residual variance σ^2. In modeling count data where a simple variance inflation factor is used, K should include the estimation of c. In logistic regression, K should include the intercept and slope parameters (i.e., there is no σ^2; instead, binomial variance is assumed). If a variance inflation factor (c) is used in logistic regression, this parameter must be counted in determining K.

Use of AIC Instead of AIC$_c$

A very common problem is the use of AIC when the sample size and/or number of estimable parameters < 40 for the model with the largest K. This error is often caused by computer software where only AIC values are computed and printed. However, even books on statistical methods written by statisticians perpetuate the use of the wrong criterion. For instance, Leonard and Hsu (1999:23–26) provide an example where AIC is used when the sample size is only 16. They did not even report results from BIC because they felt that the sample was too small. When sample size is small or the number of parameters is large, it is critical to use the small sample criteria, AIC_c or $QAIC_c$. After a proper criterion has been established (e.g., AIC_c), it should be used for all the models in the set.

Using AIC in All Subsets Selection

Several software packages allow AIC to be used in all subsets selection in which all combinations of predictor variables are used automatically. This unthinking procedure represents a "just-the-numbers" approach and violates the entire spirit of the information–theoretic approach. We should avoid believing that a computer alone can find the important relationships automatically from numerous models. The biology and underlying science should drive the hypotheses and modeling, rather than letting the computer analyze all possible models (even if a good selection criterion is used).

Few investigators realize the issues in interpreting correctly the output from a stepwise regression on a large number of variables (Miller 1990). For example, a common misuse is to assume that the variables in the selected model are important, while other variables are unimportant. Model selection uncertainty and bias are major issues in good data analysis when models with more than 3–5 parameters are estimated and sample size is small to moderate.

Of all the possible models, the stepwise procedures fit only some of these models (and often a small number). Of those fit, even fewer model results are given explicitly. For example, with 13 predictor variables there are 8,192 possible models. However, even with sample sizes in the hundreds, most stepwise routines provide results for

only, say, 3–8 models (i.e., the best model with 1, 2, ..., 8 variables, respectively). This output is very misleading to the investigator as model selection uncertainty may appear minor. In fact, all models were under consideration, subject to fitting. While stepwise procedures can be viewed as computationally efficient, they also are poor in selecting a model with good statistical properties and misleading in terms of model selection uncertainty (McQuarrie and Tsai 1998:427–429).

Information Criteria Are Not Comparable Across Different Data Sets or Different Response Variables

Two mistakes are more subtle to recognize. First, one cannot validly compare an information criterion across different data sets. The data must be considered fixed, and then models in the set can be compared and ranked. Some multiple regression software packages treat missing data in ways that invalidate the proper comparison of AIC values. Second, within the model set, it is not proper to let the response variable change. For example, 3 models of the response variable y cannot be compared with 4 models of the response variable log (y). Instead, all the models in the set must use the same response variable.

Failure of Numerical Methods to Converge

Typically, parameter estimates derived by the method of maximum likelihood are calculated numerically using a computer. Such computer routines usually are highly sophisticated but can fail to converge on the global maximum of the log-likelihood function. There are a host of technical reasons why the iterative routine might fail to converge, but generally, notice of such failure is provided in the output results. If the numerical method did not converge, then the maximum of the log-likelihood is not found and the various information criteria will be incorrect. In these cases, it often may be best to consult with an expert to determine the cause of the convergence failure.

SUMMARY

Information–theoretic methods can be misused—we hope to see their misuse minimized. We encourage reading and study to understand these approaches and their proper application. For example, if prediction is the central issue, formal inference should be based on model averaging, and this simple method has many important advantages. If the science question is simple, sample size is large, and effects are large, many

data analysis methods will allow decent inferences. In other cases, the newer methods stand apart in terms of good statistical properties of the selected model (e.g., bias, precision, achieved confidence interval coverage, prediction error).

One of the many criticisms of the null hypothesis testing approach and the associated α-levels and P-values was the incredible misuse of these procedures (Anderson et al. 2000). Such misuse is hardly the fault of the method; still, the misuse was nearly universal and this further hurt the credibility of the approach. Information–theoretic approaches are much more relevant in providing strong inference than traditional approaches so often taught in beginning classes. These new approaches arise strongly from the science of the issue and provide several pieces of quantitative evidence (e.g., ranking of all models, evidence ratios), rather than an arbitrary dichotomy (reject or not). We hope that the suggestions we offer both limit misuse of the new information–theoretic approaches as well as encourage investigators to use the full power of these approaches (e.g., formal statistical inference based on multiple models).

LITERATURE CITED

AKAIKE, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267–281 *in* B. N. Petrov and F. Csaki, editors. Second international symposium on information theory. Akademiai Kiado, Budapest, Hungary.

———. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19:716–723.

ANDERSON, D. R., K. P. BURNHAM, W. R. GOULD, AND S. CHERRY. 2001*a*. Concerns about finding effects that are actually spurious. Wildlife Society Bulletin 29:311–316.

———, ———, AND W. L. THOMPSON. 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64:912–923.

———, W. A. LINK, D. K. JOHNSON, AND K. P. BURNHAM. 2001*b*. Suggestions for presenting the results of data analysis. Journal of Wildlife Management 65:373–378.

ARNOLD, T. W., M. G. ANDERSON, M. D. SORENSON, AND R. B. EMERY. 2002. Survival and philopatry of female redheads breeding in southwestern Manitoba. Journal of Wildlife Management 66:162–169.

BREIMAN, L. 1992. The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error. Journal of the American Statistical Association 87:738–754.

BUDNIK, J. M., F. R. THOMPSON, III, AND M. R. RYAN. 2002. Effect of habitat characteristics on the probability of parasitism and predation of Bell's vireo nests. Journal of Wildlife Management 66:232–239.

BURNHAM, K. P., AND D. R. ANDERSON. 1998. Model selection and inference: a practical information–theoretic approach. Springer-Verlag, New York, USA.

———, AND ———. 2002. Model selection and multi-model inference: a practical information–theoretic approach. Second edition. Springer-Verlag, New York, USA.

———, ———, G. C. WHITE, C. BROWNIE, AND K. H. POLLOCK. 1987. Design and analysis methods for fish survival experiments based on release–recapture. American Fisheries Society Monograph 5.

CHAMBERLIN, T. C. 1965. The method of multiple working hypotheses. Science 148:754–759. [Reprint of 1890 paper in Science.]

CHERRY, S. 1998. Statistical tests in publications of The Wildlife Society. Wildlife Society Bulletin 26:947–953.

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

FLEISHMAN, E., R. MAC NALLY, J. P. FAY, AND D. D. MURPHY. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. Conservation Biology 15:1674–1685.

FRANKLIN, A. B., D. R. ANDERSON, AND K. P. BURNHAM. 2002. Estimation of long-term trends and variation in avian survival using random effects models. Journal of Applied Statistics 29:267–287.

FREEDMAN, D. A. 1983. A note on screening regression equations. American Statistician 37:152–155.

GELMAN, A., J. C. CARLIN, H. STERN, AND D. B. RUBIN. 1995. Bayesian data analysis. Chapman & Hall, New York, USA.

GUTHERY, F. S., J. J. LUSK, AND M. J. PETERSON. 2001. The fall of the null hypothesis: liabilities and opportunities. Journal of Wildlife Management 65:379–384.

HEYDE, C. C. 1997. Quasi-likelihood and its applications. Springer-Verlag, New York, USA.

JOHNSON, D. H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63:763–772.

KULLBACK, S., AND R. A. LEIBLER. 1951. On information and sufficiency. Annals of Mathematical Statistics 22:79–86.

LEONARD, T., AND J. S. J. HSU. 1999. Bayesian methods: an analysis for statisticians and interdisciplinary researchers. Cambridge University Press, Cambridge, United Kingdom.

LEOPOLD, B. D. 2001. Evolution: wildlife as an art to wildlife as a science. Wildlife Society Bulletin 29:769.

MCCULLAGH, P., AND J. A. NELDER. 1989. Generalized linear models. Chapman & Hall, London, United Kingdom.

MCQUARRIE, A. D. R., AND C.-L. TSAI. 1998. Regression and time series model selection. World Scientific, Singapore.

MILLER, A. J. 1990. Subset selection in regression. Chapman & Hall, London, United Kingdom.

MORRISON, M. L. 2001. A proposed research emphasis to overcome the limits of wildlife–habitat relationship studies. Journal of Wildlife Management 65:613–623.

O'CONNOR, R. 2000. Why ecology lags behind biology. Scientist 14:35.

ROYALL, R. M. 1997. Statistical evidence: a likelihood paradigm. Chapman & Hall, London, United Kingdom.

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

SELLKE, T., M. J. BAYARRI, AND J. O. BERGER. 2001. Calibration of p values for testing precise null hypotheses. American Statistician 55:62–71.

SHEFFERSON, R. P., B. K. SANDERCOCK, J. PROPER, AND S. R. BEISSINGER. 2001. Estimating dormancy and survival of a rare herbaceous perennial using mark–recapture models. Ecology 82:145–156.

SPIEGELHALTER, D. J., N. G. BEST, B. P. CARLIN, AND A. VAN DER LINDE. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B 64:1–34.

TUKEY, J. W. 1960. Conclusions and decisions. Technometrics 2:423–433.

WEDDERBURN, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 61:439–447.

YOCCOZ, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America 72:106–111.