REVIEW

Information-theoretic approaches to statistical analysis in behavioural ecology: an introduction

László Zsolt Garamszegi

Abstract Scientific thinking may require the consideration of multiple hypotheses, which often call for complex statistical models at the level of data analysis. The aim of this introduction is to provide a brief overview on how competing hypotheses are evaluated statistically in behavioural ecological studies and to offer potentially fruitful avenues for future methodological developments. Complex models have traditionally been treated by model selection approaches using threshold-based removal of terms, i.e. stepwise selection. A recently introduced method for model selection applies an information-theoretic (IT) approach, which simultaneously evaluates hypotheses by balancing between model complexity and goodness of fit. The IT method has been increasingly propagated in the field of ecology, while a literature survey shows that its spread in behavioural ecology has been much slower, and model simplification using stepwise selection is still more wide-spread than IT-based model selection. Why has the use of IT methods in behavioural ecology lagged behind other disciplines? This special issue examines the suitability of the IT method for analysing data with multiple predictors, which researchers encounter in our field. The volume brings together different viewpoints to aid behavioural ecologists in understanding the method, with the hope of enhancing the statistical integration of our discipline.

Communicated by P. Bednekoff

This contribution is part of the Special Issue "Model selection, multimodel inference and information-theoretic approaches in behavioural ecology".

L. Z. Garamszegi (✉)
Department of Evolutionary Ecology,
Estación Biológica de Doñana–CSIC,
c/Americo Vespucio, s/n,
41092 Seville, Spain
e-mail: laszlo.garamszegi@ebd.csic.es

Hypothesis-driven scientific thinking leads to predictions concerning the relationship between at least two variables (Popper 1963: 3–42). Process in ecology often involves multiple factors, each of which can also interact in complicated ways (Quinn and Dunham 1983). This is particularly true for behavioural ecology because behaviour is a complex and flexible trait that is mediated by several intrinsic and extrinsic factors (Krebs and Davies 1984: 15–145; Owens 2006). For example, a behavioural display can be altered by metabolic condition, hormonal state, parasite load, age and experience, territory quality, the time of the day, temperature, presence and identity of other individuals, reproductive investment in the past, or survival prospects. Moreover, these factors can interact, as the effect of one factor can depend on the effect of another. When such multi-causal venues are investigated, the key interest is to identify predictors in the most appropriate combination that are responsible for the biological phenomenon at hand (Chamberlin 1890; Platt 1964). This task often requires statistical approaches that can handle data with multiple predictors along a multi-hypothetical framework (Johnson and Omland 2004).

In this framework, competing hypotheses are evaluated in the form of statistical models that describe the variation in a response variable (y) as a function of one or more predictors ($x_1...x_n$), which are combined by different parameters (intercepts and slopes) (Hilborn and Mangel 1997: 12–38; Crawley 2007: 323–386). These models can include interaction terms and non-linear functions, of which several combinations can be defined mathematically.

However, only those models should be considered for further inference that represent biological mechanisms and are comprehensive to the human observer. Supposing that data on the response of interest as well as on factors that are hypothesised to influence that response have been gathered, statistical procedures can be applied to gauge support for each model, which can further be used to evaluate how the various models might explain the observed patterns of variation. This process also provides estimates for model parameters and their precision, which can be used to make inferences about the importance of biological effects and predict future outcomes.

Finding a support for a hypothesis that best approximates the reality calls for the statistical problem of model selection, in which the observer seeks the most appropriate model(s) for the data (McQuarrie and Tsai 1998: 1–2; Forster 2000; Zucchini 2000; Burnham and Anderson 2002: 13–14; Johnson and Omland 2004). Model selection is the task of choosing a model from a set of candidate models, which leads to a model of high predictive value at a minimal complexity for the given (finite) data. Reaching optimal complexity is integral in this process because making inferences from simple models with too few parameters is subject to bias due to the lack of information in the model (underfitting), while complex models with many parameters may be of poor precision due to the variability in the variables (overfitting) (McQuarrie and Tsai 1998: 8–9; Burnham and Anderson 2002: 31–35). Therefore, model selection approaches aim at achieving, from a biologically relevant initial set, the most parsimonious model with fewer variables that have greater generality (Ginzburg and Jensen 2004). Various procedures have been developed for this purpose, which rely on different strategies for model comparison and use different model selection criteria to estimate model fit relative to complexity (Forster 2000; Claeskens and Hjort 2008; Ward 2008). These model selection methods include, for example, the Mallows' Cp method (Mallows 1973), the Bayes information criterion (Schwarz 1978), the final prediction error method (Shibata 1981), the generalised information criterion (Rao and Wu 1989) and its analogues (Pötscher 1989), the delete-one cross-validation (Allen 1974; Stone 1974), the generalised cross-validation (Craven and Wahba 1979), minimum description length (Rissanen 1978), structural risk minimization (Vapnik and Chervonenkis 1974), and Takeuchi's information criterion (Takeuchi 1976), just to name some from the diverse list. One of these methods, the information-theoretic approach (IT) based on Akaike information criterion (AIC) (Akaike 1973; Burnham and Anderson 2002), has been receiving considerable attention in ecology and evolutionary biology (Johnson and Omland 2004; Rushton et al. 2004; Stephens et al. 2007a), and this special issue and this introduction are to examine its

relevance for behavioural ecology in association with prediction and parameter estimation.

The principles

The strategy that is followed in the AIC-based model selection exercise is a simultaneous assessment of competing biological hypotheses, in which these hypotheses are represented in the form of statistical models. Such a model evaluation begins with an initial set of models that reflect the list of the competing hypotheses that are defined a priori as separate models (Burnham and Anderson 2002: 15–19). All models considered are only abstractions of the unknown truth (Burnham and Anderson 2002: 15–19). Therefore, a certain amount of information is necessarily lost when the model is used to reflect the truth (Burnham and Anderson 2002: 20–22). The IT approach derives from the Kullback–Leibler (K-L) information, which is the *absolute* distance of models from the biological truth (Burnham and Anderson 2002: 50–57). The K-L information cannot be directly estimated, but the interest from the modeller's perspective is to find the model from a meaningful candidate set that loses the least information about full reality. Therefore, the *relative* distances between models are at the heart of the IT approach (Burnham and Anderson 2002: 60–64). The method provides a quantitative expression of these distances via AIC (Akaike 1973), or via its corrected version for small sample sizes (AICc, Sugiura 1978), or for overdispersed data (QAIC, Lebreton et al. 1992). AIC (or its variant) for each model can be mathematically computed based on how they support the data while correcting for biases that arise from model complexity. The model with the smallest criterion value represents the best fit for the predictive information in the data while ignoring the noise in the data, which is implied as the best approximation of the truth in the K-L sense. Note that given that AIC is always relative, the approximation of reality might be very poor if distances between completely meaningless models are compared. The approach always provides a "best" model with the smallest AIC, but if the initial models are all meaningless, the selected model for the data will be of little, if any, biological relevance (i.e. may explain little variation in the data).

Further inferences can be made based on the rank of models and the associated AIC values (Burnham and Anderson 2002: 60–64). Model selection will always provide an estimated best model with the lowest AIC, and it may be reasonable to focus on this model alone if other models lack support and if all models considered initially were of biological relevance. This approach philosophically implies binary thinking, in which one model is retained and the others are all rejected. However, in most of the cases, there are more than one model that receives more or

less equal support (model selection uncertainty), which requires measures of evidence on a continuous scale and making interpretations from multiple models while adjusting for their supportive value (multimodel inference) (Burnham and Anderson 2002: 149–202). The strength of the AIC–IT approach is that it offers various ways to deal with model selection uncertainty in the form Δ values, model likelihoods, Akaike weights, and evidence ratio (see Burnham and Anderson 2002: 70–79; Burnham et al. 2010; Symonds and Moussalli 2010). These continuous measures can be used to judge scientific evidence quantitatively (instead of the qualitative binary thinking). Moreover, inferences from a confidence set of equally supported models can be made by means of model averaging, in which particular estimates (and their precision) from the selected models are weighted according to their model probabilities (Burnham and Anderson 2002: 150–152). Therefore, the IT method accepts that uncertainty is an inherent feature of biological data, and brings the plausibility of different hypotheses in the focus in cases when there are at least two models to be compared. Practitioners in ecology are in favour of the IT methodology for model selection problems, because it is suitable for testing several competing hypotheses in concert, and is able to deal with model selection uncertainty and to provide quantitative scientific results based on data support (Johnson and Omland 2004; Hobbs and Hilborn 2006; Stephens et al. 2007a).

## Multiple hypotheses and the use of IT approaches in behavioural ecology

Model selection should be expected to be a widespread task in behavioural ecology because multiple hypotheses might be necessary to envisage for the interpretation of confounding effects and alternative mechanisms that can also mediate an association between the investigated traits (Towner and Luttbeg 2007; Cézilly et al. 2008; Garamszegi et al. 2009). Thinking along more than one hypothetical scenario is especially important when analysing observational data, as correlations cannot be used to reveal causality (Stephens et al. 2007b). Behavioural ecologists are often constrained to use such data for at least two reasons. First, many behaviours usually rely on a complex physiological machinery, which thus are difficult to manipulate without affecting an entire suite of unwanted phenotypic features (the "hidden treatment" problem, see Huston 1997). Second, certain evolutionary questions are investigated in an interspecific comparative context, in which species-specific traits can be considered non-manipulable because the scale of the question defies extrapolation from small-scale experimentation (see Quinn and Keough 2002: 6 on the spatial and temporal constraints of experimentation). However, even the

analysis of experimental data may require the consideration of more than hypothesis because some variables cannot fully be controlled by the study design (Huston 1997; Guthery 2007). For example, the effect of a treatment may depend on some factors (age, individual quality, mating status, time, and place of origin) that entail appropriate statistical control if they cannot be randomised (treatment-by-covariate interactions). Consequently, open-minded research thinking can call for model selection approaches in the face of competing research hypotheses in both experimental and correlational studies (Garamszegi et al. 2009). A strict focus on a single causal link may be deceptive, if there is no one-to-one relation between two variables, and thus a large proportion of biological reality remains undiscovered.

To estimate the frequency of use of methods for evaluating multiple hypotheses in our field, I screened papers published in three leading behavioural journals (*Animal Behaviour*, *Behavioral Ecology*, and *Behavioral Ecology and Sociobiology*) in 2003 and 2008, relying on at least 50 studies in each sample. I categorised studies according to whether they used any model selection procedure (i.e. intended to choose a model or models with the highest predictive value from a larger set of models) or simply made inferences from a single model.

I found that only 21% of studies (68 out of 329) opted to evaluate more than one statistical model, while 79% (261 out of 329) made inferences from a single model (Fig. 1). The frequency of the use of model selection methods has slightly increased over the last 5 years (likelihood ratio chi-square test: $\chi^2=3.76$, $df=1$, $P=0.052$, Fig. 1a; note that although for simplicity, I provide statistical results based on null hypothesis testing, different hypotheses about the sources of variation in the data could be evaluated by using IT-based model selection) and varies significantly across journals (likelihood ratio chi-square test: $\chi^2=13.47$, $df=2$, $P=0.001$, Fig. 1b). Observational studies were more likely to use model selections methods (likelihood ratio chi-square test: $\chi^2=10.10$, $df=1$, $P=0.002$, Fig. 1c), perhaps because these studies are more likely to evaluate multiple hypotheses. This pattern may indicate the researchers' preference for simple statistics when relying on experimental design. However, it is clear that the multimodel framework is not restricted to the analysis of correlative data, as more than 10% of experimental studies applied a complex statistical modelling of multiple hypotheses.

Focusing on the proportion of studies that relied on multimodel approaches, I also assessed which kind of model selection method has been used the most often. The most frequent approach for model selection was the stepwise removal of non-supported terms, or the approach in which only non-significant interactions (or only some of the variables) were removed at one step. The stepwise or single removal of terms are used three to four more
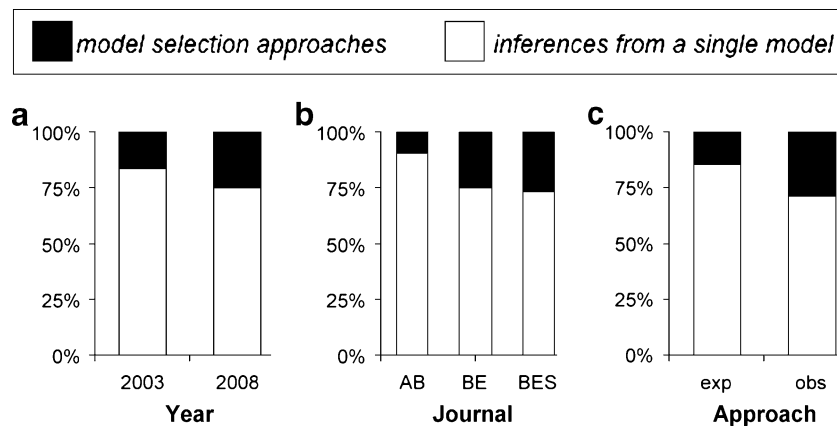
Fig. 1 The proportion of studies in behavioural ecology that dealt with multi-causal hypotheses by applying complex modelling routines in terms of model selection. Results from a literature survey that screened data-driven papers in three journals (*AB*, *Animal Behaviour*; *BE*, *Behavioral Ecology*; *BES*, *Behavioral Ecology and Sociobiology*) published in 2003 and 2008. At least 50 papers were included for each journal in each year. **a** The change of research practice for handling multiple hypotheses over 5 years. **b** Differences in the application of complex modelling between journals. **c** The frequency of multimodel approaches in experimental (*exp*) and observational (*obs*) studies

commonly than IT methods, a trend that is consistent across years, journals, and experimental vs. observational studies (likelihood ratio chi-square tests, year: $\chi^2=0.95$, $df=2$, $P=0.623$; journal: $\chi^2=6.28$, $df=4$, $P=0.180$; approach: $\chi^2=2.76$, $df=2$, $P=0.251$; Fig. 2).

In ecology and evolutionary biology, IT-based thinking serves a basis for about 20% of papers published (Johnson and Omland 2004; Guthery et al. 2005; Stephens et al. 2007a; Richards et al. 2010). A similar statistics for behavioural ecology shows that when all studies are taken together, only 6% of the recently published material in the three screened journals employed the IT approach. These figures suggest that research design along multiple hypotheses and models, and the use of IT-based model selection approaches has not been explored as fully in our field as in related disciplines.

The most common practices applied to handle multiple hypotheses in behavioral ecology

The above statistics indicate that the two preferred approaches in our discipline are the full-model approach and the stepwise model selection. The large proportion of studies do not discriminate more than one biological hypothesis, at least at the statistical level. I suspect that behavioural ecologists prefer this full-model approach over model selection because they either investigate frequently one-to-one causal relationship, or they load the competitive hypotheses within a single statistical model. The one-factorial modelling is addressed in many simple statistics such as *t* tests, ANOVAs, correlations, and linear regressions. More complex models without further simplification are employed in multiple regressions, partial correlations,
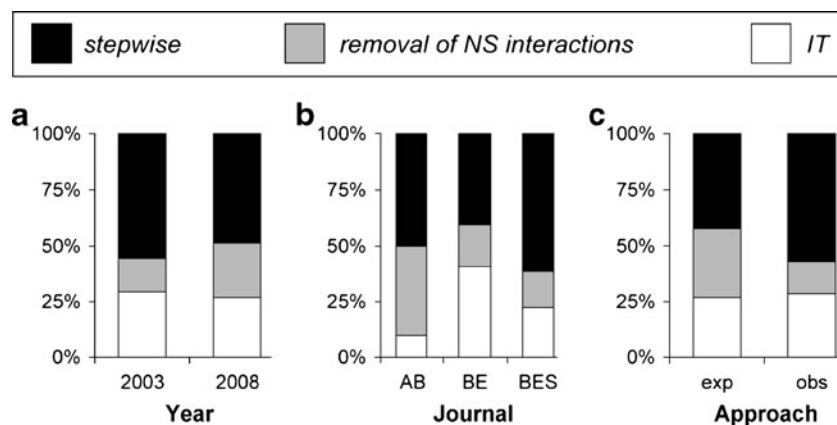
Fig. 2 The use of different approaches for multimodel inference based on model selection, as revealed by a literature survey of three journals in behavioural ecology. Figures reflect the distribution of approaches within studies that relied on the consideration of multiple models (*black columns* in Fig. 1). The differences between **a** years, **b** journals (*AB*, *Animal Behaviour*; *BE*, *Behavioral Ecology*; *BES*, *Behavioral Ecology and Sociobiology*), and **c** data types (*exp* experimental, *obs* observational)

ANCOVAs, or two-way ANOVAs or other forms of GLMs. The interest in these complex models is to test several predictors, when there is a biological reason for including all of them and their combinations within a single model. This modelling strategy aims to eliminate the possibility of variable effects that can result in a spurious relationship between the variable(s) of interest (Sokal and Rohlf 1995: 617–630). Such a design shifts the focus from the evaluation of multiple hypotheses to the control for potentially confounding variables. More discussion on the full-model approach can be found in Forstmeier and Schielzeth (2010).

In cases when authors do apply model selection methods, these are usually based on stepwise or step-down techniques, which follow a strategy that is philosophically different from that of the IT method (Draper and Smith 1981: 327–347; Derksen and Keselman 1992; Whittingham et al. 2006; Mundry and Nunn 2008). Particularly, they aim to establish one minimum adequate model that includes only those variables whose effects are supported by the data at hand. These procedures start from the full model that contains all terms considered then test several models successively (and not simultaneously), and assess the degree to which those models satisfy a certain criterion that determines data support. Simplification of the full model relies on a process in which the effect of each term is examined in turn. If the inclusion of a term is not verified by fulfilling a chosen statistical criterion (e.g. statistical significance), it is treated as having no explanatory role and thus removed from the model under the consideration that this removal does not lead to the loss of information. Note that this process can test nested models only, which are the simplified derivates of a complex model, and can fail to retain influential variables if sample size is low. Furthermore, the stepwise method does not require the a priori definition of these competing models, but instead relies on those that are visited automatically via the stepwise routine. On the contrary, the IT method can potentially evaluate the full-model space that also include non-nested model, while requiring hard thinking about the biological relevance of the candidate models (Burnham and Anderson 2002: 15–19). The stepwise model simplification traditionally relies upon a predefined level of statistical significance (e.g. $P < 0.05$) as a criterion for the exclusion or inclusion of terms, but some users apply criteria other than probability levels of significance (even AIC, against which, see warnings in Whittingham et al. 2006). Recent discussions bemoan the widespread use of stepwise approaches due to several reasons, including biased parameter estimation, inconsistencies resulting from different iteration methods and criteria, problems caused by multiple hypothesis testing, the inability to evaluate non-nested models, and the strong reliance on a single final model (Whittingham et

al. 2006; Mundry and Nunn 2008). However, based on the comparison of the predictive ability of seven model selection methods applied to real ecological datasets, Murtaugh argued that stepwise methods performed similarly to other selection algorithms (Murtaugh 2009).

Why is the IT approach not spreading in behavioural ecology?

I propose that we can understand the restricted use of IT-based inferences in our field as follows.

First, behavioural ecologists are generally unfamiliar with the new method and unaware of the benefits of the IT-based model selection (i.e. they do not know about it). Due to our general research interest, practical constraints, and statistical knowledge, the statistical integration of our community might be a generally slow process compared with other disciplines. Behavioural ecologists had especially little chance to obtain experience with the IT-based new method because it requires the understanding of a new philosophy that departs from the standard statistical training based on significance. For the general reader in our field, it might be difficult to consult with the highly technical primary literature on model selection methods. Moreover, the commonly used statistical platforms widely support the routine application of classical methods, such as stepwise regressions, while enhanced model selection methods including model averaging are not fully implemented.

Second, researchers do know what the IT method offers, but they prefer using simple statistics and traditional model selection approaches in most cases because these are believed to be more appropriate for the data and questions prevalent in behavioural ecology (i.e. they reject it). I have witnessed colleagues arguing that there are certain attributes specific to data in behavioural ecology that prohibit the application of the IT approach on the conceptual basis. For example, it is a common belief that simple causal mechanisms and elegantly performed experiments require analyses with simple statistical designs (Fig. 1c) (see also Stephens et al. 2007b). However, the IT method does not preclude experimentation, and vice versa, as the IT framework is more relevant to interpretation rather than the study design (Garamszegi et al. 2009). Richards et al. (2010) and Burnham et al. (2010) in this volume show how data from an experiment can be analysed with the IT method. Other attributes potentially preventing the use of the IT approach in association with behavioural data may include sample size issues. Although sample size affects science in general, IT modelling is sometimes thought to be sensitive to such restrictions. The number of models in the candidate set should be substantially lower than sample size (Burnham and Anderson 2002: 15–19), but this condition

might be easily violated if data on behaviour are limited, while several variables are considered to predict them during the modelling. Another reasoning I heard against IT-based model selection methods is that some statistical designs frequently used to analyse data in behavioural ecology may not be compatible with this framework (e.g. random effect models: Vaida and Blanchard 2005; phylogenetic problems: Rabosky 2006; Ripplinger and Sullivan 2008). While it may be true that we use special statistical procedures for our specific biological questions that have not been amalgamated with the IT framework, this may reflect underdevelopment rather than incompatibility. Note that certain statistical approaches we use in our field have been expanded with the IT-based model selection tool (mixed models: Fox 2002: Appendix; Liang et al. 2008; social network analysis: Whiteheat 2007; meta-regression: Jones et al. 2009; phylogenetic meta-analysis: Lajeunesse 2009).

A primer for statistical integration

In my view of the ideal, objective statistical choice requires that the investigator understands the pros and cons of all methods available for the question and data at hand. Moreover, the investigator needs to think and determine whether statistical machination yield useful information at all. In the real world, average behavioural ecologists might rarely have the background to assess the mathematical properties of the various model selection methods. Given the importance of the multiple hypothesis framework in our research, the careful use of the IT method could, in theory, offer benefits for the analysis of behavioural data with multiple predictors. Currently, according to my impression, these benefits are either completely unknown or misinterpreted, or remain unexplored. Therefore, the IT tool needs better integration within the field of behavioural ecology. This integration process involves statistical training at first, and the critical assessment of the performance of the method in association with data and design encountered in our field of research at second.

The goal of this special issue of *Behavioral Ecology and Sociobiology* is to provide an overview with illustrative examples on the properties of the IT approach that will hopefully enhance this assimilation process. There are several books available on model selection methods in general and on AIC–IT-based approaches in particular that can serve as a primary source for the detailed statistical background (Linhart and Zucchini 1986; Sakamoto 1991; Royall 1997; McQuarrie and Tsai 1998; Burnham and Anderson 2002; Massart 2007; Claeskens and Hjort 2008; Konishi and Kitagawa 2008). As this foundational literature is highly technical and may be hard to consume for the targeted readership, this volume collects papers that provide

an introductory material for the subject without requiring an expertise in mathematical statistics. Contributing papers visit some characteristics that are particularly relevant from the perspective of our research field. The majority of the authors are trained ecologists (and not statisticians), who have been extensively exposed to model selection problems in their research practice, and obtained considerable experience with the IT method. Hence, the scope of this issue offers a non-mathematical focus that might help practicing biologists obtain some familiarity with the IT method. Below, I highlight the main issues emerging from this discussion about model selection in behavioural ecology.

The introductory paper (Symonds and Moussalli 2010) and the subsequent chapter by the most influential advocates of the IT approach (Burnham et al. 2010) offer a useful guide to the application of basic concepts of IT and AIC for model selection and model averaging, which are necessary to understand before proceeding to the more advanced topics. From these opening contributions, beginners can obtain familiarity with the calculation of AIC, AICc, QAIC, model probabilities and likelihoods, as well as evidence ratio, and can learn how to and when to use them for scientific interpretation. These papers also show the principles of multimodel inference and model averaging, and describe how parameters can be estimated by incorporating model selection uncertainty. These chapters also provide some example analyses that illustrate how the IT tools perform in practice with problems and data from behavioural ecology. For more detailed elementary discussions, I suggest consulting Anderson (2008).

Hard thinking is at the heart of the multiple hypothesis approach because competing models should be considered and defined prior to data collection and analysis (Burnham and Anderson 2002: 15–19). A comparison between models makes sense only if the model space described by the candidate set covers the essentials of the causal network. IT-based model selection will always rank models, but the best model—and even the model averaged parameters—can be misleading if the original batch of models include only models with no biological relevance. This is the user's responsibility to define biologically relevant models. Burnham et al. (2010) highlight the importance of deriving alternative hypotheses based on hard thinking and representing these as probability models. In a similar vein, Dochtermann and Jenkins (2010) claim that the construction of an initial model set that will be subject to subsequent model selection is one of the more challenging aspects of multiple hypothesis-driven research. Others (Hegyi and Garamszegi 2010) emphasise that interactions between numerous variables in complex models are often cognitively intractable because biological questions and incomplete information in our field necessitate the consideration of several variables in almost all potential

combinations. Using extra-pair paternity in birds as a model, Burnham et al. (2010) give a hypothetical example and demonstrate how existing knowledge and information on the natural history of the study species can be used for the definition of the initial model set. Yet, another study points to that uncertainties during model construction can raise instabilities in model selection results and affect the precision of parameter estimates from model averaging (Hegyi and Garamszegi 2010). Dochtermann and Jenkins (2010) formulate three approaches for developing suites of hypotheses: (1) based on exploratory analyses of independent datasets, (2) based on previous research on related species, and (3) based on theory. Although none of these solutions fully guarantees that all biologically relevant interactions and causal links are included in the hypothesis set, there are ways to assess how much variance different models can explain (Dochtermann and Jenkins 2010). These diagnostics can be used to make inferences about the information content that the initial model set covers. Forstmeier and Schielzeth (2010) advocate the stronger reliance on the full model when sample size relative to the number of predictors is large enough to allow reliable parameter estimation. Under such conditions, full models can offer an objective consideration of predictors and can also avoid cryptic multiple hypothesis testing.

The IT approach is often treated as an alternative to conventional null hypothesis testing (NHT) (Anderson et al. 2000; Stephens et al. 2005; Steidl 2006; Lukacs et al. 2007). The distinction between these methods principally lies on philosophical and not mathematical grounds. In other words, NHT and IT approaches differ with regard to how inference is drawn, but not with regard to computation and model fitting (Mundry 2010). While NHT focuses on the probability of data given the null hypothesis, the IT approach measures the strength of evidence for all competing hypotheses, given the data (Burnham et al. 2010; Richards et al. 2010). Accordingly, not only multiple hypotheses testing, but also simple problems are viewed from a different perspective in an IT-based analysis. Most important, experimental data could be used to make inferences about the plausibility of the alternative hypothesis relative to the null hypothesis in the IT context. Computations of t tests or ANOVA tables include residual sums of squares that can be translated into AIC, which can be further used in multimodel inference (Burnham et al. 2010), and model averaged effect sizes can be obtained for parameter estimations that takes the plausibility of hypotheses into account (Richards et al. 2010 following the recommendations of Lukacs et al. 2007). On the other hand, NHT can only reject the null hypothesis (while the alternative hypotheses remain untested), and thus, it forces the researcher to make binary decisions and obtain parameter estimates under two extreme scenarios. Given

the contrast between IT and NHT, it is not surprising that there is a disagreement between practitioners that support one or the other approach. Behavioural ecologists can learn from this discussion however, and this special issue includes viewpoints from both sides of the statistical debate. I personally noticed that both methods are often accused mistakenly to perform badly, while they are just misused or misinterpreted (see also Stephens et al. 2007a). Statistical procedures can always provide wrong results if they are used wrongly. Therefore, the key issue is to understand these methods better, so we can learn to interpret the results more correctly.

Some emphasise that the IT-based method may have some limitations under certain circumstances, and its practical application is subject to some common misunderstandings (Anderson and Burnham 2002; Guthery et al. 2005; Steidl 2006). These pitfalls and practical mistakes include the difficulty in the definition of the models and variables in the candidate set, the temptation for data dredging (a fishing exercise for explanatory variables based on a large number of models without a priori theoretical foundation), the focus on the best models without considering their explanatory value and the *relative* nature of AIC (i.e. the lowest AIC model from poor candidate set is meaningless), mistreatment of interactions, disregarded errors in parameter estimation, imbalance between the number of cases and number of variables or models, untested assumptions about collinearity, subjective thresholds used for model comparison criteria, the sensitivity of AIC to overfitting, overparameterization, and a mixture of NHT and IT philosophy in a single analysis (Hegyi and Garamszegi 2010; Mundry 2010). Practical mistakes appear because the method is wrongly applied or the results are incorrectly interpreted, not necessarily because the method performs poorly. Any shortcoming needs to be understood and considered for an objective statistical practice, and they should also delineate the direction of future developments.

For the fruitful application of the IT method, it would also be important to examine its performance in statistical situations that researchers meet in behavioural ecology, and to develop features specific to their needs. These tasks include tests of the sensitivity of different model selection criteria and model averaging to the presence of interactions and variance–covariance structures specific to behavioural data and experimental approach, and to combine IT-based thinking with tools that are often used in behavioural ecology.

To this end, Richards et al. (2010) assess the value of IT model selection using AIC in an experimental context. Their simulations show that for simple analyses (e.g. comparisons between control and experimental groups in a t test design), the advantage of the IT approach over tests of significance is strictly philosophical, as it puts higher emphasis on biological rather than statistical significance.

However, when applied to problems of more complex causality, the model selection approach of the IT framework offers substantial enhancements in terms of the ability to handle data that violate the assumptions of the NHT (e.g. non-linear relations, non-normal distributions, and hetero-scedasticity) and to detect appropriate error structures. However, this piece failed to demonstrate in all experimental situations and sample size conditions that model averaging provides unanimously consistent improvements in terms of prediction accuracy over the best AIC model approach, when prediction accuracy was estimated as the predictive errors provided by the two approaches. This is somewhat surprising because extensive simulations of correlative patterns reveal the superiority of model averaging (Burnham and Anderson 2002: 298–301). Consequently, Richards et al. (2010) recommend that the model averaging tool might require more tests in association with experimental data.

Following a similar motivation, Freckleton (2010) examines two under-appreciated issues of major relevance in behavioural sciences: collinearity and measurement error. Collinearity describes the situation where predictor variables are correlated to some degree, making it difficult to disentangle direct and indirect relationships with the response variable (Wetherill et al. 1986: chapter 4; Quinn and Keough 2002: 127–129; Graham 2003). For example, body mass often covaries strongly with other independent variables in the model due to allometric scaling or life-history factors, which raises issues about collinearity (e.g. Graham 2003). Measurement error depicts the statistical phenomenon that data points are subject to a certain degree of uncertainty due to instrumental errors or to true within-subject variation. Increasing measurement error reduces the repeatability of behavioural traits (Bell et al. 2009; Garamszegi et al. 2009), which can cause a downward bias in parameter estimations (McArdle 2003; Adolph and Hardin 2007). Freckleton's (2010) simulations relying on four simple models that combine two predictors in different linear combinations reveal that, in the presence of collin-earity, parameter estimates are unbiased, but the sampling variance (error) increases in conventional regression analysis. The exercise implied that model averaging can, on the other hand, decrease this variance inflation for predictors, but only at the cost of increased bias, which can be of concern if collinearity is strong. When predictors were measured with error, were collinear, and had non-zero effects, parameter estimates appeared biased irrespective of the estimation method and only depended on the degree of measurement error in Freckleton's (2010) work.

Model fitting approaches assume that the data set is complete, i.e. that observations are available for each individual and for each variable (Nakagawa and Freckleton 2008, 2010). The IT method thus requires case-wise deletion of data if any missing value occurs, which can be quite common in behavioural ecological datasets (especially in comparative studies, in which data originate from different sources). In fact, it can often happen that data availability depends on individual behaviour or life history (e.g. more active or better surviving individuals are easier to observe); thus, missing data occur non-randomly with respect to a focal variable. Nakagawa and Freckleton (2010) investigate the effect of missing observations on the IT output in an example analysis. They show that different patterns of missing data can have different consequences for estimation precision (or statistical power) and parameter estimation. If missing data occur non-randomly, the relative support of particular models during model selection is altered, which can lead to incorrect ranking and equivocal model compar-ison. Therefore, the presence of missing observations can influence the model averaged results, and can also cause biased parameter estimations. The authors recommend the use of statistical approaches, such as multiple imputation and data augmentation techniques that can be used to handle the problems due to missing data even when they occur non-randomly. These imputation and augmentation techniques can effectively be incorporated in the AIC–IT framework to provide unbiased parameter estimates.

The topics visited in this volume do not provide an exhaustive list of statistical issues that can emerge in association with the spread of IT application within our discipline. For example, we have not discussed the possible contribution of the method in phylogenetic comparative studies (Harvey and Pagel 1991). This is an area where AIC-based model selection would potentially provide significant enrichment because various models considering different phylogenetic hypothesis, mode of evolution (e.g. punctuational or gradual), or evolutionary process (e.g. Brownian motion or Ornstein–Uhlenbeck process) could be fitted to the interspecific data. Further exploitations can be envisioned, and specific problems within behavioural ecology await for appropriate tests in the future.

Towards an objective statistical choice

Several problems in behavioural ecology involve more than one alternative hypothesis and therefore call for multimodel statistical inferences. The IT approach has been developed for such inferences; thus, its use may offer benefits when applied to designs with multiple predictor variables, frequently encountered in our field. An objective statistical choice should consider this as a potentially useful tool to answer questions specific to behavioural ecology. We hope that this assimilation process will be an intellectually active and critical process and not just an automatic takeover of a fashionable method from another discipline, for which this volume brings up some stimulating points.

Although this issue deals with AIC-based model selection, AIC–IT methods are not the only tools that should be considered in this respect. As indicated at the beginning, a full range of model selection methods are available that might be completely unknown to most of the readers and that use completely different strategies or/and rely on different criteria for model fit. These methods seem to be even less represented in the ecological literature than the AIC–IT approach, and I doubt that this unresponsiveness is motivated by well-founded statistical evidence. Moreover, recently advocated Bayesian approaches can also be used for model selection problems (O'Hara and Sillanpää 2009). Bayesians open up a new philosophical horizon in statistical modelling because they emphasise the importance of existing knowledge as a prior distribution, which is then updated in the light of new data to obtain the posterior probability distribution of a hypothesis (Gelman and Hill 2007; McCarthy 2007). Note that the IT method follows the likelihood principle, which asserts that all relevant information is contained in a likelihood function for the observed data (Berger and Wolpert 1984). Therefore, the AIC–IT method is not the only method for analysing data along multiple hypotheses, and there might be situations when other methodologies fit better to the data and question at hand. Some features of the AIC–IT tool may require further test and development in association to analytical problems that we encounter in behavioural ecology. However, compared with classical ways of the analyses of multi-predictor data, the IT approach is one of the improvements in the techniques available and has focussed people on new ways of thinking.

As a final word, I find it important to note that fulfilling the aims of this special issue is fraught with a great responsibility. Our goal is to stimulate statistical training within our discipline based on the taste and expertise of practicing ecologists and not statisticians. The "easy to read" aspect of this discussion may enhance the spread of statistical concepts of major importance, but it is dangerous to adopt these concepts without consulting the full details of the underlying scientific (i.e. statistical) evidence in the primary literature. Accordingly, I recommend the reading of subsequent papers as an interface between research practice and advanced mathematical statistics. Moreover, given the limits of this Special Issue, it is difficult to guarantee that we have successfully provided a balanced overview on a controversial statistical topic. However, we hope that we were able to highlight some statistical issues that will stimulate behavioural ecologists to develop and maintain a pluralistic statistical perspective.

References

Adolph SC, Hardin JS (2007) Estimating phenotypic correlations: correcting for bias due to intraindividual variability. Funct Ecol 21:178–184

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Csáki F (ed) 2nd International Symposium on Information Theory. Akadémiai Kiadó, Budapest, pp 267–281

Allen DM (1974) The relationship between variable selection and data augmentation and a method for prediction. Technometrics 16:125–127

Anderson DR (2008) Model based inference in the life sciences: a primer on evidence. Springer, New York

Anderson DR, Burnham KP (2002) Avoiding pitfalls when using information-theoretic methods. J Wildl Manage 66:910–916

Anderson DR, Burnham KP, Thompson WL (2000) Null hypothesis testing: problems, prevalence, and an alternative. J Wildl Manage 64:912–923

Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. Anim Behav 77:771–783

Berger JO, Wolpert RL (1984) The likelihood principle. Institute of Mathematical Statistics, Hayward

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York

Burnham K, Anderson D, Huyvaert K (2010) Improving inferences in ecological and behavioral science: some background, observations, and comparisons. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1029-6

Cézilly F, Danchin É, Giraldeau L-A (2008) Research methods in behavioural ecology. In: Danchin É, Giraldeau L-A, Cézilly F (eds) Behavioural ecology: an evolutionary perspective on behaviour. Oxford University Press, Oxford, pp 55–95

Chamberlin TC (1890) The method of multiple working hypotheses. Science 15:92–96

Claeskens C, Hjort NL (2008) Model selection and model averaging. Cambridge University Press, Cambridge

Craven P, Wahba G (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numerische Mathematik 31:377–403

Crawley MJ (2007) The R book. Wiley, West Sussex

Derksen S, Keselman HJ (1992) Backward, forward and stepwise automated subset-selection algorithms—frequency of obtaining authentic and noise variables. Br J Math Stat Psychol 45:265–282

Dochtermann N, Jenkins SH (2010) Developing and evaluating candidate hypotheses in behavioral ecology. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1039-4

Draper NR, Smith H (1981) Applied regression analysis, 2nd edn. Wiley, New York

Forster MR (2000) Key concepts in model selection: performance and generalizability. J Math Psychol 44:205–231

Forstmeier W, Schielzeth H (2010) Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1038-5

Fox J (2002) An R and S-PLUS companion to applied regression. Sage, Newbury Park

Freckleton RP (2010) Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1045-6

Garamszegi LZ, Calhim S, Dochtermann N, Hegyi G, Hurd PL, Jørgensen C, Kutsukake N, Lajeunesse MJ, Pollard KA, Schielzeth H, Symonds MRE, Nakagawa S (2009) Changing philosophies and tools for statistical inferences in behavioral ecology. Behav Ecol 20:1363–1375

Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge

Ginzburg LR, Jensen CXJ (2004) Rules of thumb for judging ecological theories. Trends Ecol Evol 19:121–126

Graham MH (2003) Confronting multicollinearity in ecological multiple regression. Ecology 84:2809–2815

Guthery FS (2007) Deductive and inductive methods of accumulating reliable knowledge in wildlife science. J Wildl Manage 71:222–225

Guthery FS, Brennan LA, Peterson MJ, Lusk JJ (2005) Information theory in wildlife science: critique and viewpoint. J Wildl Manage 69:457–465

Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford

Hegyi G, Garamszegi LZ (2010) Stepwise selection and information theory in ecology and behavior. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1036-7

Hilborn R, Mangel M (1997) The ecological detective: confronting models with data. Princeton University Press, Princeton

Hobbs NT, Hilborn R (2006) Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecol Appl 16:5–19

Huston MA (1997) Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. Oecologia 110:449–460

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends Ecol Evol 19:101–108

Jones KS, Nakagawa S, Sheldon BC (2009) Environmental sensitivity in relation to size and sex in birds: meta-regression analysis. Am Nat 174:122–133

Konishi S, Kitagawa G (2008) Information criteria and statistical modeling. Springer, New York

Krebs JR, Davies NB (1984) Behavioural ecology: an evolutionary approach. Blackwell Scientific, Oxford

Lajeunesse MJ (2009) Meta-analysis and the comparative phylogenetic method. Am Nat 174:369–381

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

Liang H, Wu HL, Zou GH (2008) A note on conditional AIC for linear mixed-effects models. Biometrika 95:773–778

Linhart H, Zucchini W (1986) Model selection. Wiley, New York

Lukacs PM, Thompson WL, Kendall WL, Gould WR, Doherty PF, Burnham KP, Anderson DR (2007) Concerns regarding a call for pluralism of information theory and hypothesis testing. J Appl Ecol 44:456–460

Mallows CL (1973) Some comments on Cp. Technometrics 15:661–675

Massart P (2007) Concentration inequalities and model selection: ecole d'eté de probabilités de Saint-Flour XXXIII—2003. Springer, Berlin

McArdle BH (2003) Lines, models, and errors: regression in the field. Limnol Oceanogr 48:1363–1366

McCarthy MA (2007) Bayesian methods for ecology. Cambridge University Press, Cambridge

McQuarrie ADR, Tsai C-L (1998) Regression and time series model selection. World Scientific, Singapore

Mundry R (2010) Issues in information theory based statistical inference—a commentary from a frequentist's perspective. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1040-y

Mundry R, Nunn CL (2008) Stepwise model fitting and statistical inference: turning noise into signal pollution. Am Nat 173:119–123

Murtaugh PA (2009) Performance of several variable-selection methods applied to real ecological data. Ecol Lett 12:1061–1068

Nakagawa S, Freckleton R (2008) Missing inaction: the dangers of ignoring missing data. Trends Ecol Evol 23:592–596

Nakagawa S, Freckleton RP (2010) Model averaging, missing data and multiple imputation: a case study for behavioural ecology. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1044-7

O'Hara RB, Sillanpää MJ (2009) A review of Bayesian variable selection methods: what, how, and which. Bayesian Analysis 4:85–118

Owens IPF (2006) Where is behavioural ecology going? Trends Ecol Evol 21:356–361

Platt JR (1964) Strong inference. Science 146:347–353

Popper KR (1963) Conjectures and refutations. Routledge and Keagan Paul, London

Pötscher BM (1989) Model selection under nonstationary: autoregressive models and stochastic linear regression models. Ann Stat 17:1257–1274

Quinn JF, Dunham AE (1983) On hypothesis testing in ecology and evolution. Am Nat 122:602–617

Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge

Rabosky DL (2006) Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164

Rao CR, Wu Y (1989) A strongly consistent procedure for model selection in a regression problem. Biometrika 76:369–374

Richards SA, Whittingham MJ, Stephens PA (2010) Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1035-8

Ripplinger J, Sullivan J (2008) Does choice in model selection affect maximum likelihood analysis? Syst Biol 57:76–85

Rissanen J (1978) Modeling by shortest data description. Automatica 14:465–471

Royall MR (1997) Statistical evidence: a likelihood paradigm. Chapman and Hall, London

Rushton SP, Ormerod SJ, Kerby G (2004) New paradigms for modelling species distributions? J Appl Ecol 41:193–200

Sakamoto Y (1991) Categorical data analysis by AIC. KTK Scientific, Tokyo

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Shibata R (1981) An optimal selection of regression variables. Biometrika 68:45–54

Sokal RR, Rohlf FJ (1995) Biometry, 3rd edn. Freeman, New York

Steidl RJ (2006) Model selection, hypothesis testing, and risks of condemning analytical tools. J Wildl Manage 70:1497–1498

Stephens PA, Buskirk SW, Hayward GD, Del Rio CM (2005) Information theory and hypothesis testing: a call for pluralism. J Appl Ecol 42:4–12

Stephens PA, Buskirk SW, del Rio CM (2007a) Inference in ecology and evolution. Trends Ecol Evol 22:192–197

Stephens PA, Buskirk SW, Hayward GD, Del Rio CM (2007b) A call for statistical pluralism answered. J Appl Ecol 44:461–463

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. Jean-le-Blanc Journal of the Royal Statistical Society, Series B 36:111–147

Sugiura N (1978) Further analysis of the data by Akaike's information and the finite corrections. Commun Stat A7:13–26

Symonds M, Moussalli A (2010) Model selection, multimodel inference and model averaging using Akaike's information criterion: an introduction for statistically terrified behavioural ecologists. Behav Ecol Sociobiol. doi:10.1007/s00265-010-1037-6

Takeuchi K (1976) Distribution of informational statistics and a criterion of model fitting (in Japanese). Suri-Kagaku (Mathematical Sciences) 153:12–18

Towner MC, Luttbeg B (2007) Alternative statistical approaches to the use of data as evidence for hypotheses in human behavioral ecology. Evol Anthropol 16:107–118

Vaida F, Blanchard S (2005) Conditional Akaike information for mixed-effects models. Biometrika 92:351–370

Vapnik V, Chervonenkis A (1974) Theory of pattern recognition (in Russian). Nauka, Moscow

Ward EJ (2008) A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecol Modell 211:1–10

Wetherill GB, Duncombe P, Kenward M, Kollerstrom J, Paul SR, Vowden BJ (1986) Regression analysis with applications. Chapman and Hall, London

Whiteheat H (2007) Selection of models of lagged identification rates and lagged association rates using AIC and QAIC. Commun Stat, Simul Comput 36:1233–1246

Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? J Anim Ecol 75:1182–1189

Zucchini W (2000) An introduction to model selection. J Math Psychol 44:41–46