Schneider, M. K., R. Law, and J. B. Illian. 2006. Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modelling. Journal of Ecology 94:310–321.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. Wiley, New York, New York, USA.

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

Ulanowicz, R. E. 1988. On the importance of higher-level models in ecology. Ecological Modelling 43:45–56.

Wikle, C. K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84:1382–1394.

Wikle, C. K., L. M. Berliner, and N. Cressie. 1998. Hierarchical Bayesian space-time models. Environmental and Ecological Statistics 5:117–154.

Ecological Applications, 19(3), 2009, pp. 581–584

Bayesian methods for hierarchical models: Are ecologists making a Faustian bargain?

SUBHASH R. LELE[1,3] AND BRIAN DENNIS[2]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1 Canada*
[2]*Department of Fish and Wildlife Resources and Department of Statistics, University of Idaho, Moscow, Idaho 83844-1136 USA*

It is unquestionably true that hierarchical models represent an order of magnitude increase in the scope and complexity of models for ecological data. The past decade has seen a tremendous expansion of applications of hierarchical models in ecology. The expansion was primarily due to the advent of the Bayesian computational methods. We congratulate the authors for writing a clear summary of hierarchical models in ecology.

While we agree that hierarchical models are highly useful to ecology, we have reservations about the Bayesian principles of statistical inference commonly used in the analysis of these models. One of the major reasons why scientists use Bayesian analysis for hierarchical models is the myth that for all practical purposes, the only feasible way to fit hierarchical models is Bayesian. Cressie et al. (2009) do perfunctorily mention the frequentist approaches but are quick to launch into an extensive review of the Bayesian analyses. Frequentist inferences for hierarchical models such as those based on maximum likelihood are beginning to catch up in ease and feasibility (De Valpine 2004, Lele et al. 2007). The recent "data cloning" algorithm, for instance, "tricks" the Bayesian MCMC setup into providing maximum likelihood estimates and their standard errors (Lele et al. 2007). A natural question that a scientist should ask is: if one has a hierarchical model for which full frequentist (say, based on ML estimation) as well as Bayesian inferences are available, which should be used and why? This can only be answered based on the philosophical underpinnings. The convenience criterion, commonly used to justify the use of Bayesian approach, no longer applies, given the recent advances in the frequentist computational approaches. Although the Bayesian computational algorithms made statistical inference for such models possible, it is not clear to us that the Bayesian inferential philosophy necessarily leads to good science. In return for seemingly confident inferences about important quantities in the face of poor data and vast natural uncertainties, are ecologists making a Faustian bargain?

We begin by stating our reservations about the scientific philosophy advocated by the authors. The authors claim that modeling is for synthesis of information (Cressie 2009). Furthermore, they claim that subjectivity is unavoidable. We completely disagree with both these statements. In our opinion, the fundamental goal of modeling in science is to understand the mechanisms underlying the natural phenomena. Models are quantitative hypotheses about mechanisms that help us connect our prospective understandings to observable phenomena. Certainly in the sociology of the conduct of science, subjectivity often enters in the array of mechanisms hypothesized. However, good scientists are trained rigorously toward considering as many alternative explanations as imagination allows, with the ultimate filters being consistency with observation, experiment, and previous reliable knowledge. Introducing more subjectivity into the process of acquiring reliable knowledge introduces confounding factors in the empirical filtering of hypotheses, and so as scientists, we should be striving to reduce subjectivity instead of increasing it. Just because there is subjectivity in hypothesizing mechanisms, it does not give us a free

Manuscript received 18 March 2008; revised 8 July 2008; accepted 24 July 2008. Corresponding Editor: N. T. Hobbs. For reprints of this Forum, see footnote 1, p. 551.
[3] E-mail: slele@ualberta.ca

pass to introduce subjectivity in testing those mechanisms against the data.

Obtaining hard, highly informative data requires substantial resources and time. Can expert opinion play a role in inference? Eliciting prior distributions from experts for use in Bayesian statistical analysis has often been suggested for incorporating expert knowledge. However, eliciting priors is more art than science. Aside from this operational difficulty, a far more serious problem is in deciding "who is an expert." In the news media, "experts" are available to offer sound bites favoring any side of any issue. As scientists, we might insist that expert opinion should be calibrated against reality. Furthermore, the weight the expert opinion receives in the statistical analysis should be based on the amount and quality of information such an expert brings to the table. Bayesian analysis lacks such explicit quantification of the expertness.

We do believe that expert opinion can and should be brought into ecological analyses (Lele and Das 2000, Lele 2004), although not by Bayesian methods. Recently, Lele and Allen (2006) showed how expert opinion can be incorporated using a frequentist framework by eliciting data instead of a prior. The methodology suggested in Lele and Allen (2006) automatically weighs expert opinion and hard data according to their Fisher information about the parameters of the process. The expert who brings in only "noise" and no information automatically gets zero weight in the analysis. Provided the expert opinion is truly informative, the confidence intervals and prediction intervals after incorporation of such opinion are shown to be shorter than the ones that would be obtained without its inclusion. It is thus possible to incorporate expert opinion under the frequentist paradigm.

Hierarchical models are attractive for realistic modeling of complexity in nature. However, as a general principle, the complexity of the model should be matched by the information content of the data. As the number of hierarchies in the model increases, the ratio of information content to model complexity necessarily decreases.

Unfortunately expositions of Bayesian methods for hierarchical models have tended to emphasize practice while deemphasizing the inferential principles involved. Potential consequences of ignoring principles can be severe when such data analyses are used in policy making. In the following, we discuss some myths and misconceptions about Bayesian inference.

"Flat" or "objective" priors lead to desirable frequentist properties.—Many applied statisticians and ecologists believe that flat or non-informative priors produce Bayesian credible intervals that have properties similar to the frequentist confidence intervals. This idea has been touted by the advocates of the Bayesian approach as an important point of reassurance. An irony in this claim is that Bayesian inference seems to be justified under frequentist principles. However, the claim is plain wrong. The Bayesian credible intervals obtained under flat priors can have seriously incorrect frequentist coverage properties; the actual coverage can be substantially smaller or larger than the nominal coverage (Mitchell 1967, Heinrich 2005; D. A. S. Fraser, N. Reid, E. Marras, and G. Y. Yi, *unpublished manuscript*). A practicing scientist should ask: Under Bayesian inferential principles, what statistical properties are considered desirable for credible intervals and how does one assure them in practice?

Non-informative priors are unique.—Many ecologists believe that the flat priors are the only kind of distributions that are "objective" priors. In fact, in Bayesian statistics the definition of non-informative prior has been under debate since the 1930s with no resolution. There are many different types of proper and improper distributions that are considered "non-informative." We highly recommend that ecologists read Chapter 5 of Press (2003) and Chapter 6 of Barnett (1999) for a non-mathematical, easily accessible discussion on the issue of non-informative priors. For a quick summary, see Cox (2005). Furthermore, it is also known that different "non-informative" priors lead to different posterior distributions and hence different scientific inferences (e.g., Tyul et al. 2008). The claim that use of non-informative priors lets the data speak is flatly incorrect. A scientist should ask: Which non-informative priors should ecologists be using when analyzing data with hierarchical models?

Credible intervals are more understandable than confidence intervals.—The "objective" credible intervals (i.e., formed with flat priors) do not have valid frequentist interpretation in terms of coverage. For subjective Bayesians, the interpretation of the coverage is in terms of "personal belief probability". Neither of these interpretations is valid for the "objective" Bayesian analysis. A scientist should ask: What is the correct way to interpret "objective" credible intervals?

Bayesian prediction intervals are better than frequentist prediction intervals.—Hierarchical models enable the researcher to predict the unobserved states of the system. The naïve frequentist prediction intervals, where estimated parameter values are substituted as if they are true parameter values, are correctly criticized for having incorrect coverage probabilities. These intervals can be corrected using bootstrap techniques (e.g., Laird and Louis 1987). Such corrected intervals tend to have close to nominal coverage. It is known that "objective" Bayesian credible intervals for parameters do not have correct frequentist coverage. A scientist should ask: Are prediction intervals obtained under the "objective" Bayesian approach guaranteed to have correct frequentist properties? If not, how does one interpret the "probability" represented by these prediction intervals?

As sample size increases, the influence of prior decreases rapidly.—For models that have multiple parameters, it is usually difficult to specify non-informative priors on all the parameters. The Bayesian scientists put non-informative priors on some of the parameters and informative priors on the rest. Dennis (2004) showed that the influence of an informative prior can decrease extremely slowly as the sample size increases when there are multiple parameters. We point out that the hierarchical models being proposed in ecology sometimes have enormous numbers of parameters. A scientist should ask: How does one evaluate the influence of prior specification on the final inference?

The problem of identifiability can be surmounted by specifying informative priors.—By definition, no experimental data generated by the hierarchical model in question can ever provide information about the non-identifiable parameters. If this is the case, then how can anyone have an "informed" guess about something that can never be observed? Furthermore, as noted above such "informative" priors can be inordinately influential, even for large samples. A scientist should ask: How exactly does one choose a prior for a non-identifiable parameter, and in what sense is specifying an informative prior a desirable solution for surmounting non-identifiability?

Influence of priors on the final inference can be judged from plots of the marginal posterior distributions.—The true influence of the priors on the final inference is manifested in how much the joint posterior distribution differs from the joint prior. In practice, however, influence of priors is judged by looking at the plots of the marginal priors and posteriors. The marginalization paradox (Dawid et al. 1973) suggests that the practice could fail in spectacular ways. A scientist should ask: How should one judge the influence of the specification of the prior distribution on the final inference?

Bayesian posterior distributions can be used for checking model adequacy.—Cressie et al. (2009) correctly emphasize the importance of checking for model adequacy in statistical inference. They suggest using goodness of fit type tests on the Bayesian predictive distribution. A scientist should ask: If such a test suggests that the model is inadequate, how does one know if the error is in the form of the likelihood function or in the specification of the prior distribution?

Reporting sensitivity of the inferences to the specification of the priors is adequate for scientific inference.—Cressie et al. (2009), as well as many other Bayesian analysts, suggest that one should conduct sensitivity of the inferences to the specification of the priors. In our opinion, it is not enough to simply report that the inferences are sensitive. Inferences are guaranteed to be sensitive to some priors and guaranteed to be non-sensitive to some other priors. What is needed is a suggestion as to what should be done if the inferences are sensitive. A scientist should ask: If the inferences prove sensitive to the particular choice of a prior, what recourse does the researcher have?

Scientific method is better served by Bayesian approaches.—One of the most desirable properties of a scientific study is that it be reproducible (Chang 2008). The frequentist error rates, either the coverage probabilities or probabilities of weak and misleading evidence (Royall 2000), inform the scientists about the reproducibility of their results. A scientist should ask: How does one quantify reproducibility when using "objective" Bayesian approach?

In summary, we applaud the authors for furthering the discussion of use of hierarchical models in ecology. While hierarchical models are useful, we cannot avoid questioning the quality of inference that results from hierarchy proliferation. Royall (2000) quantifies the concept of weak inference in terms of probability of weak evidence where, based on the observed data, one cannot distinguish between different mechanisms with enough confidence. We surmise that the probability of weak evidence becomes unacceptably large as the number of unobserved states increases. We feel that scientists should be wary of introducing too many hierarchies in the modeling framework. Furthermore, we also have difficulties with the Bayesian approach commonly used in the analysis of hierarchical models. We suggest that the Bayesian approach is neither needed nor is desirable to conduct proper statistical analysis of hierarchical models. The alternative approaches based on the frequentist philosophy of science should be considered in analyzing the hierarchical models.

Literature Cited

Barnett, V. 1999. Comparative statistical inference. John Wiley, New York, New York, USA.

Chang, K. 2008. Nobel winner retracts research paper. New York Times, 10 March 2008.

Cox, D. R. 2005. Frequentist and Bayesian statistics: a critique. *In* L. Louis and M. K. Unel, editors. Proceedings of the Statistical Problems in Particle Physics, Astrophysics and Cosmology. Imperial College Press, UK. ⟨http://www.physics.ox.ac.uk/phystat05/proceedings/default.htm⟩

Cressie, N. A. C., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19:553–570.

Dawid, A. P., M. Stone, and J. V. Zidek. 1973. Marginalization paradoxes in Bayesian and structural inference. Journal of the Royal Statistical Society B 35:189–233.

Dennis, B. 2004. Statistics and the scientific method in ecology. Pages 327–378 *in* M. L. Taper and S. R. Lele, editors. The nature of scientific evidence: statistical, empirical and philosophical considerations. University of Chicago Press, Chicago, Illinois, USA.

De Valpine, P. 2004. Monte Carlo state-space likelihoods by weighted kernel density estimation. Journal of the American Statistical Association 99:523–536.

Heinrich, J. 2005. The Bayesian approach to setting limits: what to avoid. *In* L. Louis and M. K. Unel, editors. Proceedings of the Statistical Problems in Particle Physics, Astrophysics and Cosmology. Imperial College Press, UK. ⟨http://www.physics.ox.ac.uk/phystat05/proceedings/default.htm⟩

Laird, N. M., and T. A. Louis. 1987. Empirical Bayes confidence intervals based on bootstrap samples. Journal of the American Statistical Association 82:739–750.

Lele, S. R. 2004. Elicit data, not prior: on using expert opinion in ecological studies. Pages 410–436 *in* M. L. Taper and S. R. Lele, editors. The nature of scientific evidence: statistical, empirical and philosophical considerations. University of Chicago Press, Chicago, Illinois, USA.

Lele, S. R., and K. L. Allen. 2006. On using expert opinion in ecological analyses: a frequentist approach. Environmetrics 17:683–704.

Lele, S. R., and A. Das. 2000. Elicited data and incorporation of expert opinion for statistical inference in spatial studies. Mathematical Geology 32:465–487.

Lele, S. R., B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecology Letters 10:551–563.

Mitchell, A. F. S. 1967. Discussion of paper by I. J. Good. Journal of the Royal Statistical Society, Series B 29:423–424.

Press, S. J. 2003. Subjective and objective Bayesian statistics. Second edition. John Wiley, New York, New York, USA.

Royall, R. 2000. On the probability of observing misleading statistical evidence. Journal of the American Statistical Association 95:760–768.

Tyul, F., R. Gerlach, and K. Mengersen. 2008 A comparison of Bayes–Laplace, Jeffreys, and other priors: the case of zero events. American Statistician 62:40–44.

Ecological Applications, 19(3), 2009, pp. 584–588

Shared challenges and common ground for Bayesian and classical analysis of hierarchical statistical models

Perry de Valpine[1]

Department of Environmental Science, Policy and Management, 137 Mulford Hall #3114, University of California Berkeley, Berkeley, California 94720 USA

Let me begin by welcoming the paper by Cressie et al. (2009) as an insightful overview of hierarchical statistical modeling that will be valuable from both Bayesian and classical perspectives. Statistical conclusions hinge on the appropriateness of the mathematical models used to represent hypotheses, and Cressie et al. explain many merits of hierarchical models. My comments will highlight the shared challenges and common ground of Bayesian and classical analysis of hierarchical models; give a likelihood theory complement to Cressie et al.'s explanations of their merits; summarize methods for maximum likelihood and "empirical Bayes" estimation and incorporation of uncertainty; and explore some of the claims of Bayesian advantages and classical limitations.

Both Bayesian and classical analyses can and must address the inherent challenges of inference and prediction from noisy data, including comparing models, selecting a best model or combination of models, deciding if a model is acceptable at all, avoiding overfitting and statistical "fishing" or "data dredging," and incorporating uncertainty. In practice, an appropriate modeling framework trumps many other issues, including choice of Bayesian or classical analysis.

The difference between Bayesian and classical philosophies is that Bayesian analysis uses the mathematics of probability distributions for model parameters, P as defined by Cressie et al., while classical analysis does not. It is generally agreed that Bayesian analysis must define "probability of P" as "degree of belief for P" (or "subjective probability of P" or other synonymous terms), whereas classical "probability" refers to a frequency of outcomes over a long run (O'Hagan 1994). Bayesian analysis makes "probability" statements about hypotheses, given data, that are formally weighted degree-of-belief comparisons, while classical analysis makes statements about frequencies with which different hypotheses would have produced the data and, in some approaches, hypothetical unobserved data. Classical analysis encompasses much more than Neyman-Pearson and/or Fisher hypothesis testing (which have been critiqued for both their actual logical limitations when correctly interpreted and their potential for misinterpretation; see Mayo and Spanos [2006] for extensions of NP logic). For example, model comparison using Akaike's Information Criterion (AIC; Burnham and Anderson 2002) takes a classical approach to parameters. While Bayesian and classical approaches are philosophically different, one can interpret results from them in tandem (Efron 2005).

It is worth emphasizing that choosing a hierarchical model is separate from choosing a Bayesian or classical approach for parameters. In a hierarchical model, one has parameters (P, either Bayesian or classical) that define the distribution of unknown ecological states (E) that define the distribution of data (D). I will call E the latent variables and/or random effects, which are among the many names for random variables whose values are not directly known but in theory define relationships among data values, to emphasize that hierarchical models are mixed models (McCulloch and Searle

Manuscript received 20 March 2008; revised 8 July 2008; accepted 24 July 2008. Corresponding Editor: N. T. Hobbs. For reprints of this Forum, see footnote 1, p. 551.

[1] E-mail: pdevalpine@nature.berkeley.edu