Journal of Wildlife Management 74(6):1175–1178; 2010; DOI: 10.2193/2009-367

Commentary

Uninformative Parameters and Model Selection Using Akaike's Information Criterion

TODD W. ARNOLD,[1] *Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, MN 55108, USA*

ABSTRACT As use of Akaike's Information Criterion (AIC) for model selection has become increasingly common, so has a mistake involving interpretation of models that are within 2 AIC units (ΔAIC $\leq$ 2) of the top-supported model. Such models are <2 ΔAIC units because the penalty for one additional parameter is +2 AIC units, but model deviance is not reduced by an amount sufficient to overcome the 2-unit penalty and, hence, the additional parameter provides no net reduction in AIC. Simply put, the uninformative parameter does not explain enough variation to justify its inclusion in the model and it should not be interpreted as having any ecological effect. Models with uninformative parameters are frequently presented as being competitive in the *Journal of Wildlife Management*, including 72% of all AIC-based papers in 2008, and authors and readers need to be more aware of this problem and take appropriate steps to eliminate misinterpretation. I reviewed 5 potential solutions to this problem: 1) report all models but ignore or dismiss those with uninformative parameters, 2) use model averaging to ameliorate the effect of uninformative parameters, 3) use 95% confidence intervals to identify uninformative parameters, 4) perform all-possible subsets regression and use weight-of-evidence approaches to discriminate useful from uninformative parameters, or 5) adopt a methodological approach that allows models containing uninformative parameters to be culled from reported model sets. The first approach is preferable for small sets of a priori models, whereas the last 2 approaches should be used for large model sets or exploratory modeling.

KEY WORDS Akaike's Information Criterion (AIC), Akaike-best model, model averaging, model selection, parameter selection, uninformative parameters.

In the last decade, information-theoretic approaches have largely supplanted null hypothesis testing in the wildlife literature (Anderson and Burnham 2002, Burnham and Anderson 2002). Although this is a largely constructive paradigm shift, I nevertheless share concerns that one statistical ritual has replaced another and that comparative ranking of models now overshadows ecological interpretation of those models (Guthery et al. 2005, Chamberlain 2008, Guthery 2008). One small but incessantly common problem that contributes to this is the reporting and interpretation of models that are not truly competitive with top-ranking models, but appear competitive by virtue of low Akaike's Information Criterion (AIC) scores. This occurs whenever a variable with poor explanatory power is added to an otherwise good model and the result is a model with ΔAIC < 2, a distance widely interpreted as indicating a "substantial level of empirical support" (Burnham and Anderson 2002:170). However, this is an erroneous interpretation, and Burnham and Anderson (2002:131) found this issue important enough to put inside a text box (something they did only 29 times in 454 text pages):

> Models having Δi [ΔAIC] within about 0–2 units of the best model should be examined to see whether they differ from the best model by 1 parameter *and* have essentially the same values of the maximized log-likelihood as the best model. In this case, the larger model is not really supported or competitive, but rather is 'close' only because it adds 1 parameter and therefore will be within 2 Δ_i units, even though the fit, as measured by the log-likelihood value, is not improved.

Obviously, a similar caveat would apply to models with 2 extra parameters that fall within approximately 4 ΔAIC units of the best model, or 3 extra parameters that fall within approximately 6 ΔAIC units of the best model, distances that are often interpreted as meaningful.

A WORKED EXAMPLE

I illustrate the problem of uninformative parameters using a recently published data set on detection probabilities of breeding waterfowl pairs in North Dakota, USA (Pagano and Arnold 2009). Model selection in that study was based on AIC, which is defined as $-2\log L(\theta \| y) + 2K$, where $\log L(\theta \| y)$ is the maximized log-likelihood of the model parameters given the data and K is the number of estimable parameters (Burnham and Anderson 2002:61). For any well-supported approximating model, it is possible to add any single parameter and achieve a new model that is $\leq$2 AIC units from the well-supported model, because even if the additional parameter has no explanatory ability whatsoever (i.e., log-likelihood is unchanged), AIC will only increase by 2 due to the 1-unit increase in K. For example, Pagano and Arnold (2009, table 2) reported a 16-parameter model where detection probabilities (p) of breeding duck pairs were described by a factorial combination of 2 observers (obs) and 8 species (spp). Pagano and Arnold (2009) considered additional covariates that might affect detection probabilities and modeled these covariates to have an additive effect over both observers and all species (i.e., ΔK = 1). Effective sample size (n) for this data set was 6,162, so the small sample adjustment to AIC$_c$ of 17 versus 16 parameters is a nearly negligible 0.01. Hereafter I will use AIC and assume n/K large and overdispersion (c) negligible, but these criticisms also apply to model selection based on AIC$_c$ and QAIC$_c$, although the boundaries are no longer precisely restricted to <2 ΔAIC units, but may be somewhat larger depending on values of n/K. Based on their review of

[1] *E-mail: arnol065@umn.edu*

Table 1. Models examining effects of various covariates on detection probabilities of indicated breeding pairs of waterfowl in North Dakota, USA (from Pagano and Arnold 2009). I added single parameters assuming an additive effect to the base model, which included $K = 16$ parameters (8 species × 2 observers). Three of these covariates were considered biologically feasible (total ducks, vegetative cover, and cover type), 6 were not (random 5, 4, 8, and 1; not Sunday, Monday, or Wednesday; and last duck seen a mallard), and I excluded 6 additional nonsense or random variables (ΔAIC = 0.64–2.00) from presentation. I evaluated all models compared to the base model using Akaike's Information Criterion (AIC), ΔAIC, and changes in model deviance (Dev).

Model	AIC	ΔAIC	K	Dev
Total ducks	4,426.71	−16.62	17	4,392.71
Random 5	4,439.95	−3.38	17	4,405.95
Random 4	4,442.20	−1.13	17	4,408.19
Vegetative cover	4,442.65	−0.68	17	4,408.65
Random 8	4,442.81	−0.52	17	4,408.80
Random 1	4,442.90	−0.43	17	4,408.90
Base model	4,443.34	0	16	4,411.33
Not Sunday, Monday, or Wednesday	4,445.09	1.75	17	4,411.08
Cover type	4,445.25	1.92	17	4,411.25
Last duck seen a mallard	4,445.33	1.99	17	4,411.32

the literature, Pagano and Arnold (2009) considered 12 additional covariates that they believed might affect detection probabilities and found that 7 of them were supported by net reductions in AIC_c, whereas all 12 variables produced models that were ≤1.92 ΔAIC_c units from model p[obs × spp]. Indeed, so were 4 nonsensical variables that I considered specifically for this commentary, such as whether the last duck seen was a mallard (*Anas platyrhynchos*), whether the next duck seen was a northern pintail (*A. acuta*), whether the survey was conducted on a day that included the letter n (i.e., Sunday, Monday, or Wednesday), and log[(standardized temp/standardized wind speed)^2], plus 8 completely random variables generated using Z-distributions (Table 1; ΔAIC ≤ 2.00 for all 12 variables, with 4 of them leading to net reductions in AIC).

The ultimate objective of Pagano and Arnold (2009) was to assess whether double-observer methodologies provided enhanced prediction of breeding duck pairs. Selection of top-ranked models is only the first step in this process; biological interpretation of parameter effects is an essential second step. Total ducks had the largest influence on detection probabilities (ΔAIC = 16.62); model-based detection probabilities for mallards were 0.87 if there were no other ducks on the wetland, versus 0.75 if there were 60 other ducks on the wetland, which represents a substantial reduction in sightability, and this effect was even larger for cryptic species like ruddy ducks (*Oxyura jamaicensis*). Extent of vegetative cover on surveyed wetlands led to a much lower 0.68-unit reduction in AIC; mallards on wetlands completely ringed by tall emergent vegetation had 0.84 detection probabilities, whereas mallards on wetlands with no tall emergent vegetation had 0.86 detection probabilities, but wetlands with less than half of their perimeters surrounded by tall emergent comprised <20% of sampled wetlands. Clearly, vegetative cover could be ignored without introducing important bias, even though its effect was supported by lower AIC. But if we do include covariates such as vegetative cover, we would by the same ΔAIC criterion also include the clearly spurious random variable numbers 1, 8, 4, and 5 (Table 1). An underappreciated facet of AIC-based model selection is that it has about a 1 in 6 chance of admitting a spurious variable based on lower AIC, as opposed to a 1 in 20 chance based on traditional hypothesis testing at $\alpha = 0.05$. When sample sizes are large as in Pagano and Arnold (2009), even AIC-supported variables can have minimal biological effect (Guthery 2008). Interpreting variables that are not supported by lower AIC would further exacerbate this problem.

EXTENT OF THE PROBLEM

I reviewed all papers published in Volume 72 (2008) of the *Journal of Wildlife Management* (*JWM*) looking for evidence that authors were interpreting models that were <2 ΔAIC units from the best-approximating model and differed only in having one additional parameter. Of 60 papers that provided tables of AIC-ranked models, 43 (72%) reported hierarchically more complex models (i.e., models containing ≥1 additional parameters not found in the best model) that were <2 ΔAIC units from the top-ranking model and 35 of these 43 papers (81%) contained interpretation errors involving these additional parameters. These errors ranged from egregious (e.g., 15 papers that drew biological inference from the additional parameters), to disconcerting (e.g., 30 papers that considered these models to be competitive with the top-ranked model), to benign (e.g., 18 papers that model-averaged these models with better supported models). If using valuable journal space to summarize noncompetitive models qualifies as an error (Guthery 2008), many additional papers could have been labeled erroneous. Only 4 papers explicitly identified the additional variables as uninformative (Bentzen et al. 2008, Devries et al. 2008, Koneff et al. 2008, Odell et al. 2008) without also resorting to a criterion such as 95% confidence intervals that could have also rejected legitimate parameters.

POTENTIAL SOLUTIONS

There are ≥5 potential solutions to the 2 ΔAIC problem, and authors of 2008 *JWM* articles employed all of them, oftentimes in combination.

Full reporting.—If a truly limited set of a priori models are considered from the outset, then it probably makes sense to report and discuss all models, including those with one additional but uninformative parameter. However, the reporting should not be that these models are competitive with the higher ranked models, but rather that the additional variable(s) received little to no support, depending on the level of reduction in deviance versus the top-supported model (see also Anderson and Burnham 2002:916). For example, Odell et al. (2008) considered just 7 models to discriminate active versus inactive black-tailed prairie dog (*Cynomys ludovicianus*) colonies and although those authors included all 7 models in their table of results, they correctly ignored their second- and third-ranked models as being unsupported embellishments of their top-ranked model. Koneff et al. (2008) went one step further and devoted additional text to explain that an uninformative parameter in their analysis (group size of indicated waterfowl pairs) had no discernable effect on detection probabilities. If the a priori model set is small enough that information from all a priori models can be readily presented and authors also describe the lack of effect for uninformative parameters, then this seems like an ideal solution to the problem. Full reporting is also warranted for studies testing specific hypotheses about impacts of certain predictor variables, at least with respect to the variables of interest (e.g., Vercauteren et al. [2008] on testing efficacy of dogs at deterring deer from interacting with cattle). However, this approach becomes unworkable for model sets that are too large to justify full reporting, which includes most of the papers I reviewed.

Model averaging.—An especially common practice in *JWM* articles was to model average over all models, over all models within some cumulative weight (typically 90% or 95%), or over all models within some range of ΔAIC (typically 2, 4, or 7). One of the apparent benefits of model averaging was that it minimized the effect of uninformative parameters, particularly if coefficients for these variables were assumed to be zero in models where those variables were absent (Burnham and Anderson 2002:151–153). And if uninformative parameters are truly independent (i.e., uncorrelated with other, more useful variables), model averaging will typically have little impact on the bias and precision of the more useful parameter estimates (T. L. Shaffer, United States Geological Survey, personal communication). However, in many cases where investigators used model averaging, if models that included uninformative parameters had been ignored, the top model would have received 80–90% of model weight and there would have been little or no model-selection uncertainty. Model averaging is probably best employed as a tool to deal with legitimate model-selection uncertainty (e.g., $\geq$2 unnested models, all with substantial support) and when the primary goal is prediction rather than variable selection. Although several authors used model averaging to deal with model-selection uncertainty, I could only identify one instance where it seemed particularly useful (Saracco et al. 2008).

Confidence intervals.—Several authors discounted the importance of uninformative parameters, but only after determining that 95% confidence intervals included zero. The main problem with this solution is that it can also discard variables in best-approximating models that are supported by lower AIC values. For $n/K > 40$, AIC-based model selection will support additional variables whose approximately 85% confidence intervals exclude zero (i.e., if likelihood-ratio $\chi^2 > 2$ on 1 degree of freedom, then $P < 0.157$). It makes little sense to select variables at $P < 0.157$ using AIC and then turn around and dismiss them at $P > 0.05$ using 95% confidence intervals. A couple of authors made an important step in the right direction by using 90% confidence intervals for their parameter estimates (Hein et al. 2008, Long et al. 2008); those authors just needed to take it 5% further and use 85% confidence intervals and they would have been fully AIC compatible. If an ability to generate 85% confidence intervals were widely available in computer programs like MARK (White and Burnham 1999), then this might be a more highly favored solution. But using 95% confidence intervals with information-theoretic approaches leads to variable-selection ambivalence when β/standard error (SE)(β) = 1.4–2.0, and ambivalence is not a hallmark of good scientific writing.

Relative variable importance.—If the primary objective of modeling is to evaluate the relative importance of many potential predictor variables, such as in many habitat-selection studies, then summing Akaike model weights across all models that include that variable can be a useful approach (Burnham and Anderson 2002:167–169). When comparing summed model weights it is important that each of *j* variables be included in an equal number of models and the easiest way to achieve this is by considering all possible combinations of 2^j models (even more combinations are possible with interactions and quadratic terms). But unless all variables lead to lower AIC, this approach of considering all possible combinations will produce many models that are within 2–4 AIC units of the model with minimal AIC (i.e., ranges of ΔAIC that are frequently used as cut-offs for interpretation). But there is simply no compelling reason to put all of these models and their AIC scores into a table for publication, because finding an AIC-best model was not the objective. A table that includes a list of individual variables, their cumulative model weights, and model-averaged parameter estimates (or some other indication of biological effect size) is all that is really required (e.g., Tipton et al. 2008, table 1). However, if *j* is large, this approach misses much of the elegance of the modeling philosophy originally advocated by Burnham and Anderson (2002:147): "just because AIC was used as a selection criterion does not mean that valid inference can be expected. The primary mistake here is a common one: the failure to posit a small set of a priori models, each representing a *plausible* research hypothesis."

Discarding models with uninformative parameters.—When a sequential modeling approach is used to evaluate a large suite of potential models, as is often done in an exploratory context after first considering a more limited set of a priori models, some authors have adopted an a priori modeling approach that allows models with uninformative parameters to be discarded without further consideration. Fondell et al. (2008) adopted a hierarchical modeling approach wherein they retained only the AIC–best-ranked model from the previous step when they moved on to consider a new suite of covariates. Although models with uninformative parameters were reported at each stage (Fondell et al. 2008, table 1), they were not allowed to propagate in subsequent steps. Devries et al. (2008:1793) included an even more eloquent recognition of the problem: "Among ranked models, we considered a model to be a competitor for drawing inference if parameters in the top model were not simply a subset of those in the competing model (Burnham and Anderson 2002)." Models that failed this test were excluded from tables of competitive models, but a careful reading of the methods of Devries et al. (2008) nevertheless allows identification of all models they considered. Pagano and Arnold (2009:394) conducted an exploratory analysis of covariates affecting detection probabilities by fitting a full model that included all covariates, from which those authors "sequentially eliminated the least important covariate (as identified by minimal absolute value of *b*/SE)... If eliminating a covariate led to a reduction in AIC_c we discarded the higher order model from our model set. We continued this approach, sequentially deleting the least important covariate, until no additional covariate could be eliminated without leading to an increase in AIC_c." Models that were hierarchically more complex versions of the top model were not reported, and valuable journal space was not wasted on models that were not actually competitive, nor were these models allowed to cannibalize model weight that legitimately belonged to the hierarchically simpler model. However, critiques of sequential model fitting include its ad hoc approach and the potential for model selection bias (Burnham and Anderson 2002:43–45).

RECOMMENDATIONS

Recognition of the 2 ΔAIC problem as it applies to uninformative parameters is an important first step, but published errors still abound even though Burnham and Anderson (2002) called explicit attention to this problem (see also Anderson and Burnham 2002, Guthery et al. 2005). I reviewed 5 potential solutions to this problem, but each solution had weaknesses, and none provided a universal solution to the problem. This is actually a beneficial outcome because it requires researchers to carefully consider which approach to use and does not allow statistical ritual to replace the practice of careful thinking (Guthery 2008).

For studies employing truly limited sets of a priori models (e.g., $n \leq 10$), I recommend reporting all models and taking care to explain to readers that models with AIC scores near the top-ranked model might not be competitive as based on consideration of model deviance (Burnham and Anderson 2002:131). I also recommend full reporting of any models that represent experimental manipulations of key variables or tests of clearly articulated a priori objectives. In both cases, further discussion of parameter estimates, their uncertainty, and their biological interpretation is warranted and investigators might consider using 85% confidence intervals so that model-selection and parameter-evaluation criteria are more congruent. For exploratory approaches that involve many variables, I recommend using balanced variable sets and summed Akaike model weights if the primary goal is variable ranking and identification (Burnham and Anderson 2002:167–169) or a sequential modeling approach that allows unsupported variables to be eliminated without further reporting if the primary objective is to identify a most parsimonious model (Devries et al. 2008, Pagano and Arnold 2009). In either case, there is no need to include models with uninformative parameters in tables of model rankings. Whatever method is ultimately adopted, the primary objective should be to move beyond model ranking to model interpretation (Guthery et al. 2005), and having a smaller subset of models that are deemed to be competitive would represent a small but important step in the right direction.

ACKNOWLEDGMENTS

I thank F. S. Guthery, D. H. Johnson, P. M. Lukacs, T. L. Shaffer, and 2 anonymous reviewers for helpful comments on earlier drafts of this manuscript.

LITERATURE CITED

Anderson, D. R., and K. P. Burnham. 2002. Avoiding pitfalls when using information-theoretic methods. Journal of Wildlife Management 66:912–918.

Bentzen, R. L., A. N. Powell, and R. S. Suydam. 2008. Factors influencing nesting success of king eiders on northern Alaska's coastal plain. Journal of Wildlife Management 72:1781–1789.

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.

Chamberlain, M. J. 2008. Are we sacrificing biology for statistics? Journal of Wildlife Management 72:1057–1058.

Devries, J. H., L. M. Armstrong, R. J. MacFarlane, L. Moats, and P. T. Thoroughgood. 2008. Waterfowl nesting in fall-seeded and spring-seeded cropland in Saskatchewan. Journal of Wildlife Management 72:1790–1797.

Fondell, T. F., D. A. Miller, J. B. Grand, and R. M. Anthony. 2008. Survival of dusky Canada goose goslings in relation to weather and annual nest success. Journal of Wildlife Management 72:1614–1621.

Guthery, F. S. 2008. Statistical ritual versus knowledge accrual in wildlife science. Journal of Wildlife Management 72:1872–1875.

Guthery, F. S., L. A. Brennan, M. J. Peterson, and J. J. Lusk. 2005. Information theory in wildlife science: critique and viewpoint. Journal of Wildlife Management 69:457–465.

Hein, C. D., S. B. Castleberry, and K. V. Miller. 2008. Male Seminole bat winter roost-site selection in a managed forest. Journal of Wildlife Management 72:1756–1764.

Koneff, M. D., J. A. Royle, M. C. Otto, J. S. Wortham, and J. K. Bidwell. 2008. A double-observer method to estimate detection rate during aerial waterfowl surveys. Journal of Wildlife Management 72:1641–1649.

Long, R. A., J. L. Rachlow, and J. G. Kie. 2008. Effects of season and scale on response of elk and mule deer to habitat manipulation. Journal of Wildlife Management 72:1133–1142.

Odell, E. A., F. M. Pusateri, and G. C. White. 2008. Estimation of occupied and unoccupied black-tailed prairie dog colony acreage in Colorado. Journal of Wildlife Management 72:1311–1317.

Pagano, A. M., and T. W. Arnold. 2009. Detection probabilities for ground-based breeding waterfowl surveys. Journal of Wildlife Management 73:392–398.

Saracco, J. F., D. F. DeSante, and D. R. Kaschube. 2008. Assessing landbird monitoring programs and demographic causes of population trends. Journal of Wildlife Management 72:1665–1673.

Tipton, H. C., V. J. Dreitz, and P. F. Doherty, Jr. 2008. Occupancy of mountain plover and burrowing owl in Colorado. Journal of Wildlife Management 72:1001–1006.

Vercauteren, K. C., M. J. Lavelle, and G. E. Phillips. 2008. Livestock protection dogs for deterring deer from cattle and feed. Journal of Wildlife Management 72:1443–1448.

White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation for populations of marked animals. Bird Study 46(Supplement):120–138.

Associate Editor: Brennan.