Multimodel inference and adaptive management

Sarah E. Rehme [a,*], Larkin A. Powell [b], Craig R. Allen [c]

[a] Nebraska Cooperative Fish and Wildlife Research Unit, School of Natural Resources, University of Nebraska − Lincoln, NE 68583-0984, USA
[b] School of Natural Resources, University of Nebraska − Lincoln, NE 68583-0974, USA
[c] U.S. Geological Survey − Nebraska Cooperative Fish and Wildlife Research Unit, School of Natural Resources, University of Nebraska − Lincoln, NE 68583-0984, USA

A B S T R A C T

Ecology is an inherently complex science coping with correlated variables, nonlinear interactions and multiple scales of pattern and process, making it difficult for experiments to result in clear, strong inference. Natural resource managers, policy makers, and stakeholders rely on science to provide timely and accurate management recommendations. However, the time necessary to untangle the complexities of interactions within ecosystems is often far greater than the time available to make management decisions. One method of coping with this problem is multimodel inference. Multimodel inference assesses uncertainty by calculating likelihoods among multiple competing hypotheses, but multimodel inference results are often equivocal. Despite this, there may be pressure for ecologists to provide management recommendations regardless of the strength of their study's inference. We reviewed papers in the Journal of Wildlife Management (JWM) and the journal Conservation Biology (CB) to quantify the prevalence of multimodel inference approaches, the resulting inference (weak versus strong), and how authors dealt with the uncertainty. Thirty-eight percent and 14%, respectively, of articles in the JWM and CB used multimodel inference approaches. Strong inference was rarely observed, with only 7% of JWM and 20% of CB articles resulting in strong inference. We found the majority of weak inference papers in both journals (59%) gave specific management recommendations. Model selection uncertainty was ignored in most recommendations for management. We suggest that adaptive management is an ideal method to resolve uncertainty when research results in weak inference.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Ecology is an inherently complex science studying phenomena characterized by nonlinear interactions that make it difficult to understand basic relationships and responses to management. Most ecological field research is conducted in relatively short, small scale studies (Wiens, 1989) which are often inadequate to untangle ecological complexity. Wildlife managers and policy makers, whose decisions affect ecosystems at larger scales in space and time, rely on ecologists to provide management recommendations drawn from these short, small scale studies. To cope with the difficulties associated with drawing conclusions from such studies, ecologists are, with increasing frequency, using alternatives to traditional statistical null hypothesis testing in order to disentangle the underlying trends in complex data (Anderson et al., 2000; Johnson and Omland, 2004; Stephens et al., 2007).

Strong inference, where multiple alternative hypotheses are tested with experiments to falsify those hypotheses (Platt, 1964), and adaptive inference, an iterative process of investigation that alternates between minimizing Type I and Type II errors at different places in the investigative process (Holling and Allen, 2002) have been suggested as approaches appropriate to understanding complex problems. Both approaches pose and test branch points in a tree of logically alternative hypotheses. But strong inference relies on situations where causes can be single and separable and where discrimination between pair-wise alternative hypotheses can be determined experimentally by a simple yes or no answer. As Platt (1964) demonstrates, strong inference is a powerful and rapid way to deal with questions in molecular biology, cell biology and physiology. Strong inference is less applicable in ecological systems, where causes are not entirely separable (Hilborn and Stearns, 1982; Pickett et al., 1994). Frequently, competing hypotheses cannot be distinguished by a single unambiguous test or set of controlled experiments, but only by a suite of tests that accumulate a body of evidence supporting one line of argument and not others. Instead of pitting hypotheses against each other, adaptive inference relies on multiple, competing hypotheses followed by tests that develop

* Corresponding author. Tel.: +1 402 472 0449; fax: +1 402 472 2722.
E-mail addresses: ser@huskers.unl.edu (S.E. Rehme), lpowell3@unlnotes.unl.edu (L.A. Powell), callen3@unlnotes.unl.edu (C.R. Allen).

a consistency of pattern lending support to a particular line or lines of argument.

Strong inference and adaptive inference are useful, but not appropriate in all situations. One method that is increasing in prevalence within the fields of ecology and conservation is multi-model inference (Guthery et al., 2005; Hobbs and Hilborn, 2006). Multimodel inference is a statistical technique where alternative plausible models are assessed given the data, based on relative likelihoods (Anderson et al., 2000). These models are selected a priori based on thoughtful, science-based consideration of the problem to be answered and hypotheses about the causal effects behind this problem. These plausible models are then analyzed simultaneously as a set to determine the best approximating model or set of models using information theoretic approaches (Burnham and Anderson, 2002). However, model results are often equivocal due to uncertainty in model selection (Guthery et al., 2005), and researchers are left with the resulting weak inference, with multiple models plausible given the data at hand. Researchers are thus faced with the dilemma of providing management recommendations to managers based on weak inference.

When researchers are required to draw conclusions from multiple plausible models, they have at least three alternatives open to them. One method is to average otherwise equivocal results. Model averaging uses model weights to derive more robust model parameters or model estimates (Johnson and Omland, 2004). Another alternative is to repeat the experiment and post-pone initiating a management regime. However, when manage-ment decisions must be made and it is not feasible to repeat the experiment, a third option, adaptive management, is a logical follow up for researchers and managers when drawing conclusions from research with weak multimodel inference. Adaptive management permits management to continue while managers increase their knowledge through monitoring coupled with well designed management experiments. Management is able to continue because in adaptive management uncertainty is acknowledged, management is designed to reduce sources of uncertainty over time, and management actions are designed to be optimal within the current state of uncertainty (Holling, 1978; Walters, 1986).

The use of adaptive management has been increasing over the last decade (McFadden et al., 2011). Given the changing paradigms in ecological research, that is, the increasing prevalence of multi-model inference, we sought to document the use of multimodel inference in two top management and conservation journals, and the pervasiveness of weak inference resulting from its use. Where weak inference was present in the results from reported field studies, we sought to determine if authors were communicating the uncertainty underlying weak inference to managers, and the type of recommendations that followed from results. Specifically, we evaluated peer-reviewed papers in two journals to (1) quantify the prevalence of multimodel inference, (2) quantify the prevalence of weak inference, and (3) determine what type of management recommendations authors draw from multimodel inference results. We expected weak inference to be abundant within papers that used multimodel inference, and therefore, given the increasing use of adaptive management, we specifically searched within the management recommendations for the endorsement of an adaptive management approach.

2. Methodology

2.1. Inference strength

We reviewed articles in the 2008 issues of the Journal of Wildlife Management (volume 72) and Conservation Biology (volume 22).

We selected these journals because their target readership includes managers and conservationists, and we wished to understand our objectives within the context of the literature available to these interest groups. Papers were included in our review if (1) data reported were collected from field studies, (2) data were analyzed using multimodel inference (MMI) or statistical null hypothesis testing, and (3) management or conservation predictions or recommendations were drawn from the reported statistical analyses. We excluded commentaries, literature reviews, statistical theory papers, and papers where the objective was to theoretically develop or test a specific type of model (e.g., population growth models) without testing multiple competing statistical hypotheses. Subsequent analyses were restricted to papers that used MMI as a method of comparing hypotheses (Burnham and Anderson, 2002). In the reported results of MMI papers, we determined the number of models in a confidence set of models based on the minimum cutoff point suggested by Royall (1997) where models in the confidence set are within 10% of the Akaike weight of the top model. Models within the confidence set are considered to be the best supported given the data and the models selected for analysis. It is important to define the confidence set because these models should be taken into consideration when model averaging or dis-cussing model selection results. Where papers did not report Akaike weights, or where Akaike weights were not applicable (i.e. Schwartz's criterion (Schwarz, 1978) and deviance information criterion, (DIC) (Spiegelhalter et al., 2002)), we designated the confidence set as the set of models within 2 ΔAIC or ΔDIC of the top model (Burnham and Anderson, 2002). We categorized papers with only one model supported in all model analyses (a confidence set of one) as strong inference and papers with > 1 top model in all model analyses as weak inference. We selected this narrow definition because it most closely approximates the unequivocal conclusion of the null hypothesis test as described by Platt (1964). If some model analyses contained one top model and other analyses within the same paper contain >1 top model, we classified the paper as including both types of inference. Papers that did not provide sufficient information to determine confidence sets were catego-rized as "unknown" inference.

2.2. Management recommendations

We categorized each paper's recommendations as non-management, vague, specific, or adaptive. Some papers did not provide explicit management recommendations but predicted how factors beyond local management control (e.g., climate change, urban expansion) may change ecosystems or organisms. Vague recommendations listed how the ecosystem needed to be struc-tured or what changes needed to occur without providing managers with explicit actions to implement. Specific recommen-dations were explicit in what actions managers needed to take and how these actions would directly affect the organism or ecosystem in question. Adaptive recommendations explicitly evoke the implementation of management actions while reducing uncer-tainty through monitoring in an iterative, learning process.

2.3. Uncertainty

To determine if authors acknowledged model selection uncer-tainty, we searched each paper containing MMI for the term "uncertainty" and recorded the context in which it was used. Authors that did not use the term uncertainty or used the term outside of their model selection results were categorized as not acknowledging uncertainty. If authors mentioned uncertainty as the reason for model averaging or explicitly stated their model selection as having uncertainty, we categorized them as

acknowledging uncertainty. Although authors may have used other means to acknowledge the uncertainty in their model selection, the term "uncertainty" is the most clearly defined and least ambiguous (Regan et al., 2002). Model averaging is one way in which to deal with uncertainty without having to explicitly use the word "uncertain", so we also quantified how many papers calculated model-averaged estimates.
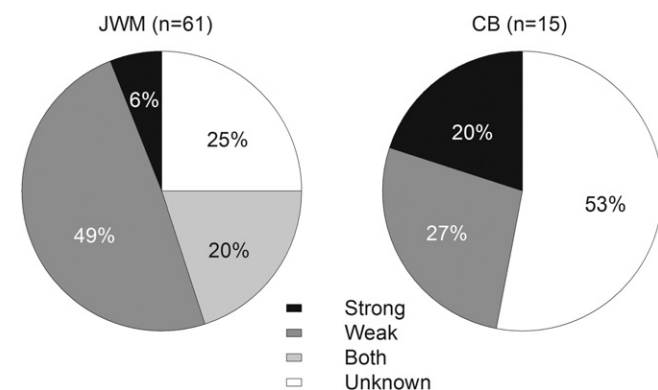
3. Results and discussion
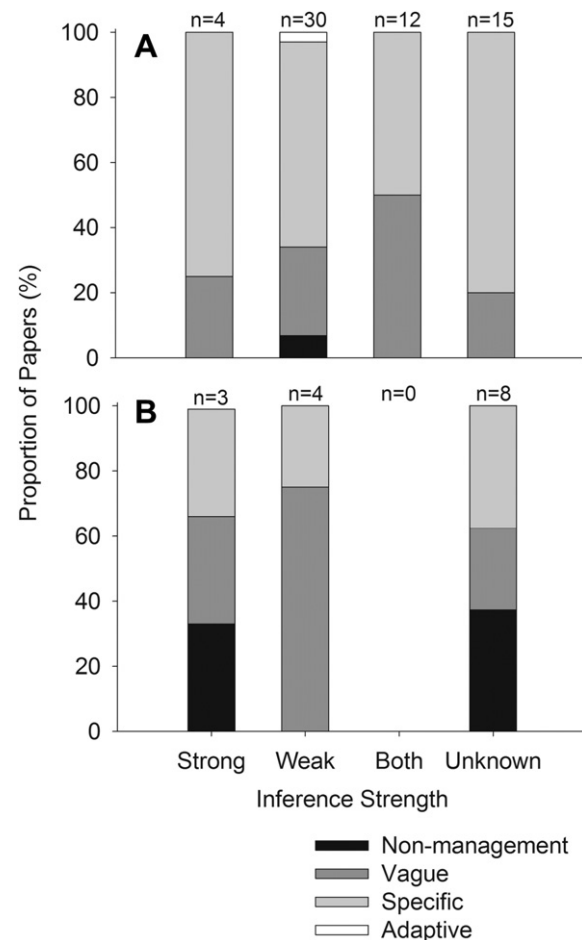
3.1. Inference strength

We reviewed 159 articles in the 2008 issues of the Journal of Wildlife Management (JWM) and 105 articles in Conservation Biology (CB) that met our specific criteria. Thirty-eight percent (61 of 159) and 14% (15 of 105) of articles in JWM and CB, respectively, utilized multimodel inference (Rehme, 2010), with model fit assessed with AIC, second-order pseudo AIC (pAIC), quasi-likelihoods AIC (qAIC), AIC adjusted for small sample sizes (AIC$_c$), Bayesian information criterion (BIC), or DIC. The majority of MMI papers contained either weak inference or did not provide sufficient information for us to determine the strength of their inference (Fig. 1).

We encountered a surprising lack of necessary information to properly understand the authors' analysis methods and results, which hindered our ability to interpret the inference strength of many of the reviewed papers. Thirty papers from both journals did not report sufficient information for us to determine what models they considered. Another set of thirty papers reported only a portion of the information needed to interpret their process for model selection. Twelve percent (9 of 76) of all MMI papers reviewed reported no means of assessing model fit (*e.g.*, AIC values or weights). Twenty-eight percent (21 of 76) of all MMI papers reported incomplete multimodel inference results (*i.e.* only the top models, some sets of models but not others).

We encourage editors and reviewers to respond to the call by Anderson et al. (2001b) to provide results of multimodel inference, such as the model set and associated AIC values and model weights. In some cases, the number of models compared or the number of different analyses were too large to reasonably report all models and corresponding information criterion outputs. However, these results could be provided in supplemental material, but no JWM

Fig. 2. Percentage of papers in each multimodel inference strength class according to the type of recommendation provided in (A) the Journal of Wildlife Management (JWM, *n* = 61) and (B) Conservation Biology (CB, *n* = 15) in 2008. Only 1 paper in either journal provided an adaptive management recommendation. No CB papers were categorized as "both" inference strength (papers that contained both strong inference analyses and weak inference analyses).

papers and only one CB paper (using an on-line supplement) provided missing information in such a manner. Many journals offer on-line resources for supplemental material, and editors should remind authors of this option, so that they may present their model selection results in full. In some cases, the number of tested models was too large to report because the authors chose to include all measured variables and almost all possible interactions. It is unlikely that every combination of variables and interactions represents a set of *plausible* models (see Anderson et al., 2001a). Trivial null hypotheses have been criticized in null hypothesis testing (Anderson and Burnham, 2002; Robinson and Wainer, 2002) and models including variable interactions with no biological basis are no less trivial (Guthery et al., 2005).

3.2. Management recommendations and uncertainty

Where inference was weak, authors in our sample often provided specific management recommendations (Fig. 2), but the majority of papers failed to acknowledge the resulting uncertainty by using the term "uncertainty" (Table 1). Due to the type of journals we selected, our results exclude journals that do not require authors to propose management recommendations. Therefore, it is possible that management recommendations following weak inference are less pervasive in journals that do not require such

Fig. 1. Percentage of multimodel inference papers in the 2008 issues of the Journal of Wildlife Management (JWM) and Conservation Biology (CB) in each multimodel inference strength class. The majority of papers contained weak inference, with >1 model in the confidence set of models, or unknown inference, where authors did not provide sufficient information to determine the confidence set. No CB papers were categorized as "both" inference strength (papers that contained both strong inference analyses and weak inference analyses).

Table 1
The use of the term "uncertainty" in relation to multimodel inference results (number that performed model averaging) in papers in Journal of Wildlife Management (JWM, $n = 61$) and Conservation Biology (CB, $n = 15$) in 2008. The majority of weak inference papers did not use the term "uncertainty", and the majority of these papers did not model average.

Use of term uncertainty	Strong[a]		Weak[b]		Both[c]		Unknown[d]		
	JWM	CB	JWM	CB	JWM	CB	JWM	CB	Total
No mention of uncertainty	3	1	23 (9)	2 (1)	10 (1)	0	14 (4)	7	60
Term used, but unrelated to model selection	1	1	0	0	1	0	1 (1)	1	5
Term used as reason for model averaging	0	0	4 (4)	1 (1)	1 (1)	0	0	0	6
Term used when talking about model selection	0	1	3 (1)	1	0	0	0	0	5
Total	4	3	30	4	12	0	15	8	76

[a] Confidence set = 1 top model.
[b] Confidence set > 1 model.
[c] Article contains both strong inference analyses and weak inference analyses.
[d] Not enough information provided to determine confidence set of models.

recommendations. However, we selected JWM and CB because they are regarded as prominent in the fields of wildlife and habitat management and are read by managers.

When specific management recommendations were suggested without acknowledging uncertainty, authors failed to provide managers and policy makers with complete information on the consequences of management decisions. Further, when authors do not acknowledge the inherent uncertainty in weak inference, they may set unrealistic expectations on the part of those adopting the management recommendations. We may have underestimated the number of papers that implicitly acknowledged the concept of uncertainty, for we focused on the explicit use of the term "uncertainty". As such, our results may be biased against authors that used an alternate term or implicitly acknowledged uncertainty. None-the-less, we feel this was the best method for taking the authors' meaning at face-value and mimicking a manager's perception of the acknowledgement of model selection uncertainty within the article. "Uncertainty" is an established term within adaptive management and is easily recognized by readers. We chose not to attempt to infer authors' implicit acknowledgement of uncertainty because a measurement of variation in subjective judgment of implicit acknowledgement was beyond the scope of this review (see Regan et al., 2002). Regardless, it is clear that authors, reviewers and editors should be open to the explicit acknowledgement of uncertainty in peer-reviewed papers so that scientists can maintain the transparency that is important to facilitate open communication between scientists and managers.

An important element of effective management of natural resources is the continuing dialog between ecologists and managers (Gunderson et al., 1995; Holling, 1978). For managers to effectively use the results of ecological field studies, managers must understand the limitations of the study so that they may properly assess risk in decision making. The appropriate level of risk for any given decision can only be evaluated by the manager and stakeholders. Therefore, ecologists should not presuppose risk is not a factor in the application of their management recommendations. Ecologists can avoid this presupposition by acknowledging any model selection uncertainty.

Burnham and Anderson (2002) advocated the use of quantitative evidence to allow decision-makers to assess what is important; authors should be encouraged to provide such evidence. Only 14% ($n = 10$) of papers in both journals that did not have a strong inference chose to use the term "uncertainty" in relation to their multimodel selection (Table 1). However, 25% of these chose to model average parameters of interest as a way of dealing with multimodel selection uncertainty. To model average, authors must select a confidence set of models across which to average parameter estimates, or they must average across all models. We identified 14 methods by which authors determined their confidence set of models (Table 2). The subjectivity with which authors selected their confidence set becomes problematic when readers wish to compare the parameters derived from model averaging among studies. Burnham and Anderson (2002) recommended that models within 2 ΔAIC of the top model be considered as competitive with the top model, models within 2–4 ΔAIC of the top model be considered as plausible, and models > 4 ΔAIC be considered unlikely. The majority of authors chose to work within this recommendation, but not all authors rationalized their reasoning behind selecting the method

Table 2
Proportion of papers in each journal according to 14 methods used by authors to select models for their confidence set of models for multimodel inference in the 2008 issues of the Journal of Wildlife Management (JWM) and Conservation Biology (CB). Akaike's information criterion (AIC) was most commonly used, though variations of AIC, Bayesian information criterion, and deviance information criterion were also employed.

Method for Determining Confidence Set	JWM ($n = 61$)	CB ($n = 15$)
Models within 2 ΔAIC of top model	33%	20%
No criteria for confidence set reported	28%	47%
Authors used weights comparatively ("best", "better", or "more weight")	5%	7%
Lowest AIC value (only one model in confidence set)	7%	13%
Lowest ΔAIC (no specific ΔAIC value provided)	5%	0
Models within 4 ΔAIC of top model	5%	0
Listed and discussed evidence ratios	5%	0
Models <2 ΔAIC of top model are "competitive", 2–4 ΔAIC are "plausible", >4 ΔAIC are "unlikely"	3%	0
Models that add up to 95% of total weight	3%	0
Models within 10% of the weight of top model	2%	7%
Models within 10% of the weight of top model or 4 best models	2%	0
Models 0–2 ΔAIC of top model have "substantial support", 4–7 ΔAIC have "considerably less support", >10 ΔAIC have essentially no support	2%	0
Models within 10 ΔAIC of top model	0	7%
Models that add up to 90% of total weight	2%	0

that they used for determining the confidence set. The ecological community needs to establish a consistent method for determining confidence sets, and editors can be a part of the solution by restricting the variability allowed among papers.

3.3. Adaptive management

Simply acknowledging uncertainty and model averaging parameters of interest does not fully solve the dilemma faced by managers and policy makers when ecological studies fail to result in strong inference. When strong inference and statistical null hypothesis testing fails or is inapplicable, adaptive inference is a logical alternative course of investigation for understanding complex ecological interactions (Holling and Allen, 2002). Multimodel inference is a tool that can be used within adaptive inference. However, adaptive inference does not solve the manager's predicament of how to continue to make management decisions when scientific investigation is weak and uncertain or still in progress. Meta-analyses can also provide better understanding within adaptive inference by coalescing weak inferences from multiple studies to build evidence. But meta-analyses are limited to topics for which there have been many independent studies. Managers who need to make decisions from one or two weak inference studies are thus at an impasse without adaptive management.

Adaptive management provides a means by which managers can move forward with management despite the uncertainty in weak inference sometimes inherent in statistical methods, including multimodel inference. We suggest that the type of recommendation must also be tailored to the strength of inference from which it is being drawn. However, only one JWM paper and zero CB papers out of our sample recommended an adaptive approach to management. This one JWM paper was classified as a weak inference paper, but, given the abundant use of MMI and the pervasiveness of weak inference, we feel increased acknowledgement of the utility of adaptive management is needed.

4. Conclusion

Our results demonstrate that weak inference is prevalent in the use of multimodel inference and that authors are failing to acknowledge the resulting uncertainty in their specific management recommendations. Authors and editors should be aware of the importance of acknowledging uncertainty both in explicit terms and through methods such as model averaging, but acknowledgement can only take us so far. We suggest that editors must be open to not requiring specific management recommendations from authors when the research results do not permit strong inference. However, when management recommendations are required, adaptive management is an ideal method for dealing with uncertainty resulting from weak inference.

The strength in adaptive management is that it is a method that can be used despite uncertainty and weak inference and permits the continuation of management in such situations, without spurious certitude. Continued management provides information about the system that reduces uncertainty and improves future management decisions. Working scientists and resource managers can interact transparently and more effectively to move forward in understanding the ecosystem in question when they are open about the uncertainty, and adaptive management provides a framework in which to do this.

However, merely adding the words "adaptive management" to any set of management recommendations is not enough. Even if their inference is weak, authors can continue to draw conclusions and develop hypotheses from their results. We suggest that authors consider how these hypotheses might be incorporated into and

tested using an adaptive management plan. The strength of adaptive management is the ability to take uncertainty about hypotheses or processes and build a management plan that works toward the reduction of uncertainty in the underlying ecological processes and the effects of management actions. In this way, weak inference and the resulting management recommendations can still be useful to managers and policy makers through adaptive management.

References

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

Anderson, D.R., Burnham, K.P., Gould, W.R., Cherry, S., 2001a. Concerns about finding effects that are actually spurious. Wildl. Soc. Bull. 29, 311–316.

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

Anderson, D.R., Link, W.A., Johnson, D.H., Burnham, K.P., 2001b. Suggestions for reporting the results of data analyses. J. Wildl. Manage. 65, 373–378.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: a Practical Information-theoretic Approach, second ed. Springer-Verlag, New York.

Gunderson, L.H., Holling, C.S., Light, S.S., 1995. Barriers broken and bridges built: a synthesis. In: Gunderson, L.H., Holling, C.S., Light, S.S. (Eds.), Barriers and Bridges to the Renewal of Ecosystems and Institutions. Columbia University Press, New York, pp. 489–532.

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

Hilborn, R., Stearns, S.C., 1982. On inference in ecology and evolutionary biology: the problem of multiple causes. Acta Biotheor. 31, 145–164.

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

Holling, C.S., Allen, C.R., 2002. Adaptive inference for distinguishing credible from incredible patterns in nature. Ecosystems 5, 319–328.

Holling, C.S., 1978. Adaptive Environmental Assessment and Management. John Wiley and Sons, Chichester.

Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108.

McFadden, J.E., Hiller, T.L., Tyre, A.J., 2011. Evaluating the efficacy of adaptive management approaches: is there a formula for success? J. Environ. Manage. 92 (5), 1354–1359.

Platt, J.R., 1964. Strong inference. Science 146, 347–353.

Pickett, S.T.A., Kolasa, J., Jones, C.G., 1994. Ecological Understanding: the Nature of Theory and The Theory of Nature, first ed. Academic Press, San Diego.

Rehme, S.E., 2010. Ecology, stable isotopes, and management of grassland songbirds at National Park Service properties on the Great Plains. Master's Thesis, University of Nebraska – Lincoln.

Regan, H.M., Colyvan, M., Burgman, M.A., 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecol. Appl. 12, 618–628.

Robinson, D.H., Wainer, H., 2002. On the past and future of null hypothesis significance testing. J. Wildl. Manage. 66, 263–271.

Royall, R.M., 1997. Statistical Evidence: a Likelihood Paradigm. Chapman and Hall, New York.

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

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. Roy. Stat. Soc. B 64, 583–639.

Stephens, P.A., Buskirk, S.W., Martínez del Rio, C., 2007. Inference in ecology and evolution. Trends Ecol. Evol. 22, 192–197.

Walters, C.J., 1986. Adaptive Management of Renewable Resources. Macmillan Publishing Company, New York.

Wiens, J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3, 385–397.