

Model structure adequacy analysis: selecting models on the basis of their ability to answer scientific questions

Mark L. Taper · David F. Staples ·
Bradley B. Shepard

Received: 9 November 2007 / Accepted: 9 November 2007 / Published online: 19 February 2008
© Springer Science+Business Media B.V. 2008

Abstract Models carry the meaning of science. This puts a tremendous burden on the process of model selection. In general practice, models are selected on the basis of their relative goodness of fit to data penalized by model complexity. However, this may not be the most effective approach for selecting models to answer a specific scientific question because model fit is sensitive to all aspects of a model, not just those relevant to the question. Model Structural Adequacy analysis is proposed as a means to select models based on their ability to answer specific scientific questions given the current understanding of the relevant aspects of the real world.

Keywords Model selection · Model identification · Model development · Multi-model inference · Bull trout · *Salvelinus confluentus* · Monitoring

1 Introduction

1.1 Models, model selection, and science

Models are an integral part of science. They are used for illustrating our understanding of the world, theoretically exploring complex system dynamics, making inferences

M. L. Taper (✉) · B. B. Shepard
Department of Ecology, Montana State University, 310 Lewis Hall,
Bozeman, MT 59717-3460, USA
e-mail: taper@rapid.msu.montana.edu

D. F. Staples
Fish & Wildlife Biometrics Unit, Minnesota Department of Natural Resources,
5463-C West Broadway, Forest Lake, MN 55025, USA

B. B. Shepard
Montana Cooperative Fishery Research Unit, Montana Fish, Wildlife & Parks,
1400 South 19th Avenue, Bozeman, MT 59718-5417, USA

from observations, and for determining appropriate actions for managing complex systems (Nichols 2001). Generally, a model is an abstraction or simplification of a real-world system, and as such, it has been argued that models carry the meaning of science (Giere 2004). Taper (2004) and Taper and Lele (2004a, b) argued that science and scientists makes progress by successively replacing existing models with new models of increasingly accurate approximation to nature's underlying reality. Chatfield (1995) believes that the major source of error in scientific analysis is using the wrong model. Model choice thus has immense influence on our understanding and decision-making capabilities.

In the sciences with which the authors are most familiar (ecology, evolutionary biology, environmental science, conservation biology, and fish & wildlife management), model development and choice has generally followed two pathways. In the first, complex theoretical models are developed that represent the current biological understanding of the system in question. While these complex models are useful for exploring system responses and trajectories under various hypotheses, they present difficulties for estimating parameter values (Wiegand et al. 2003), often fit poorly to data (Mangel et al. 2001), give unreliable predictions (Berryman 1991; Liebholt 1994), and lack ongoing testing and development (Logan 1994). Alternatively, selecting from a suite of relatively simple candidate models using information criteria (Akaike 1973; Akaike 1987; Taper 2004) is becoming increasingly common (Hobbs and Hilborn 2006; Johnson and Omland 2004; Stephens et al. 2005). The use of information criteria has benefited science by fostering the comparison of multiple models and thereby reducing model selection error (Taper and Lele 2004b).

The information criteria approach to model selection has a major drawback in that all comparisons are relative to the candidate model set. You may find a model is the best in your set without realizing that it is only the best of a bad lot. For instance, if the full model does not include important covariates, any model selected from this candidate set will be missing those important covariates. Lindsay (2004) has recently proposed testing for what he calls "Model Adequacy". Model adequacy is a class of generalized goodness of fit measures that compare the distribution of observations predicted by a model to a non-parametric estimate of the true distribution. These ideas are developed more fully in the forthcoming book by Lindsay and Markatou titled *Statistical distances: a global framework to inference* (Springer in Press).

In addition, the ability of a model to fit past data does not say everything about whether the model will make correct policy predictions (Hilborn and Walters 1992). Likelihood (and other goodness of fit measures) fit all aspects of the data, but not all aspects of data are equally relevant to a scientific question. Consider Fig. 1, where a is the object in the real world and b and c are models. Model b is clearly a much better fit than model c as many more of the features of a are matched. However, if your scientific question is whether the object will fall over or not, than model c is more useful than model b because its broad base matches the base of a .

There is a growing recognition that models should be selected based on their ability to answer questions of interest (Lele 2004; Lindsay 2004; Minin et al. 2003; Reynolds and Ford 1999; Taper 2004). These questions can vary widely over theoretical and applied science; consequently, model selection techniques should be adaptable and

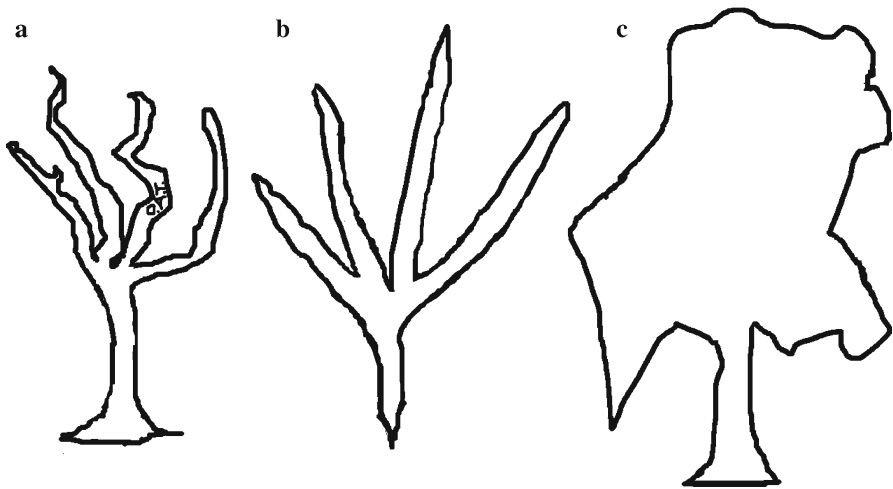


Fig. 1 (a) Represents an object in the real world, (b) and (c) represent models of (a)

not conform to a “one size fits all” dogma. It is possible that the model that will best answer the question of interest is not the one with the best fit to the data.

The information criteria approach essentially takes a statistical viewpoint to model selection. The investigator begins by assessing what model is best supported by data at hand, and then sees what scientific questions can be answered using this model. Thus, the focus moves from data to models to scientific question. This approach seems particularly germane for retrospective analyses of the many long-term, time-series data sets that have been archived over the years.

As useful as the statistical viewpoint has been to ecology, a scientific viewpoint to model development and selection may also be important. With a scientific viewpoint, the researcher begins with a focus on delineating the scientific question of interest, then searches for models to represent or probe this question, and finally considers what data is needed to parameterize the models. This approach is proactive and relevant to the design of monitoring protocols, experiments, and survey designs. To effectively select models, the researcher needs to answer three levels of questions: first, what is the fundamental scientific question to be answered? Second, what model structure, or degree of complexity, is needed to adequately answer this question? And third, what data observed at what precision, are needed to parameterize the selected model? This trichotomy corresponds strongly to Mayo’s (1996, 2004) hierarchy of models.

In this paper, we present a simulation-based methodology for model selection from a scientific viewpoint, which we call model structural adequacy (MSA) analysis. MSA analysis is used to evaluate the model structure, or complexity, that is necessary for usefully answering scientific and/or management questions. This method can be used to prospectively determine what model structure and data are necessary, or if useful answers are even possible, given resource constraints.

The purpose of this paper is to develop this useful mode of thinking and to make it concrete by giving it a name. We demonstrate MSA’s utility with an example that

investigates bull trout monitoring procedures. This MSA analysis generated surprising results that led to our development and recommendation of a monitoring program that is both accurate and economical.

1.2 Model structural adequacy procedures

The seven steps of the basic MSA approach are as follows:

1. Formulate an explicit scientific question of interest that addresses ultimate study goals. It is critical to link the scientific question to a measurable feature of the system and translate it into specific statistical questions.
2. Construct mechanistic simulation models of the underlying process using the best available information and expert opinion. We refer to these as complex process models or CPMs. This is not a trivial step, and many of these models may be largely speculative; however, model construction is, at a minimum, an opportunity to explicitly state what is known and unknown about the true process.
3. Construct a suite of candidate models. The goal is to find a model that: (a) captures the essential features of the underlying process as they relate to the question of interest, and (b) can be parameterized with the resources available. The candidate model set will be most useful if it contains a variety of model structural forms and complexities.
4. Generate data sets of realistic size from the CPM and calculate the metric of interest (i.e. define the true answer to question of interest in the modeled ‘world’).
5. Use the data generated in step 4 to estimate the parameters for each of the candidate models.
6. Use the constructed data from step 4 and the parameter estimates from step 5 to estimate/predict the metric of interest using each candidate model. For each candidate model, compare the estimate of the metric of interest to the ‘true’ value from step 4.
7. Repeat steps 4–6 a large number of times accumulating results to build inductive support for candidate models based on how well each model answers the question of interest. Generally, this can be characterized by the distributions of the estimates of the metric of interest calculated using each candidate model.

1.3 Model structural adequacy and the identification of errors

As Mayo (1996) points out, it is critical to be able to identify the kinds of errors that can occur in an analysis. Model selection is subject to two kinds of errors: structural errors (errors of approximation) and estimation errors (uncertainty in parameter estimation) (Bozdogan 1987; Taper 2004). Structural, or approximation, error is the discrepancy between the underlying natural distribution and the best possible approximation from a model of given structural complexity. Invariably, the discrepancy between the underlying distribution and the distribution predicted by a model will be greater than the approximation error because parameter estimates will differ from their optimal values. This added discrepancy is the estimation error.

The impact both of these types of errors have on the adequacy of the candidate models in MSA analysis can be investigated through simulation. The discrepancy found in step 6 of the basic MSA between the candidate models' estimates of the metric of interest and the true value of the metric are the total discrepancies and contain both types of errors. The approximation error can be determined by repeating the analysis, but with the large data sets generated in step 4. These data sets should be sufficiently large that parameter estimation error is minimal. The difference between the total discrepancy and the approximation discrepancy is the estimation error.

Large approximation errors indicate that other candidates should be sought. On the other hand, large estimation errors indicate that larger sample sizes or the development of more efficient statistical estimation techniques should be considered.

Model selection in MSA is also subject to a third type of error. MSA inference is dependent on the degree to which the complex process models we formulate are representative of true process. The CPM that is used for generating the data is itself only approximating the true underlying process. Thus, computed model adequacy from the basic MSA analysis will only be an estimate of the adequacy of the candidate models when applied to real data. We term this third source of error formulation error.

There are two sources of formulation error. First, the investigators are likely to be uncertain as to what values of the parameters in the CPM will best represent the underlying process. And second, the investigators may be uncertain regarding important features of the underlying process. The impact of both of these sources of formulation error can also be investigated.

Parameter uncertainty can be evaluated by drawing estimates of the parameters used by the CPM in step 4 from reasonable parameter distributions. Alternative CPMs, each representing alternative process uncertainty, can be used to evaluate effects that process uncertainty has on the adequacy of candidate models.

Below, we illustrate MSA analysis with an example from our research on monitoring trout populations in the Flathead Basin of NW Montana, USA. We used MSA to evaluate whether current monitoring data coupled with simple scalar population dynamic models are useful for estimating a population's risk of declining below a conservation threshold, or if more detailed demographic models offer justifiable improvement in risk estimates, given that estimating demographic rates will be costly.

2 Example application

2.1 Question of interest

Bull trout, *Salvelinus confluentus*, is a native top-level predator species in the Flathead basin, but its abundance has severely declined primarily due to a major community shift in Flathead Lake (Deleray et al. 1999) and habitat changes in the natal tributaries (Rieman et al. 1997). Two fundamental management goals for this threatened population are to prevent further declines below sustainable levels and facilitate increases to some, as yet undefined, secure abundance level. Given these objectives, an estimate of the risk of the population declining is a useful monitoring indicator that consolidates relevant information about the population's status (Staples et al. 2005). Monitoring

the population's risk is also more concordant with the definitions of “endangered” (‘in danger of extinction’) and “threatened” (‘likely to become endangered in foreseeable future’) in the Endangered Species Act (U.S. ESA of 1973 [as in U.S. Code 2000]) than is monitoring the population's trend or abundance. Thus, estimating risk of decline for the population is a primary question of interest for the management agencies charged with managing this species.

Next, we formulated an explicit scientific question that addresses the monitoring goal: to estimate risk of decline. For bull trout, the resulting scientific question was: “What is the probability that adult abundances will decline below an *a priori* threshold within 10 years?” This probability is hereafter denoted as the PLT10. To lessen extrapolation errors, we used a ‘quasi-extinction’ threshold, i.e. an abundance threshold, higher than actual extirpation, which ideally represents some biologically relevant abundance level. Relatively short 10-year prediction intervals were used to further decrease extrapolation errors, and are more relevant to management goals and timeframes. Support for candidate models was evaluated based on how well a given model estimates the PLT10 from simulated data.

For risk-based monitoring, a model is needed that can deliver useful estimates of risk of decline from data that can be realistically obtained. In our case, risk was defined as declines in adult abundance because of the biological and practical significance of adults to the Flathead population. Adult bull trout have completed the demanding migratory life history, and have the highest reproductive values in the population. Further, data on spawning adults (i.e. redd counts) have been collected consistently for over 25 years for this bull trout population, so there are ample data to parameterize count-based population models. Redd numbers can be estimated with moderately high precision (Muhlfeld et al. 2006), and may be more indicative of trends in adult numbers than other available data series such as lake netting data. (DFS unpublished analyses) Utilizing redd data in count-based models represents an ideal way to monitor the bull trout population as redds are relatively easy to collect and non-intrusive to this threatened population.

2.2 Complex process models

In the MSA simulations, mechanistic models of a bull trout population (i.e. the complex process models) represented the ‘true’ population, and were used to generate time-series data and calculate the true probability of the population reaching quasi-extinction starting from the last observation of the data series. The CPMs were age-structured (age 0–9), spatially explicit process models of the bull trout population based on information available in the literature, government reports, university theses, and expert opinion of Flathead field biologists (see acknowledgements) educed in a series of meetings (see Staples 2006 for model details).

2.3 Candidate models

Five candidate models were used to estimate the probability of decline based on simulated abundance data from the process models. We deliberately used candidate models

with different structures from the process models for insight into how well we could approximate risk with imperfect knowledge of the complex population dynamics. Three of the candidate models were based on count data, and could be used with existing redd data. The three count-based models are discrete-time versions of the density independent exponential growth model (DI; [Dennis et al. 1991](#)), a Ricker density dependent model (RD), and a Gompertz density dependent model (GD). Parameter estimation from time-series data for all three models was done using methods of [Dennis and Taper \(1994\)](#).

The remaining two candidate models, a stage (SG) and an age-stage model (AS), used demographic information on vital rates and age class abundances to make population predictions, but these models would require a significant amount of additional sampling effort for implementation. These models were developed by aggregating components of the age-structured life history through elasticity and life stage simulation analyses ([Caswell 2001](#); [Wisdom et al. 2000](#)) guided by biological intuition. To evaluate structural errors, we used the ‘true’ vital rate distributions from the process models for estimating PLT10 with the SG and AS models.

2.4 Simulations

To evaluate the ability of the candidate models to reasonably estimate the distribution of PLT10, we simulated 1,000 time-series of bull trout abundance data with a CPM, which was then used to calculate the corresponding true PLT10 from the end of the data series. These simulated data series were used to estimate parameter values for the candidate models, which were then used to estimate PLT10. The distributions of the PLT10 estimates from the candidate models were compared to the true PLT10 distribution. The simulation process was repeated for different CPM models representing different hypotheses about the true life history of the bull trout population (see below).

2.5 Structural and estimation errors

Estimates of the PLT10 from the count-based candidate models have several sources of variation: (1) errors resulting from structural differences between the model and the true process, (2) sampling error in model parameter estimates, and consequently predictions, due to observing the population for a finite number of time periods, and (3) variation due to measurement error in abundance data. We focused on the first two sources of variation, structural and estimation errors, in count-based model predictions for this study. Methods to accommodate measurement error in abundance data for the DI and GD are covered in [Staples et al. \(2004\)](#) and [Dennis et al. \(2006\)](#); measurement error in the RD model is covered in De Valpine and Hastings ([De Valpine and Hastings 2002](#)). To evaluate structural error for count-based models (DI, RD, GD), each simulated data series had length 1,000 to minimize sampling error in parameter estimation. To evaluate the effect of sampling error in parameter estimates, of the simulated data series had a more realistic length 20.

Variation in candidate demographic model predictions is caused by: (1) structural differences between the model and the true process, (2) measurement error in ini-

tial age/stage abundance estimates, and (3) uncertainty in candidate model vital rate estimates. We focused on the first two sources, and did not evaluate the effect of error in vital rate estimates; in general, higher uncertainty in vital rates will increase the estimated risk of decline (Staples 2006). To evaluate structural errors in predictions, initial age and stage abundances were assumed to be the exact abundances from the simulated data series. To represent sampling error in abundance estimates, initial stage and age abundances were contaminated with an error term resulting in a coefficient of variation (CV) of 0.25 for initial abundances.

2.6 Formulation errors

While the general life history configuration of the Flathead bull trout population is known, there are key uncertainties concerning vital rate distributions and possible density dependent interactions among juvenile age-classes. Consequently, we used three process models incorporating different hypotheses about density-dependence among juvenile cohorts. The three process models (denoted DD1, DD2, and DD3) were parameterized using five different suites of vital rate distributions: a base set of vital rate values representing the consensus values plus 4 sets of alternate distributions in which adult and sub-adult survival rates, key survival rates as determined by elasticity analyses, were perturbed from their base values (both increased, both decreased, one increased/one decreased, and vice-versa).

2.7 Results

The distributions of PLT10 for both structural and sampling simulations are given in Fig. 2 for the three process models and candidate models using the base set of vital rate distributions. In the base simulations, the ‘true’ distribution of risk had a strong positive skew for all three CPMs. Of the count-based candidate models, PLT10 estimates from the GD model had the highest correlation to true risk in all simulations (Table 1). The GD model estimated true risk much better than the RD model for the DD1 sampling simulations, though both gave similar PLT10 estimates for the DD2 and DD3 simulations. Sampling error in parameter estimation increased the variability of risk estimates for the RD and GD models, but had little effect on predictions from the DI model.

Like the true risk distribution, PLT10 estimates from the demographic SG candidate model had positive skew but tended to be biased low and less variable than the true risk (Fig. 2). The distribution of PLT10 estimates from the AS model were close to the true distribution of risk for simulations using the base vital rate values, and had the highest correlation to the true risk of decline (Table 1). Sampling error in abundance estimates had little effect on risk estimates from the SG and AS models.

Changing vital rate distributions in the CPMs affected the adequacy of candidate model predictions, most notably for the demographic candidate models. However, for simulations in which sub-adult survival rates were increased, the SG and AS demographic candidate models underestimated the PLT10 regardless of whether the mean adult survival rate was increased or decreased (Fig. 3). When the means of

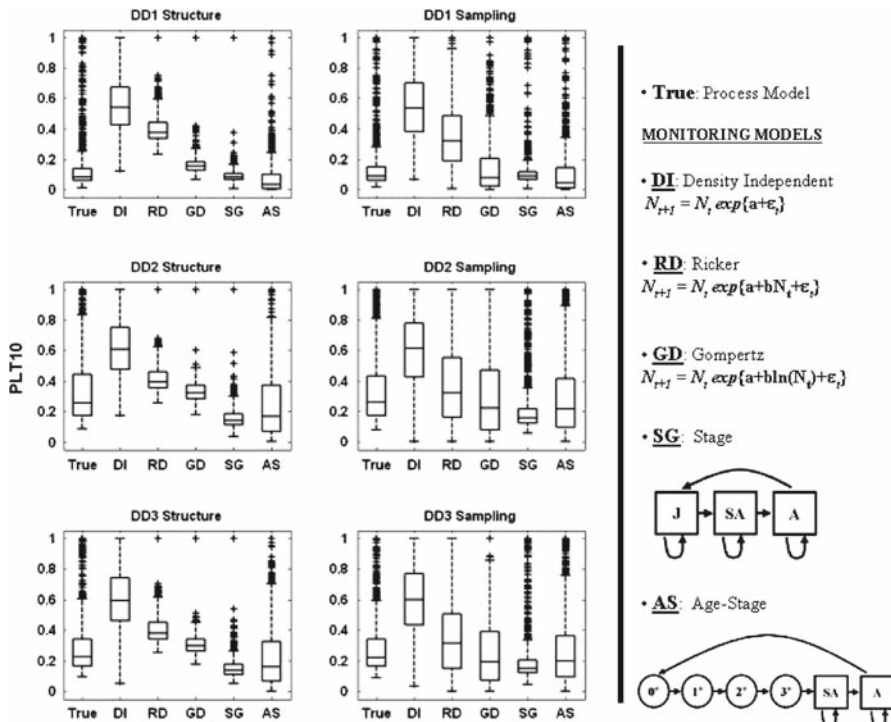


Fig. 2 Estimated PLT10 for structural and sampling simulations. Models: True-process model, DI—Density independent, RD—Ricker density dependent, GD—Gompertz density dependent, SG—Stage, AS—Age-stage

Table 1 Correlation between candidate model risk predictions and true PLT10 from process models using the base set of demographic rates

| | | Correlation with true PLT10 | | | | |
|-----|-----------|-----------------------------|------|------|------|------|
| | | DI | RD | GD | SG | AS |
| DD1 | Structure | 0.27 | 0.46 | 0.55 | 0.59 | 0.68 |
| | Sample | 0.31 | 0.35 | 0.45 | 0.55 | 0.63 |
| DD2 | Structure | 0.39 | 0.58 | 0.59 | 0.62 | 0.75 |
| | Sample | 0.38 | 0.45 | 0.47 | 0.60 | 0.73 |
| DD3 | Structure | 0.40 | 0.62 | 0.64 | 0.64 | 0.72 |
| | Sample | 0.37 | 0.43 | 0.44 | 0.62 | 0.71 |

both survival rates were increased, the GD model slightly underestimated risk, but the SG and AS models rarely showed any risk of decline. If sub-adult survival was high, but adult survival was low, the true PLT10 distribution was wide and positively skewed. In this case, the GD model gave reasonable estimates of risk, though these risk estimates were less variable than the true risk of decline. The SG and AS demographic candidate models, on the other hand, rarely showed any measurable risk for these simulations even though the demographic candidate models used the same shifted

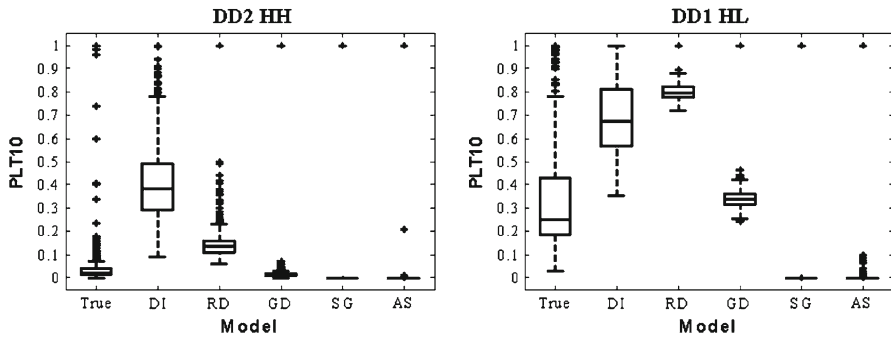


Fig. 3 Estimated PLT10 for simulations of process models using alternate survival rates for sub-adults and adults. HH denotes both mean survival rates were increased; HL denotes sub-adult survival increased and adult survival decreased

vital rate distributions as the process models. This demonstrates potential structural errors in risk predictions resulting from aggregating age classes into “stages”.

2.8 Implications for bull trout monitoring

Though differences in juvenile dynamics or survival rates have important effects on patterns of adult abundance and risk dynamics, this study offers general insights into monitoring an age-structured population like bull trout. The DI model, which has been used in previous bull trout analyses (Staples *et al.* 2005), tends to be imprecise and overestimate risk. Of the count-based models, the Gompertz (GD) performed the best over the range of process models and vital rate distributions examined. The distributions of PLT10 estimates from the GD model were similar to that of the true process, and the GD model approximated the mean PLT10 for the populations reasonably well. The GD model parameterized with redd data from the Flathead is likely the best available means for approximating risk of decline for the bull trout population.

Demographic models incorporating age-structure for juveniles potentially could increase the ability to detect problems in bull trout populations and may be useful for populations for which there are no historical time-series data available; however, changes in vital rate distributions detrimentally affected demographic model risk predictions. PLT10 estimates from the demographic models, using the ‘true’ vital rate distributions, were inadequate due to structural differences between the demographic candidate models and complex process models. The sensitivity of demographic model risk estimates to changes in survival rates suggests caution is warranted in implementing demographic models for monitoring the population. These models must be carefully developed and their sensitivity to structural assumptions evaluated because they can give misleading results, even when the correct vital rate values are known.

3 Discussion

In general practice, model selection is based on finding the ‘best’ fitting model. A model’s fit to data is often penalized by its complexity as in information criteria

(Taper 2004). Thus, as Hobbs and Hilborn (2006) say: “[...]the level of detail in the model is decided by the data available to estimate the parameters in the model.” Scientific inference under the model selection paradigm assumes that parameters in the selected model reflect similar features acting in the real world. For example, Hooten (1995), Zeng et al. (1998), and Taper and Gogan (2002) use the presence of an intraspecific competition coefficient in selected models as evidence for the influence of density dependence on population dynamics of the studied species.

The model structural adequacy analysis described in this paper is a very different approach from information criteria based model selection, which has no formal means of differentiating between fit to features of interest or fit to epiphenomena. In MSA analysis, model evaluation is based on the ability of a model to give answers to specific questions. For MSA, a good model structure doesn’t need to match underlying truth perfectly or even be the model in a candidate suite that explains the most variance. A good model should fit or capture variance relative to the scientific question being asked.

Another distinction between MSA and information criteria based model selection, is that MSA analysis is prospective and is undertaken before “confronting models with data” (Hilborn and Mangel 1997). Consequently, the risk of model selection bias that Chatfield (1995) warns of is greatly reduced.

Simulations with the CPMs can give insight into emergent properties resulting from the stochastic relationships in a specified mechanistic structure, allowing the modelers to evaluate the performance of complicated model for which field biologists have yet to collect necessary data. In particular, CPM simulations can address the multiple sources of error independently, including: formulation error (error in the CPM itself), process stochasticity, model structural error, sampling error (due to limited amounts of data), and measurement error. Thus, MSA analysis holds a relationship to inference through model selection, similar to that power analysis holds to inference through significance testing.

A MSA analysis can have a variety of beneficial effects on the research process. It creates an opportunity for modelers, managers, and biologist to interact for increased understanding of the system under consideration (Clark and Schmitz 2001). The process of formulating the complex process models forces the modelers and biologists to explicitly consider what is known and unknown about the system. Answers to these questions can help understand the impact of the misspecification of currently used models, help plan future data collection, and build support for the model ultimately determined as most reasonable by the MSA analysis.

MSA analysis as described here has its limitations. MSA is based on the fundamental assumption that the complex process models (CPM’s) accurately represents the real world. But, of course, all models are approximations. As the CPM will be based to a varying degree on expert opinion, subjectivity is part of formulation error. MSA analysis is only useful in situations where the CPM is (with the right parameterization) a good approximation of the real world. This requires extensive knowledge about the biology of the situation. Fortunately, perfect knowledge is not needed; uncertainty can be explored via alternative CPM’s and parameter distributions. Researchers may need to use a wide range of models to bracket the uncertainty regarding the underlying process. Nevertheless, not knowing everything is not an excuse to do nothing. A final

check that CPM formulation error has not distorted the MSA analysis would be to conduct a model adequacy test when the MSA selected model is applied to real data. Of course, the loss function (Lindsay 2004; Minin et al. 2003) for model adequacy should match that used during the MSA analysis.

The concept of a model structural adequacy analysis seems straightforward. Nevertheless, we are unaware of any other cases where it has been applied. Perhaps, this is not surprising, as MSA is step beyond power analysis, which despite being universally recommended in statistics courses, is itself almost never applied in practice. Certainly, experienced modelers apply a MSA analysis in an intuitive fashion; asking themselves what model features they need to include to effectively represent real world behavior. We make such questions a cornerstone of our modeling practice. However, as useful as such musings are, we have found they do not substitute for a formal MSA analysis. We find that even with decades of modeling experience, our intuitions about model behavior are not invariably correct. Models still surprise us.

Acknowledgments We thank Mark Delaray, Jim Dunnigan, Les Everts, Wade Fredenberg, Dan Goodman, Barry Hanson, Mike Hensler, Brian Marotz, Clint Muhlfeld, Drew Parkin, David Rockwell, Scott Rumsey, Bob Snyder, and Tom Weaver for sharing their knowledge of bull trout biology with us. Jim Nichols, Scott Mills, Prasanta Bandyopadhyay, and two anonymous reviewers usefully commented on earlier versions of the manuscript. This work was supported in part by Northwest Power & Conservation Council grant #PTMM11. Montana Fish, Wildlife & Parks grant #Z0862 and by Montana Fish, Wildlife, and Parks contract #060327 all to M. L. Taper.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood. In B. N. Petrov & F. Cs'aki (Eds.), *Proceedings of the 2nd International Symposium on Information theory* (pp. 267–281). Budapest: Akademia Kaído.
- Akaike, H. (1987). Factor-analysis and AIC. *Psychometrika*, 52, 317–332.
- Berryman, A. A. (1991). Population theory: An essential ingredient in pest prediction, management, and policy making. *American Entomologist*, 37, 138–142.
- Bozdogan, H. (1987). Model selection and Akaike information criterion (AIC)—the general-theory and its analytical extensions. *Psychometrika*, 52, 345–370.
- Caswell, H. (2001). *Matrix population models: Construction, analysis, and interpretation*. Sunderland, MA: Sinauer Associates.
- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society A*, 158, 419–466.
- Clark, W. R., & Schmitz, R. A. (2001). When modelers and field biologists interact: Progress in resource science. In T. Shenk & A. Franklin (Eds.), *Modeling in natural resource management: Development, interpretation, and application*. Washington, DC: Island Press.
- De Valpine, P., & Hastings, A. (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs*, 72, 57–76.
- Deleray, M., Knotek, L., Rumsey, S., & Weaver, T. (1999). Flathead lake and river fisheries status report: DJ Report No. F-78-R-1 through 5. Montana Department of Fish, Wildlife and Parks, Kalispell.
- Dennis, B., Munholland, P. L., & Scott, J. M. (1991). Estimation of growth and extinction parameters for endangered species. *Ecological Monographs*, 61, 115–143.
- Dennis, B., Ponciano, J. M., Lele, S. R., Taper, M. L., & Staples, D. F. (2006). Estimating density dependence, process noise, and observation error. *Ecological Monographs*, 76, 323–341.
- Dennis, B., & Taper, M. L. (1994). Density-dependence in time-series observations of natural-populations—estimation and testing. *Ecological Monographs*, 64, 205–224.
- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, 71, 742–752.

- Hilborn, R., & Mangel, M. (1997). *The ecological detective: Confronting models with data*. Princeton: Princeton University Press.
- Hilborn, R., & Walters, C. J. (1992). Quantitative fisheries stock assessment—choice, dynamics and uncertainty. Routledge: Chapman & Hall.
- Hobbs, N. T., & Hilborn, R. (2006). Alternatives to statistical hypothesis testing in ecology: A guide to self teaching. *Ecological Applications*, 16, 5–19.
- Hooten, M. (1995). *Distinguishing forms of statistical density dependence and independence in animal time series data using Information Criterion*. Bozeman, Montana: Department of Biology, Montana State University.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19, 101–108.
- Lele, S. R. (2004). Evidence functions and the optimality of the law of likelihood. In M. L. Taper & S. R. Lele (Eds.), *The nature of scientific evidence: Statistical, philosophical and empirical considerations*. Chicago: The University of Chicago Press.
- Liebold, A. M. (1994). Use and abuse of insect and disease models in forest pest management: Past, present, and future. In W. W. Covington & L. F. DeBano (Eds.), *Sustainable ecological systems: Implementing an ecological approach to land management*. U.S. Forest Service Technical Report RM-247.
- Lindsay, B. G. (2004). Statistical distances as loss functions in assessing model adequacy. In M. L. Taper & S. R. Lele (Eds.), *The nature of scientific evidence: Statistical, philosophical and empirical considerations* (pp. 439–488). Chicago: The University of Chicago Press.
- Logan, J. A. (1994). In defense of big ugly models. *American Entomologist*, 40, 202–207.
- Mangel, M., Fiksen, O., & Giske, J. (2001). Theoretical and statistical models in natural resource management and research. In T. Shenk & A. Franklin (Eds.), *Modeling in natural resource management: Development, interpretation, and application*. Washington, DC: Island Press.
- Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. Chicago: The University of Chicago Press.
- Mayo, D. G. (2004). An error-statistical philosophy of evidence. In M. L. Taper & S. R. Lele (Eds.), *The nature of scientific evidence: Statistical, philosophical and empirical considerations* (pp. 79–118). Chicago: The University of Chicago Press.
- Minin, V., Abdo, Z., Joyce, P., & Sullivan, J. (2003). Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology*, 52, 674–683.
- Muhlfeld, C. C., Taper, M. L., Staples, D. F., & Shepard, B. B. (2006). Observer error structure in bull trout redd counts in Montana streams: Implications for inference on true redd numbers. *Transactions of the American Fisheries Society*, 135, 643–654.
- Nichols, J. D. (2001). Using models in the conduct of science and management of natural resources. In T. Shenk & A. Franklin (Eds.), *Modeling in natural resource management: Development, interpretation, and application*. Washington, DC: Island Press.
- Reynolds, J. H., & Ford, E. D. (1999). Multi-criteria assessment of ecological process models. *Ecology*, 80, 538–553.
- Rieman, B., Lee, D. C., & Thurow, R. F. (1997). Distribution, status, and likely future trends of bull trout within the Columbia River and Klamath Basins. *North American Journal of Fisheries Management*, 17, 1111–1125.
- Staples, D. F. (2006). Viable population monitoring: Risk-based population monitoring for threatened and endangered species with application to bull trout, *salvelinus confluentus*. Ph.D. Dissertation, Department of Ecology, Montana State University, Bozeman, MT.
- Staples, D. F., Taper, M. L., & Dennis, B. (2004). Estimating population trend and process variation for PVA in the presence of sampling error. *Ecology*, 85, 923–929.
- Staples, D. F., Taper, M. L., & Shepard, B. B. (2005). Risk-based viable population monitoring. *Conservation Biology*, 19, 1908–1916.
- Stephens, P. A., Buskirk, S. W., Hayward, G. D., & Del Rio, C. M. (2005). Information theory and hypothesis testing: A call for pluralism. *Journal of Applied Ecology*, 42, 4–12.
- Taper, M. L. (2004). Model identification from many candidates. In M. L. Taper & S. R. Lele (Eds.), *The nature of scientific evidence: Statistical, philosophical and empirical considerations* (pp. 448–524). Chicago: The University of Chicago Press.
- Taper, M. L., & Gogan, P. J. P. (2002). The Northern Yellowstone elk: Density dependence and climatic conditions. *Journal of Wildlife Management*, 66, 106–122.

- Taper, M. L., & Lele, S. R. (2004a). Dynamical models as paths to evidence in ecology. In M. L. Taper & S. R. Lele (Eds.), *The nature of scientific evidence: Statistical, philosophical and empirical considerations* (pp. 275–297). Chicago: The University of Chicago Press.
- Taper, M. L., & Lele, S. R. (2004b). The nature of scientific evidence: A forward-looking synthesis. In M. L. Taper & S. R. Lele (Eds.), *The nature of scientific evidence: Statistical, philosophical and empirical considerations* (pp. 527–551). Chicago: The University of Chicago Press.
- Wiegand, T., Jeltsch, F., Hanski, I., & Grimm, V. (2003). Using pattern-oriented modeling for revealing hidden information: A key for reconciling ecological theory and application. *Oikos*, 100, 209–222.
- Wisdom, M. J., Mills, L. S., & Doak, D. F. (2000). Life stage simulation analysis: Estimating vital-rate effects on population growth for conservation. *Ecology*, 81, 628–641.
- Zeng, Z., Nowierski, R. M., Taper, M. L., Dennis, B., & Kemp, W. P. (1998). Complex population dynamics in the real world: Modeling the influence of time-varying parameters and time lags. *Ecology*, 79, 2193–2209.