Improving data analysis in herpetology: using Akaike's Information Criterion (AIC) to assess the strength of biological hypotheses

Marc J. Mazerolle[1]

Abstract. In ecology, researchers frequently use observational studies to explain a given pattern, such as the number of individuals in a habitat patch, with a large number of explanatory (i.e., independent) variables. To elucidate such relationships, ecologists have long relied on hypothesis testing to include or exclude variables in regression models, although the conclusions often depend on the approach used (e.g., forward, backward, stepwise selection). Though better tools have surfaced in the mid 1970's, they are still underutilized in certain fields, particularly in herpetology. This is the case of the Akaike information criterion (AIC) which is remarkably superior in model selection (i.e., variable selection) than hypothesis-based approaches. It is simple to compute and easy to understand, but more importantly, for a given data set, it provides a measure of the strength of evidence for each model that represents a plausible biological hypothesis relative to the entire set of models considered. Using this approach, one can then compute a weighted average of the estimate and standard error for any given variable of interest across all the models considered. This procedure, termed model-averaging or multimodel inference, yields precise and robust estimates. In this paper, I illustrate the use of the AIC in model selection and inference, as well as the interpretation of results analysed in this framework with two real herpetological data sets. The AIC and measures derived from it is should be routinely adopted by herpetologists.

Keywords: data analysis, estimation, hypothesis testing, model averaging, regression, significance, stepwise, variable selection.

Out with the old?

In our attempt to explain reality through experiments or observational studies, we must choose how best to answer particular questions. Though elaborating a sound design is paramount, the analysis phase of an investigation is an important, and often complex issue. Sometimes, we may strive to estimate the effect (magnitude) of a given variable on a response variable. In other instances, we wish to assess whether the effect is sufficiently important to include the variable in a model to make predictions, an issue of model selection. The latter often happens in observational studies, where a number of variables are believed to explain a given ecological process or pattern. Whereas classical techniques

such as tests of null hypotheses are well-suited for manipulative experiments, their widespread use and abuse to tackle issues such as parameter estimation and model selection only reflects the slow migration of superior techniques from the distant world of statistics into ecological disciplines.

Hypothesis testing (null hypothesis testing) indirectly addresses issues of estimation and model selection by chasing statistically significant effects (i.e., the effect is or is not significant), and provides little information on the size of the effect (Yoccoz, 1991; Cherry, 1998; Goodman, 1999). Hypothesis testing has generated schisms in some fields (Abelson, 1997; Shrout, 1997; Batanero, 2000), with very heated exchanges between devotees and "unbelievers" of this approach (in favor: Harris, 1997; Chow, 1998; Robinson and Wainer, 2002; Mogie, 2004; against: Hunter, 1997; Cherry, 1998; Goodman, 1999; Guthery et al., 2001). Regardless, hypothesis testing does not perform particularly well in model selection (e.g., variables selected by forward, backward, or stepwise ap-

1 - Centre de recherche en biologie forestière, Pavillon Abitibi-Price, Faculté de Foresterie et de Géomatique, Université Laval, Québec, Québec G1K 7P4, Canada
e-mail: mmazerolle@usgs.gov
Present address: USGS Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, Maryland 20708-4017, USA.

proaches), and alternatives do exist (Anderson et al., 2000, 2001a; Guthery et al., 2001; Johnson, 1999, 2002). Alternative approaches include Bayesian methods and model selection based on information theory (Ellison, 1996; Wade, 2000; Burnham and Anderson, 2002, 2004).

One such approach, developed in the early 1970's, rests on Akaike's information criterion (AIC) and its associated measures. This framework is also known as the information-theoretic approach, as it has arisen from information theory, a field encompassing a number of methods and theories pivotal to many of the sciences. Because information theory *per se* goes beyond the scope of the present paper, the reader should consult Kullback and Leibler (1951), Cover and Thomas (1991), and Burnham and Anderson (2002: 49) for further discussions on the issue. In ecology, the AIC and its related measures were first applied almost exclusively in the context of model selection in capture-recapture analyses (Lebreton et al., 1992; Anderson et al., 1994), but have gained popularity since the last decade in more general situations (Johnson and Omland, 2004).

The AIC is increasingly prevalent in published papers on various taxa in the leading ecological journals such as *Ecology*, *Ecological Applications*, *Oikos*, *Journal of Wildlife Management, and Journal of Applied Ecology*. However, by looking deeper into the bulk of practitioners and the taxa-oriented journals, a different picture arises. Indeed, some fields such as herpetology, still seem reticent to use these techniques. This becomes evident when comparing the proportion of papers published in the leading journals of ornithology, herpetology, and ichthyology. Indeed, based on the first two issues of 2004, 10.8% of the papers published in ornithology used this approach, whereas only 3.5% and 0.8% of papers dealt with this method in the ichthyology and herpetology journals, respectively (fig. 1). Though this review is superficial and does not discriminate between types of studies (mark-recapture or other designs),
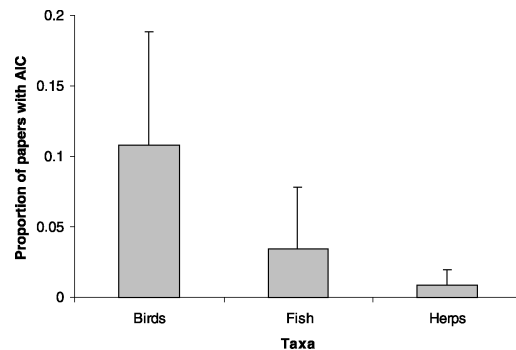
Figure 1. Mean proportion (± 1 SD) of papers dealing with the information-theoretic approach and that were published in the first two 2004 issues of the leading herpetological (*Amphibia-Reptilia*, *Copeia*, *Herpetologica*, *Herpetological Monographs*, *Journal of Herpetology*; $n = 150$ papers searched), ichthyological (*Canadian Journal of Fisheries and Aquatic Sciences*, *Ecology of Freshwater Fish*, *Fisheries Research*, *Journal of Fish Biology*, *Transactions of the American Fisheries Society*; $n = 165$ papers searched), and ornithological journals (*The Auk*, *Bird Study*, *The Condor*, *Ibis*, *Journal of Avian Biology*; $n = 183$ papers searched).

it indicates nonetheless that this approach is not taking off in herpetology. Some herpetological papers have used the approach in general journals (e.g., Van Buskirk and Arioli, 2002; Weyrauch and Grubb, 2004; Bailey et al., 2004), but most herpetologists seem unaware of these methods. In this paper, I illustrate with simple examples, the use and interpretation of information-theoretic approaches with real herpetological data sets. Whereas Schmidt and Anholt (1999) and Schmidt et al. (2002) briefly mentioned these approaches in the context of capture-recapture data, I provide a comprehensive and general illustration of the AIC and related measures for all types of data analyses.

First things first

To be most efficient in our investigations, we should keep three principles in mind when making inferences (Burnham and Anderson, 2001): 1) simplicity and parsimony, 2) several working hypotheses, and 3) strength of evidence. Simplicity and parsimony, is a concept that suggests that the simplest explanation is probably the most likely. Parsimony is a component of

model building, where we must compromise between model bias and variance. Here, bias corresponds to the difference between the estimated value and true unknown value of a parameter, whereas variance reflects the precision of these estimates; a common measure of precision is the standard error (SE) of the estimate. Thus, a model with too many variables will have low precision whereas a model with too few variables will be biased (Lehmann, 1990; Burnham and Anderson, 2002: 31).

The formulation of several working hypotheses consists in evaluating the plausibility of a series of research hypotheses (or models) with an experiment or study, then according to the results of the analyses, dropping some of the least likely hypotheses (or models) and formulating new ones to test with new data (Chamberlin, 1965). At any given time, several models are considered. Following the analyses, we then require a measure of the strength of evidence in favour of each model we considered. Information-theoretic approaches adhere in part to all three principles, which make them quite attractive.

Minimizing the loss of information

Before constructing a model (e.g., a linear regression model or any generalized linear model), we must accept that no single model yields the whole truth or complete information about the phemonena under study. Indeed, models only approximate reality. The question then is to find which model would best approximate reality given the data we have recorded. In other words, we are trying to minimize the loss of information. Kullback and Leibler (1951) addressed such issues and developed a measure, the Kullback-Leibler information, to represent the information lost when approximating reality (i.e., a good model minimizes the loss of information). A few decades later, Akaike (1973) established a relationship between the maximum likelihood, which is one of the most widespread estimation methods used in statistical

analyses, and the Kullback-Leibler information. In essence, he developed an information criterion to estimate the Kullback-Leibler information. Very useful in model selection, this criterion was later termed Akaike's information criterion (AIC), defined as

$$ \text{AIC} = -2(\log\text{-likelihood}) + 2K $$

where K is the number of estimated parameters included in the model (i.e., the number of variables $+1$, to include the intercept). Note that when the variance is estimated, such as in models using the normal distribution, it must be included in the count of parameters, K (Anderson and Burnham, 2002: 12). The log-likelihood of the model given the data, is readily available in statistical output, and reflects the overall fit of the model (smaller values indicate worse fit). The AIC penalizes for the addition of parameters, and thus selects a model that fits well but has a minimum number of parameters (i.e., the principle of simplicity and parsimony). The AIC can also be easily calculated from the output of conventional least-squares regression for normally-distributed errors (see Burnham and Anderson, 2002: 63).

Sugiura (1978) later developed the second-order Akaike Information Criterion (AIC$_c$) for small sample sizes

$$ \text{AIC}_c = -2(\log\text{-likelihood}) + 2K \\ + \frac{2K(K+1)}{(n-K-1)} $$

where n is the effective sample size. As sample size increases, the last term of the AIC$_c$ approaches zero, and the AIC$_c$ tends to yield the same conclusions as the AIC (Burnham and Anderson, 2002: 66). Thus, it is strongly recommended to routinely use the AIC$_c$. To simplify the text in this paper, I will use the term AIC as a generic term to denote all information criteria derived from it (i.e., AIC, AIC$_c$, QAIC, QAIC$_c$, see below).

Model selection

The value of the AIC for a given model is a measure of the loss of information which results from the use of the model to explain a particular variable or pattern. Though interesting, it is of little use on its own. An AIC is most useful when compared to the AIC of other models for a given data set: the model with the lowest AIC will be the "best" model among all models specified for the data set. It remains that if only poor models are considered, the AIC will select the best of the poor models. This highlights the importance of spending time to determine the set of candidate models.

The importance of *a priori* model specification

Before undertaking the analyses, one must determine the set of candidate models to consider. Each candidate model should represent a biological hypothesis. Model specification is the hardest part of the AIC framework, as it requires a lot of thought, and very few investigators have addressed this issue (but see Lehmann, 1990; Chatfield, 1991, 1995). Pertinent models may be suggested by previous investigations in similar conditions (e.g., habitat type or organism), as well as by judgement and knowledge of the system under study. One should be able to defend having included or excluded a given model for consideration. The reader is directed to Franklin et al. (2000) for a detailed account of model specification and justification. It is important to grasp that formulating models *a posteriori*, such as when following an initial round of analyses, is unsuitable. Indeed, such practices termed "data mining" or "data dredging", can lead to spurious results (Freedman, 1983; Chatfield, 1995; Anderson et al., 2001b), and if any do occur, they must be clearly outlined in the paper (Anderson et al., 2001b).

After having specified the set of plausible models to explain the data and before conducting the analyses (e.g., multiple linear regression), one should assess the fit of the global model, defined as the most complex model of the set. In other words, for a given data set, the global model is the one including all the variables of interest, whereas simpler models only contain a subset of these variables (see "An example" below). We generally assume that if the global model fits, simpler models also fit because they originate from the global model (Burnham and Anderson, 2002: 305; Cooch and White, 2001). Many model diagnostics are available such as residuals plotted against predicted values or goodness of fit statistics, and the suitability of each depends on the type of analysis (see Hosmer and Lemeshow, 1989; McCullagh and Nelder, 1989; Williams et al., 2002).

Once the appropriate transformations have been conducted (if warranted) and the global model fits the data, one can run each of the models and compute the AIC (or AIC_c). The models can then be ranked from best to worse (i.e., low to high AIC values). One should ensure that the same data set is used for each model, i.e., the same observations must be used for each analysis. Missing values for only certain variables in the data set can also lead to variations in the number of observations. Furthermore, the same response variable (y) must be used for all models (i.e., it must be identical across models, consistently with or without transformation). Nonetheless, one may specify different link functions or distributions to compare different types of models (e.g., normal, Poisson, logistic; see McCullagh and Nelder, 1989).

Comparing models

Two measures associated with the AIC can be used to compare models: the delta AIC and Akaike weights. These are easy to compute, as calculations remain the same regardless of whether the AIC or AIC_c is used, and are easy to interpret. The simplest, the delta AIC (Δ_i), is a measure of each model relative to the best

model, and is calculated as

$$\text{Delta AIC}_i = \Delta_i = \text{AIC}_i - \min \text{AIC}$$

where AIC_i is the AIC value for model i, and min AIC is the AIC value of the "best" model. As a rule of thumb, a $\Delta_i < 2$ suggests substantial evidence for the model, values between 3 and 7 indicate that the model has considerably less support, whereas a $\Delta_i > 10$ indicates that the model is very unlikely (Burnham and Anderson, 2002: 70).

Akaike weights (w_i) provide another measure of the strength of evidence for each model, and represent the ratio of the delta AIC (Δ_i) of a given model relative to the whole set of R candidate models:

$$\frac{\exp(-\Delta_i/2)}{\sum_{r=1}^{R} \exp(-\Delta_r/2)}.$$

Essentially, we are simply changing the scale of the Δ_i's to compare them on a scale of 1 (i.e., so that the sum of the w_i equals 1). The interpretation of Akaike weights (w_i) is straightforward: they indicate the probability that the model is the best among the whole set of candidate models. They are also equivalent to Bayesian posterior probabilities (Burnham and Anderson, 2002: 302; Burnham and Anderson, 2004). For instance, an Akaike weight of 0.75 for a model, indicates that given the data, it has a 75% chance of being the best one among those considered in the set of candidate models. In addition, one can compare the Akaike weights of the "best" model and competing models to determine to what extent it is better than another. These are termed evidence ratios and are calculated as

$$\text{Evidence ratio} = \frac{w_j}{w_i}$$

where model j, the "best" model, is compared against model i. For example, an evidence ratio of

$$\frac{w_j}{w_i} = \frac{0.55}{0.40} = 1.375$$

would indicate that model j is only 1.38 more likely than model i to be the best, given the set

of R candidate models and the data. This suggests that the rank of model j might change if we were to take a series of independent samples of identical size (Burnham and Anderson, 2002: 77). In other words, there would be a high degree of uncertainty regarding the best model.

We can also measure the relative importance of a variable with Akaike weights: one simply sums the w_i of the models including the variable and compares it to the sum of the w_i for the models that do not (for an example, see Burnham and Anderson, 2002: 167). However, one should only contrast the relative importance of variables when there is an equal number of models for both situations (i.e., models with variable vs without). Instead of relying on the relative importance of a variable, a superior approach consists in formally assessing the magnitude of the effect of the explanatory variable on the response variable with an estimate averaged across all models (see *multimodel inference* below).

AIC vs H$_o$ in model selection

The AIC is not a hypothesis test, does not have an α-value, and does not use notions of significance. Instead, it focuses on the strength of evidence (i.e., Δ_i and w_i), and gives a measure of uncertainty for each model. Thus, we can determine how likely a model is to be the best given the data and models at hand. In contrast, conventional (variable) model selection approaches such as backward, forward, or stepwise selection procedures are generally based on hypothesis tests, where at a certain P-value, a variable is included or excluded (Zar, 1984; Hosmer and Lemeshow, 1989; Afifi and Clark, 1996; Kleinbaum et al., 1998). These techniques can yield different conclusions depending on the order in which the models are computed, whereas the AIC approach yields consistent results, regardless of the order in which the models are computed and, more importantly, does not require that models be nested (Anderson et al., 2000, 2001b; Burnham and Anderson, 2002: 88).

Table 1. AIC_c of the multiple linear regression models of mass lost by frogs after 2 h according to substrate type, shade treatment, and weather variables. A total of 121 individuals were retained for analysis.

Model[a]	Model ID	Log-likelihood	K[b]	AIC_c	Δ_i	w_i
Shade subst mass mass2	1	−38.89	7	92.77	26.64	0.00
Shade subst mass mass2 shade*subst	2	−35.18	9	89.98	23.86	0.00
Mass mass2	3	−64.90	4	138.14	72.02	0.00
Shade mass mass2	4	−55.25	5	121.02	54.90	0.00
Subst mass mass2	5	−51.39	6	115.52	49.39	0.00
Shade subst shade*subst	6	−124.05	7	263.09	196.97	0.00
Shade subst	7	−125.24	5	261.00	194.88	0.00
Subst	8	−128.33	4	265.00	198.88	0.00
Shade	9	−131.35	3	268.91	202.78	0.00
Shade subst mass mass2 air wind cloud	10	−24.09	10	70.18	4.06	0.12
Shade subst mass mass2 air wind cloud shade*subst	11	−19.62	12	66.12	0.00	0.88
Shade subst mass mass2 air shade*subst	12	−31.07	10	84.15	18.03	0.00
Shade subst mass mass2 air	13	−34.90	8	87.08	20.95	0.00
Shade subst mass mass2 wind shade*subst	14	−29.08	10	80.16	14.04	0.00
Shade subst mass mass2 wind	15	−33.02	8	83.32	17.19	0.00
Shade subst mass mass2 cloud shade*subst	16	−28.65	10	79.31	13.18	0.00
Shade subtype mass mass2 cloud	17	−32.76	8	82.80	16.67	0.00

[a] Subst: substrate type; shade: trials shielded from the sun or not; mass: standardized initial mass of frogs; mass2: square of initial mass; air: air temperature; wind: wind velocity (low wind vs high wind), cloud: percent cloud cover.
[b] Parameter count includes intercept and variance.

An example. Now, let's illustrate the use of the AIC_c in a real data set adapted from Mazerolle and Desrochers (unpublished data) which dealt with the mass lost by frogs on substrates associated with environments having undergone anthropic disturbance. In this field experiment, we submitted 126 green frogs (*Rana clamitans melanota*) to dehydration on three different substrates (i.e., *Sphagnum* moss, soil, or peat) in or out of the shade (shade provided by an opaque tarpaulin): 21 frogs were submitted to each combination of treatments. The initial mass (g) before dehydration was measured, as well as air temperature (°C), percent cloud cover, and wind velocity (low-no wind vs strong wind) at the start of each trial. See Mazerolle (2004) for further details on the methods. The mass lost in water after 2 h was modeled with a generalized linear model with normally-distributed errors (i.e., multiple linear regression) fitted with maximum likelihood. Before the analyses, 5 cases with missing data were deleted (to avoid variations in the number of observations used in the analyses), a log transformation (base 2) was applied to the dependent

variable to homogenize variances, and the initial frog mass was standardized (i.e., values were centered around each mean and reduced by the standard deviation).

For the purpose of this example, I chose a set of 17 ($R = 17$) candidate models (table 1). Because small individuals have a greater surface/volume ratio and lose water faster than large frogs (Thorson, 1955; Schmid, 1965), I expected a curved response between the water lost to dehydration and frog initial mass. Thus, I included the square of initial mass as a covariate in models. I also expected that frog water loss on a substrate would depend on the amount of shade (shade vs no shade). To do so, I considered the substrate × shade interaction in some models. Models 1-9 excluded weather variables. I determined the effect of weather variables on mass loss after accounting for the other variables. Models 10-17 consisted of initial mass and substrate type (with or without the interaction), with one or all of the weather variables. The global model (model 11) suggested good fit, based on visual inspection of the residuals plotted against the predicted values.

The results in table 1 indicate that model 11 with an Akaike weight of 0.88 is the best given the set of 17 candidate models. Model 10, which does not include the shade $\times$ substrate interaction, follows second, but relatively far behind. Indeed, model 10 has a Δ_i of 4.06 and an Akaike weight of 0.12. Thus, model 11 is 7.3 times more likely to be the best model than model 10 (evidence ratio = 0.88/0.12), and reveals a relatively low amount of uncertainty regarding the best model. One could base his conclusions on model 11, given its very high ranking, but this practice is generally not recommended unless $w_i \geqslant 0.90$ (Burnham and Anderson, 2002: 150). In most instances, several models will compete for first place, and it will be inappropriate to base predictions on the model ranked in first place. Fortunately, as highlighted in the next section, there are ways to address the issue.

When several models compete for first place: multimodel inference

As noted above, in some instances, the "best" model may have competitors for the top rank (i.e., $\Delta_i < 2$, or equivalently, evidence ratios < 2.7). A solution to this problem is to base the inference on the entire set of models, an approach termed multimodel inference or model averaging. Indeed, instead of relying solely on the estimates of the best model, we compute a weighted average of the estimates incorporating model uncertainty. In essence, we are using all the information available from the entire set of models to make inference and it is a very elegant way of tackling the problem.

For a given parameter, the first step consists in rearranging the AIC table with the models containing the parameter of interest. Delta AIC and Akaike weights are then recomputed for this subset of the models. To conduct model averaging, the estimate (i.e., the regression coefficient) of the parameter for each model is then weighted by the Akaike weights, as follows

$$\hat{\bar{\theta}} = \sum_{i=1}^{R} w_i \hat{\theta}_i$$

where $\hat{\theta}_i$ denotes the estimate for model i.

Similarly, one computes the precision (SE) of the model-averaged estimate, termed the unconditional SE (i.e., a SE not restricted to a single "best" model but based on the whole set)

Unconditional SE
$$= \sum_{i=1}^{R} w_i \sqrt{\widehat{\mathrm{var}}(\hat{\theta}_i|g_i) + (\hat{\theta}_i - \hat{\bar{\theta}})^2}$$

where $\widehat{\mathrm{var}}(\hat{\theta}_i|g_i)$ represents the variance of the estimate $\hat{\theta}_i$ given model g_i. Note that $\widehat{\mathrm{var}}(\theta_i|g_i)$ equates to squaring the SE of $\hat{\theta}_i$. Returning to our example with dehydrated frogs, we can easily compute the model-averaged estimate of cloud cover to determine its effect on frog water loss (table 2).

In many cases, model averaging reduces bias and increases precision (Burham and Anderson, 2002). Once the model-averaged estimates and unconditional SE are calculated, we can use confidence intervals to assess the magnitude of the effect. For a 95% confidence interval,

Upper 95% confidence limit
$$= \text{estimate} + (1.96)\text{SE}$$

and

Lower 95% confidence limit
$$= \text{estimate} - (1.96)\text{SE}.$$

Narrow intervals indicate precise estimates. We conclude that the effect is different from 0 (i.e., there is a strong effect) when the confidence interval excludes 0. In our dehydrated frog example, the 95% confidence interval for the model-averaged regression estimate of cloud cover would be $(-0.513, -0.146)$. In other words, we can be 95% confident that the true estimate of cloud cover is between -0.513 and -0.146. The confidence interval excludes 0 and indicates that cloud cover has a negative effect on frog

Table 2. AIC_c and associated measures recomputed to obtain the model-averaged estimate and precision (i.e., unconditional SE) of cloud cover. Values are based on multiple linear regression models of mass lost by 121 frogs after 2 h according to substrate type, shade treatment, and weather variables.

Model[a]	Model ID	AIC_c	Δ_i	w_i	Regression estimate of cloud cover	SE of estimate of cloud cover
Shade subst mass mass2 air wind cloud	10	4.06	0.13	0.12	−0.328	0.097
Shade subst mass mass2 air wind cloud shade*subst	11	0.00	1.00	0.88	−0.330	0.093
Shade subst mass mass2 cloud shade*subst	16	13.18	0.00	0.00	−0.349	0.094
Shade subst mass mass2 cloud	17	16.67	0.00	0.00	−0.349	0.097
Model-averaged estimate					−0.330	
Unconditional SE						0.094

[a] Subst: substrate type; shade: trials shielded from the sun or not; mass: standardized initial mass of frogs; mass2: square of initial mass; air: air temperature; wind: wind velocity (low wind vs high wind), cloud: percent cloud cover.

water loss: the amount of water lost increases with decreasing cloud cover.

Special issues with count data

Discrete data (i.e., data occurring as integers) such as the number of individuals in a trap can be modeled using Poisson regression (McCullagh and Nelder, 1989; Agresti, 1996). However, it is common to encounter overdispersion in such data. In other words, data vary more than expected from data following a Poisson distribution (McCullagh and Nelder, 1989). Poisson-distributed data have a mean equal to the variance (i.e., $\mu = \sigma^2$), whereas overdispersion occurs when the mean exceeds the variance (i.e., $\mu > \sigma^2$). Overdispersion may arise due to biological phenomena such as aggregation, or may be a sign of inadequacy of a model. To detect whether overdispersion occurs in a data set subjected to Poisson regression, we can estimate the variance inflation factor, $\hat{c}$ (also called the dispersion parameter), with the ratio of the deviance over the residual degrees of freedom (McCullagh and Nelder, 1989),

$$\hat{c} = \frac{\text{Residual deviance}}{\text{Residual df}} = \frac{\chi^2 \text{ goodness-of-fit of global model}}{df}.$$

Note that the deviance of Poisson models sometimes does not approximate well the χ^2 distribution, and as an alternative, one may estimate $\hat{c}$ with the parametric bootstrap (Cooch and White, 2001). Regardless of how it is computed, if $\hat{c} = 1$, then no overdispersion occurs. If $\hat{c}$ exceeds 1, then there is indication of overdispersion; values < 1 may suggest underdispersion but often hint inadequate model structure. A model with $\hat{c} \ll 1$ or $\hat{c} \gg 4$ suggests that a Poisson model is probably not adequate in explaining the data at hand (Burnham and Anderson, 2002: 68). Alternatively, a negative binomial model could be used to account for overdispersion (McCullagh and Nelder, 1989; White and Bennetts, 1996). We can easily account for overdispersion in the AIC as follows,

$$QAIC = \frac{-2(\text{log-likelihood})}{\hat{c}} + 2K.$$

Similarly, the AIC_c can also be adjusted for overdispersion:

$$QAIC_c = QAIC + \frac{2K(K+1)}{n - K - 1}.$$

Note that $\hat{c}$ is an additional parameter to estimate. Thus, it must be included in the count of parameters. As the estimated $\hat{c}$ will vary from model to model, it is advised to use the $\hat{c}$ of the global model (i.e., the most complex model) and use it consistently for the other models.

Table 3. Candidate set of Poisson regression models for the number of metamorphosed anurans captured in minnow traps in 12 peatland and 12 upland ponds.

Model[a]	Log-likelihood	K[b]	$QAIC_c$[c]	Δ_i	w_i
Type invertpred	−12.97	4.00	33.47	0.00	0.56
Type logperimeter invertpred	−12.78	5.00	36.36	2.90	0.13
Logperimeter invertpred	−14.90	4.00	36.96	3.49	0.10
Invertpred	−16.58	3.00	37.07	3.61	0.09
Type	−17.13	3.00	38.07	4.60	0.06
Type logperimeter	−15.64	4.00	38.29	4.82	0.05
Intercept only	−21.01	2.00	42.44	8.97	0.01
Logperimeter	−20.90	3.00	44.86	11.39	0.00

[a] Type: peatland vs upland; logperimeter: log of pond perimeter; invertpred: proportion of visits with water scorpions (*Ranatra* sp.) captured.
[b] Parameter count includes intercept and $\hat{c}$.
[c] $QAIC_c$ values are based on the variance inflation factor of the global model ($\hat{c} = 1.11$).

The logic being that the most complex model will yield the best estimate for $\hat{c}$. Burnham and Anderson (2002: 305) discuss further issues in overdispersion, especially regarding model averaging and estimating $\hat{c}$ when no global model exists. Because a $\hat{c} > 1$ implies data vary more than they should according to the Poisson distribution, we must adjust the SE of the regression estimates by multiplying them by the square-root of $\hat{c}$ to reflect this variability. Note that adjusting with $\hat{c}$ can change the ranking of the models. As $\hat{c}$ increases, the $QAIC_c$ will select models with fewer parameters. Using $\hat{c}$ will also inflate the unconditional SE of the estimates and widen confidence intervals.

An example with overdispersion in Poisson regression. Consider a second example based on minnow trapping data in ponds within peatlands and upland ponds. This data set stems from Mazerolle (unpublished data, contact author to obtain data files) in which minnow traps were deployed for three consecutive nights and checked daily in 12 peatland ponds and 12 upland ponds in eastern New Brunswick, Canada, to determine the number of metamorphosed anurans at each pond each day. The number of traps in each pond was proportional to pond size: I started with two traps for the first 25 m² of pond surface and I added an additional trap each time the area doubled *sensu*

Adams et al. (1997). A number of pond attributes were measured, but we will limit ourselves to three explanatory variables for the sake of clarity in this example: pond type (peatland vs upland), pond perimeter, and presence of an invertebrate predator, water scorpions (*Ranatra* sp.), expressed as the proportion of visits during which *Ranatra* were caught in minnow traps. Because trapping effort varied between ponds, I added the log of trap nights as an offset variable. An offset variable can be used to account for data recorded as units of time, or effort, such as trap rates in our example, but does not add to the parameter count as it is included in the dependent variable (see Agresti, 1996 for more details).

Table 3 illustrates the set of 8 candidate models considered. Given the sample of 24 and to avoid overfitting (too many variables for the number of observations), I kept the models simple. These consisted in a global model consisting of pond type, log of perimeter and presence of water scorpions. Models only included main effects, as I did not anticipate interactions between these variables. I did not include the square of logperimeter because abundances were low. I added the intercept-only model to check whether it was better than more complex models. The global model indicates that there is overdispersion in the data, as the variance in-

Table 4. Model-averaging of the regression estimate of pond type based on Poisson regressions of the number of metamorphosed anurans captured in minnow traps in 12 peatland and 12 upland ponds.

Model[a]	$QAIC_c$	Δ_i	w_i	Regression estimate of type	SE of type	Adjusted SE of type[b]
Type logperimeter	38.29	4.82	0.06	-1.696	0.587	0.618
Type	38.07	4.60	0.07	-1.420	0.559	0.589
Type logperimeter invertpred	36.36	2.90	0.16	-1.188	0.618	0.651
Type invertpred	33.47	0.00	0.70	-1.348	0.560	0.590
Model-averaged estimate				-1.349		
Unconditional SE						0.611

[a] Type: peatland vs upland; logperimeter: log of pond perimeter; invertpred: proportion of visits with water scorpions (*Ranatra* sp.) captured.
[b] Value of SE mutliplied by the square-root of $\hat{c}$, where $\hat{c} = 1.11$.

flation factor $(\hat{c})$ = Residual deviance/df = 22.1115/20 = 1.11. Thus, we use the $QAIC_c$ for model selection.

Table 4 shows that the model consisting of pond type and presence of water scorpions is the best given the data, but is followed moderately closely by the global model (i.e., $\Delta_i = 2.89$). If we want to conduct model-averaging for the estimate of any variable in this analysis, we must first adjust the SE of the estimate(s) of interest for overdispersion (i.e., by multiplying each SE by the square-root of $\hat{c}$). This is illustrated for the estimate of pond type in table 4. We can then compute the model-averaged estimate and its unconditional SE as in the first example. Finally, we conclude that the number of anurans captured in minnow traps is greater in upland ponds than peatland ponds. Indeed, there is strong evidence for an effect of pond type as the 0 is excluded from the 95% confidence interval $(-2.546, -0.151)$.

Advantages and limitations of AIC

The AIC provides an objective way of determining which model among a set of models is most parsimonious. It is rigorous, founded on solid statistical principles (i.e., maximum likelihood), yet easy to calculate and interpret (but see Anderson and Burnham, 2002 for common mistakes when using the AIC approach). All the elements required to compute the AIC (i.e., log-likelihood, number of parameters, effective sample size) can be obtained from most statistical analysis software, such as SAS, R, S-PLUS, or SPSS. Some software report the AIC directly, but in certain cases, it is inaccurate (e.g., SAS proc MLR, Stafford and Strickland, 2003). Therefore, practicioners should double-check the AIC values in the output of their program of choice to ensure that they are correctly computed.

The AIC has very attractive properties. First, as discussed earlier, it is very proficient in model selection. Second, and perhaps most revolutionary, one can base inferences on the whole set of models (multimodel inference) and incorporate model uncertainty in the estimates and precision of parameters (see example above), instead of drawing conclusions on a single model. Third, concepts of statistical significance become superfluous with this approach. Thus, much is gained in objectivity. Anderson et al. (2001c) suggest using it to solve conflicts in the applied sciences. Though information theoretic approaches have been developed for larger sets of candidate models, the AIC can also be used in cases where only two models are considered (Burnham and Anderson, 2002: 184; Buckley and Beebee, 2004).

Despite all its advantages, the AIC approach is not a panacea. As in all other modeling situations, the usefulness of a model relies on the

quality of the data used to generate it. Moreover, the conclusions of a study will depend on the set of candidate models one has specified before conducting the analyses: a better model will remain unknown unless it is specified in the candidate set. Though in many situations the AIC is to be preferred over hypothesis-testing, the latter still has its place in true experiments, that are controlled, randomized, replicated, and with few explanatory variables. But even when testing hypotheses in the context of experiments, investigators should routinely report the magnitude of the effect of the variable and its precision, as this yields more information than P-values and simple declarations of significance. Providing estimates and their associated SE's greatly improves the value of a study, as these can be used subsequently in meta-analyses.

A flash in the pan or here to stay?

In conclusion, the information-theoretic approach revolving around the AIC shows great promise for various applications in ecology, conservation biology, behavioral ecology, and physiology. It is an efficient tool in model selection, for situations generated by observational studies conducted in the field, where regressions are sought to model a given pattern or process as a function of a number of explanatory variables. In addition, this approach allows one to compute estimates for variables of interest across the whole set of models (multimodel inference), which is fundamentally superior than basing inferences on a single "best" model. Straightforward in computation and interpretation, the AIC and its associated philosophy should be seriously considered by herpetologists and any other biologists faced with the task of analysing empirical data.

References

Abelson, R.P. (1997): On the surprising longevity of flogged horses: why there is a case for the significance test. Psychol. Sci. **8**: 12-15.

Adams, M.J., Richter, K.O., Leonard, W.P. (1997): Surveying and monitoring amphibians using aquatic funnel traps. In: Sampling Amphibians in Lentic Habitats: Methods and Approaches for the Pacific Northwest, pp. 47-54. Olson, D.H., Leonard, W.P., Bury, R.B., Eds, Society for Northwestern Vertebrate Biology, Olympia.

Afifi, A.A., Clark, V. (1996): Computer-aided Multivariate Analysis, 3rd edn. Chapman and Hall/CRC, New York.

Agresti, A. (1996): An Introduction to Categorical Data Analysis. Wiley, New York.

Akaike, H. (1973): Information theory as an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory, pp. 267-281. Petrov, B.N., Csaki, F., Eds, Akademiai Kiado, Budapest.

Anderson, D.R., Burnham, K.P. (2002): Avoiding pitfalls when using information-theoretic methods. J. Wildl. Manage. **66**: 912-918.

Anderson, D.R., Burnham, K.P., White, G.C. (1994): AIC model selection in overdispersed capture-recapture data. Ecology **75**: 1780-1793.

Anderson, D.R., Burnham, K.P., Thompson, W.L. (2000): Null hypothesis testing: problems, prevalence, and an alternative. J. Wildl. Manage. **64**: 912-923.

Anderson, D.R., Link, W.A., Johnson, D.H., Burnham, K.P. (2001a): Suggestions for presenting the results of data analyses. J. Wildl. Manage. **65**: 373-378.

Anderson, D.R., Burnham, K.P., Gould, W.R., Cherry, S. (2001b): Concerns about finding effects that are actually spurious. Wildl. Soc. Bull. **29**: 311-316.

Anderson, D.R., Burnham, K.P., White, G.C. (2001c): Kullback-Leibler information in resolving natural resource conflicts when definitive data exist. Wildl. Soc. Bull. **29**: 1260-1270.

Bailey, L.L., Simons, T.R., Pollock, K.H. (2004): Estimating site occupancy and species detection probability parameters for terrestrial salamanders. Ecol. Appl. **14**: 692-702.

Batanero, C. (2000): Controversies around the role of statistical tests in experimental research. Math. Think. Learn. **2**: 75-97.

Buckley, J., Beebee, T.J.C. (2004): Monitoring the conservation status of an endangered amphibian: the natterjack toad *Bufo calamita* in Britain. Anim. Conserv. **7**: 221-228.

Burnham, K.P., Anderson, D.R. (2001): Kullback-Leibler information as a basis for strong inference in ecological studies. Wildl. Res. **28**: 111-119.

Burnham, K.P., Anderson, D.R. (2002): Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd edn. Springer, New York.

Burnham, K.P., Anderson, D.R. (2004): Multimodel inference: understanding AIC and BIC in model selection. Sociol. Method. Res. **33**: 261-304.

Chamberlin, T.C. (1965): The method of multiple working hypotheses. Science **148**: 754-759 (reprint of original 1890 paper in Science **15**: 92).

Chatfield, C. (1991): Avoiding statistical pitfalls. Stat. Sci. **6**: 240-252.

Chatfield, C. (1995): Model uncertainty, data mining and statistical inference. J. Roy. Stat. Soc. A **158**: 419-466.

Cherry, S. (1998): Statistical tests in publications of the Wildlife Society. Wildl. Soc. Bull. **26**: 947-955.

Chow, S.L. (1998): Précis of statistical significance: rationale, validity, and utility. Behav. Brain Sci. **21**: 169-239.

Cooch, E., White, G. (2001): Program MARK: Analysis of Data from Marked Individuals, "A Gentle Introduction", 2nd edn. http://www.cnr.colostate.edu/~gwhite/mark/mark.html.

Cover, T.M., Thomas, J.A. (1991): Elements of Information Theory. Wiley, New York.

Ellison, A.M. (1996): An introduction to Bayesian inference for ecological research and environmental decision-making. Ecol. Appl. **6**: 1036-1046.

Franklin, A.B., Anderson, D.R., Gutiérrez, R.J., Burnham, K.P. (2000): Climate, habitat quality, and fitness in northern spotted owl populations in northwestern California. Ecol. Monogr. **70**: 539-590.

Freedman, D.A. (1983): A note on screening regression equations. Am. Stat. **37**: 152-155.

Goodman, S.N. (1999): Toward evidence-based medical statistics: the *P*-value fallacy. Ann. Intern. Med. **130**: 995-1004.

Guthery, F.S., Lusk, J.J., Peterson, M.J. (2001): The fall of the null hypothesis: liabilities and opportunities. J. Wildl. Manage. **65**: 379-384.

Harris, R.J. (1997): Significance tests have their place. Psychol. Sci. **8**: 8-11.

Hosmer, D.W., Lemeshow, S. (1989): Applied Logistic Regression. Wiley, New York.

Hunter, J.E. (1997): Needed: a ban on the significance test. Psychol. Sci. **8**: 3-7.

Johnson, D.H. (1999): The insignificance of statistical significance testing. J. Wildl. Manage. **63**: 763-772.

Johnson, D.H. (2002): The role of hypothesis testing in wildlife science. J. Wildl. Manage. **66**: 272-276.

Johnson, J.B., Omland, K.S. (2004): Model selection in ecology and evolution. Trends Ecol. Evol. **19**: 101-108.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A. (1998): Applied Regression Analysis and Other Multivariable Methods, 3rd edn. Duxbury Press, Toronto.

Kullback, S., Leibler, R.A. (1951): On information and sufficiency. Ann. Math. Stat. **22**: 79-86.

Lebreton, J.-D., Burnham, K.P., Clobert, J., Anderson, D.R. (1992): Modeling survival and testing biological hypotheses using marked animals: a unified approach with case-studies. Ecol. Monogr. **62**: 67-118.

Lehmann, E.L. (1990): Model specification: the views of Fisher and Neyman, and later developments. Stat. Sci. **5**: 160-168.

Mazerolle, M.J. (2004): Mouvements et reproduction des amphibiens en tourbières perturbées. Unpubl. Ph. D. thesis, Université Laval, Québec.

McCullagh, P., Nelder, J.A. (1989): Generalized Linear Models, 2nd edn. Chapman and Hall, New York.

Mogie, M. (2004): In support of null hypothesis significance testing. Proc. Roy. Soc. Lond. B **271(Suppl)**: S82-S84.

Robinson, D.H., Wainer, H. (2002): On the past and future of null hypothesis significance testing. J. Wildl. Manage. **66**: 263-271.

Schmid, W.D. (1965): Some aspects of the water economies of nine species of amphibians. Ecology **46**: 261-269.

Schmidt, B.R., Anholt, B.R. (1999): Analysis of survival probabilities of female common toads, *Bufo bufo*. Amphibia-Reptilia **20**: 97-108.

Schmidt, B.R., Schaub, M., Anholt, B.R. (2002): Why you should use capture-recapture methods when estimating survival and breeding probabilities: on bias, temporary emigration, overdispersion, and common toads. Amphibia-Reptilia **23**: 375-388.

Shrout, P.E. (1997): Should significance tests be banned? Introduction to a special section exploring the pros and cons. Psychol. Sci. **8**: 1-2.

Stafford, J.D., Strickland, B.K. (2003): Potential inconsistencies when computing Akaike's Information Criterion. Bull. Ecol. Soc. Am. **82**: 68-69.

Sugiura, N. (1978): Further analysis of the data by Akaike's information criterion and the finite corrections. Comm. Statist. Theory Methods **A7**: 13-26.

Thorson, T.B. (1955): The relationship of water economy to terrestrialism in amphibians. Ecology **36**: 100-116.

Van Buskirk, J., Arioli, M. (2002): Dosage response of an induced defense: how sensitive are tadpoles to predation risk? Ecology **83**: 1580-1585.

Wade, P.R. (2000): Bayesian methods in conservation biology. Conserv. Biol. **14**: 1308-1316.

Weyrauch, S.L., Grubb, T.C., Jr. (2004): Patch and landscape characteristics associated with the distribution of woodland amphibians in an agricultural fragmented landscape: an information-theoretic approach. Biol. Conserv. **115**: 443-450.

White, G.C., Bennetts, R.E. (1996): Analysis of frequency count data using the negative binomial distribution. Ecology **77**: 2549-2557.

Williams, B.K., Nichols, J.D., Conroy, M.J. (2002): Analysis and Management of Animal Populations. Academic Press, New York.

Yoccoz, N.G. (1991): Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bull. Ecol. Soc. Am. **72**: 106-111.

Zar, J.H. (1984): Biostatistical Analysis, 2nd edn. Prentice Hall, Englewood Cliffs.