Ecology, 90(1), 2009, pp. 265–277

Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models

Samantha Low Choy,[1] Rebecca O'Leary, and Kerrie Mengersen

School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

Abstract. Bayesian statistical modeling has several benefits within an ecological context. In particular, when observed data are limited in sample size or representativeness, then the Bayesian framework provides a mechanism to combine observed data with other "prior" information. Prior information may be obtained from earlier studies, or in their absence, from expert knowledge. This use of the Bayesian framework reflects the scientific "learning cycle," where prior or initial estimates are updated when new data become available. In this paper we outline a framework for statistical design of expert elicitation processes for quantifying such expert knowledge, in a form suitable for input as prior information into Bayesian models. We identify six key elements: determining the purpose and motivation for using prior information; specifying the relevant expert knowledge available; formulating the statistical model; designing effective and efficient numerical encoding; managing uncertainty; and designing a practical elicitation protocol. We demonstrate this framework applies to a variety of situations, with two examples from the ecological literature and three from our experience. Analysis of these examples reveals several recurring important issues affecting practical design of elicitation in ecological problems.

Key words: design; ecology; expert elicitation; framework; informative Bayesian analysis; prior information; protocol.

Introduction

The benefits of Bayesian statistical modeling for ecological applications are now well established (Ellison 1996, 2004). One major benefit is that the Bayesian approach embodies a natural cycle of learning that is well suited to the ecological context (Wade 2001). It provides a framework where current knowledge can be updated by new information, so that the results (posterior) of one study can be used as the starting point (prior) for the next study. This representation of the scientific method permits more accelerated and integrated assessments: by enabling a series or sequence of smaller studies to replace a single large scale study (Fleischman and Burwen 2003, McCarthy and Masters 2005), or allowing progressive improvements to study, data design or measurement (Cummings et al. 2002, Chao 2003).

As for frequentist statistics, the Bayesian learning cycle starts with formulation of a statistical model, known as the likelihood $p(y\,|\,\theta)$, which describes the chance of observing data y given a model with parameters θ. What is often required, however, is a reverse of this logic (Crome et al. 1996, Wade 2001, Prato 2005). This is called the posterior $p(\theta\,|\,y)$ and instead provides a basis for inference about θ while conditioning on the data y that have been observed. This is useful for preliminary inferences based on limited data, such as non-replicated experiments which are common in ecology (Reckhow 1990). The reversal is achieved via Bayes theorem:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{1}$$

but requires specification of a prior distribution $p(\theta)$ of the model parameters. Priors reflect information about parameters, which must be independent of the observed data y that are used to construct the likelihood (ter Braak and Etienne 2003).

Initially, priors can be constructed as an end in themselves (e.g., Alho et al. 1996) to represent the current state of knowledge (e.g., Low Choy et al. 2005), or to inform design of data collection (Kadane 1990). Prior knowledge can be incorporated from previous experimentation (Anholt et al. 2000), meta-analysis across previous studies (e.g., Link and Sauer 1996), or expert knowledge (Garthwaite and O'Hagan 2000). Bayesian modeling with informative priors based on expert opinion can provide a useful "bridge" for ecologists, from purely conceptual models to statistical models that are calibrated to observed data. Conceptual models, where relationships and/or parameters are posited by experts but not validated against data, are widespread in ecology: multiple criterion analysis for habitat scoring (e.g., Roloff and Kernohan 1999) or conservation prioritization (e.g., Clark et al. 2006); sediment transport modeling (Merritt et al. 2003); bioregionalization (e.g.,

Manuscript received 29 November 2007; revised 12 May 2008; accepted 14 May 2008. Corresponding Editor: A. M. Ellison.

[1] E-mail: S.LowChoy@qut.edu.au

Accad et al. 2005); estimating parameters for deterministic models (e.g., Boyce et al. 2005).

There has been considerable debate about using subjective opinion to construct priors (Cox 2000, O'Hagan et al. 2006). Indeed, the use of prior information from previous studies is considered by some to be the only "objective" approach to Bayesian modeling (Clyde 1999, Hobbs and Hilborn 2006). However, representation of probabilities and uncertainty under both Frequentist and Bayesian paradigms share a subjective element (Lindley 2000, Dawid 2004), and other choices such as model and data are similarly subjective (Pearce et al. 2001, Ferrier et al. 2002, Burgman 2004). An advantage of the Bayesian framework is that it requires subjective information in the form of priors to be stated explicitly and precisely before modeling (Wintle et al. 2003). This avoids the possibility of expert opinion being used post-hoc to modify model results or tuning parameters, which can be of limited benefit (Pearce et al. 2001).

Statistical design underlying elicitation of this prior information from experts is therefore a crucial step. In this paper, we describe the steps involved in designing elicitation, we illustrate these steps briefly using two examples from the literature, and in more depth using three case studies from our recent experience. These case studies reveal several issues commonly encountered during design of elicitation in ecological contexts.

Elicitation

In Bayesian statistical modeling, "expert elicitation" refers to the process of obtaining expert opinion, together with uncertainty, which is then carefully formulated into informative prior distributions (Spetzler and Staël von Holstein 1975, O'Hagan et al. 2006). The main steps involved in elicitation *as experienced by the expert* are well documented (Hogarth 1975, Spetzler and Staël von Holstein 1975, Shepherd and Kirkwood 1994, Garthwaite and O'Hagan 2000, Clemen and Reilly 2001, Renooij 2001, Walls and Quigley 2001, Jenkinson 2005). We are more concerned in this paper with the main steps required by the statistician, in particular initial efforts required to *design elicitation*. Several steps are involved:

- E1) Determine purpose and motivation for using prior information (*Introduction*).
- E2) Specify available prior knowledge from experts or other sources, to define an appropriate and achievable goal of elicitation.
- E3) Formulate a statistical model representing the ecological conceptual model. Define the likelihood $p(y \mid \theta)$ characterizing the data model; and the prior $p(\theta)$ reflecting available prior knowledge.
- E4) Design numerical encoding (measurement technique) for effective elicitation of prior information and representation as a statistical distribution.
  - a) Make a crucial decision: whether a structural (direct) or indirect approach to elicitation is more appropriate.
  - b) Select summary statistics to be elicited, their order and appropriate units.
  - c) Determine communication method(s) appropriate for eliciting required information (visual, tabular, verbal including wording of questions, and so on).
  - d) Specify the estimation method to calculate prior model parameters from elicited information; this is especially important for indirect methods.
  - e) Determine the relative weight of the prior and the likelihood, e.g., by specifying effective prior sample size or other function of prior variance.
- E5) Manage uncertainty for accurate and robust elicitation.
  - a) Consider eliciting from multiple experts.
  - b) Condition experts to potential biases.
  - c) Design elicitation to minimize significant biases and verify elicited information.
  - d) Perform sensitivity analysis to assumptions (e.g., priors) governing elicitation.
  - e) Validate elicited results, e.g., via provision of feedback or calibration.
- E6) Design an elicitation protocol to manage logistics of implementing elicitation.
  - a) Select and motivate contributors and other sources of prior information.
  - b) Determine relative contribution of experts and other sources, e.g., by pooling or more sophisticated methods such as meta-analysis.
  - c) Choose delivery mechanisms, e.g., questionnaires, individual interview, panels.
  - d) Carefully design preparation of experts, including motivation and training.
  - e) Consider judicious use of technology to facilitate or streamline delivery.

We briefly define these steps. Various motivations (E1) for including prior information were reviewed in the *Introduction*. Key questions are whether this prior information provides the basis for standalone analysis or combination, with existing or yet-to-be-collected data, within a Bayesian statistical or other model? Is expert knowledge the sole or best source of information available, or does it complement information gaps in existing data sets?

The goal (E2) focuses on expert knowledge available. Consider the application of logistic regression (E3) to model species occurrence based on habitat (*Elicitation in Ecology*, case C [below]). Several elicitation methods are available (O'Leary et al. 2008*a*), each based on different expert knowledge. Experts could be asked directly to estimate the effects of habitat factors (Fleishman et al. 2001), or more simply to indicate whether these factors simply increase, decrease or have no effect on the response (Kuhnert et al. 2005, Martin et al. 2005, O'Leary et al. 2008*a*). Those familiar with species response curves could be asked to plot probability of presence (median and quartiles; see Plate 1) against values of a single habitat factor, assuming that all other factors are held constant at their optimum (Kynn 2006) or at some other reference values (Al-Awadhi and Garthwaite 2006). Alternatively experts could be asked to predict probability of presence at sites given their covariate values (Denham and Mengersen 2007; A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*). Other models (E3) with expert-defined priors include: Bayesian classification trees with elicitation of rules for habitat suitability (O'Leary et al. 2008*b*), or Bayesian Networks with elicitation of a hierarchy of relationships relating the probability of presence to habitat factors (Smith et al. 2007).

By mathematical definition, the posterior balances both prior information and knowledge gained from the data, since very strong prior information can be downweighted by a weak likelihood based on the data, or vice versa. Thus the distribution of the prior (E3) is an important aspect of model formulation (E3). Conjugate choices are popular and ensure the same distribution for the prior and posterior, improving interpretability and mathematical tractability (e.g., Pascual and Kareiva 1996). However care is required in specification of conjugate priors, particularly of variance parameters (Gelman 2006). Non-conjugate choices are also available and may more closely represent the information or assumptions about the parameters of interest (e.g., Kadane 1990). Non-informative priors have minimal contribution compared to the data; these include Jeffreys' priors (Jeffreys 1961) and objective priors (Box and Tiao 1973). In this paper, we focus on informative priors, which do convey some a priori information about θ.

Encoding methods (E4) underpinning elicitation can be divided into direct and indirect approaches (Winkler 1967, Kadane 1980, Weber and Borcherding 1993) (E4a). Direct approaches ask experts directly about parameters in the model, so experts not only require adequate statistical understanding of the role of parameters in the underlying model, but their knowledge should also be easily communicated in this way. This "analytic" approach suits experts who have analyzed their knowledge in this way previously. In contrast, indirect approaches take a more "holistic" approach by asking experts only about what they have observed. This typically involves asking experts to predict the response given particular scenarios, e.g., in a regression for known covariate values, or alternatively to impute values of covariates corresponding to particular responses. Predictive elicitation may reduce motivational biases (E5) since the expert is not necessarily aware of the link between their answers and the encoded prior distribution.

PLATE 1. Young male brush-tailed rock-wallaby at home on steep sedimentary escarpment in northeast New South Wales, Australia. Experts were asked to assess their probability of presence in different environments, which was formulated as a prior distribution for input into a Bayesian habitat model. Photo credit: Justine Murray.

Encoding (E4) is the process of translating elicited opinion into statistical statements (E3), and is therefore highly dependent on the goal of elicitation (E2). A plethora of encoding practices (e.g., O'Hagan et al. 2006) have been tailored to particular situations (Kadane and Wolfson 1998). Common approaches elicit quantiles at fixed probabilities or alternately elicit probabilities of fixed quantiles (Spetzler and Staël von Holstein 1975, O'Hagan 1998) (E4b). Hybrid methods oscillate between these two approaches (Dickey and Jiang 1998). Other summary statistics (E4b) may be elicited, such as moments and the mode or changes to estimates in light of hypothetical new information (Winkler 1967). Once the summary statistics about the unknown quantity θ have been quantified using expert knowledge (E4b), then it is necessary to estimate the prior distribution about θ (E4d). In most cases additional information about expert uncertainty is required, such as the equivalent sample size (Winkler 1967) of their knowledge (E4e), in order to estimate the variance of prior distributions with more than two parameters.

Elicitation is potentially subject to several sources of uncertainty (E5), including biases, being *conscious and* *subconscious discrepancies between the subject's responses and an accurate description of his underlying knowledge* (Spetzler and Staël von Holstein 1975). Accurate elicitation can be managed before elicitation, via design, or after elicitation, via verification. Experts can be conditioned to be aware of common biases so that they are more easily avoided (E5b). These include: displacement bias when experts over- or underestimate; variability bias or conservatism where typically experts underestimate the variability in the quantity of interest; and motivational biases due to the expert's lack of neutrality (Spetzler and Staël von Holstein 1975, Tversky and Kahneman 1981).

Cognitive biases attributable to misunderstanding what is required can also be managed via design of elicitation. Biases where experts anchor or adjust estimates with respect to available values (Tversky and Kahneman 1974) can be minimized by choosing the order of questions (E5c) (Garthwaite and Dickey 1985, Phillips and Wisbey 1993). Representativeness bias occurs when the probability of an event is confused with its representativeness or similarity to some major characteristic in the population, e.g., confusing variability in the population with accuracy of the average (Kahneman and Tversky 1972). This can be reduced by structuring the elicitation to avoid tacit conditioning on other quantities (E5c) (Spetzler and Staël von Holstein 1975). Tacit assumptions are often involved when experts estimate extreme probabilities near zero or one (Siu and Kelly 1998). These may arise from implicit conditioning or other representativeness biases. These biases may be minimized by asking experts to list reasons for elicited probabilities, especially extremely high or low values (E5c, E4c) (Kynn 2008).

Verification of elicited quantities can be achieved using various methods (E5e). The use of technology (E6e) can help avoid coherence biases by automatically checking and alerting for logical inconsistencies (E5c, E5e) (Kadane and Wolfson 1998). Feedback is useful for helping experts maintain self-consistency (E5c, E5e) (Spetzler and Staël von Holstein 1975, Kynn 2006). Calibration, where expert estimates are compared to actual results (Dawid 1982, Gneiting and Raftery 2007), can be immaterial if it is accepted that "what is being elicited is *expert,* not perfect, opinions, and thus they should not be adjusted" (Kadane and Wolfson 1998), or that expert knowledge captures information not adequately represented by observed data (E5e). Using different encoding techniques, and then comparing or combining prior or posterior models, may also improve accuracy (E5d, E5e) (Gavasakar 1988, Accad et al. 2005, Denham and Mengersen 2007, O'Leary et al. 2008*a*). Questioning multiple experts can lead to more representative results (E5a), although Clemen and Winkler (1999) caution that in some cases groups may only marginally outperform individuals, and the "best" experts may outperform the group (E5d, E5e).

Information elicited can be affected by how the elicitor communicates what is required, and how the expert communicates their knowledge. Thus exact wording of questions should be selected carefully, and define clearly and precisely what is required, minimizing ambiguity (E5b, E4c). In particular questions should refer to units and scales familiar to the expert (Spetzler and Staël von Holstein 1975) (E4c). Where possible it is advisable to refer to frequencies instead of probabilities as these are generally more accurately elicited (Kynn 2008: Recommendations 4,6). This has promoted investigation of visual and interactive approaches (Kadane et al. 1980, Kynn 2006; Appendix) (E4c).

Reporting an elicitation protocol (E6), like a survey protocol, ensures a transparent, repeatable and therefore scientifically justifiable process. Such details facilitate quality control and enable peer review, both essential for a robust process (E5e). The protocol should detail elicitation design especially major issues of practical implementation and logistical decisions, including expert selection, delivery, efficiency, and preparation. Motivational and selection biases are often unavoidable, but can be managed if known. Selection of experts impacts on representativeness, accuracy and credibility of elicited opinions (E6a). Pooling expert opinion, or hierarchical models such as meta-analysis, can be used to reweight expert opinions when combined (Clemen and Winkler 1999) (E6b).

Adequate preparation (E6d) contributes greatly to consistency and reliability of elicited quantities, and should include motivating experts to participate with diligence, training them in relevant concepts (Kynn 2008: Recommendation (1) including a "dry run" of elicitation, and conditioning them to potential biases (E5b). In addition experts may be prepared by gathering and listing all potential sources of expertise (experience, literature, and so on), which can help reduce availability bias, being the tendency to recall recent or important information (Tversky and Kahneman 1973). The delivery (E6c) of elicitation should be tailored to the experts and logistical constraints, and can range from a questionnaire (either by interview or post) to technology-assisted delivery (O'Leary et al. 2008*a*) (E6e). For more details on software tools see the Appendix. Overall the elicitation process should minimize effort required by the expert to reduce potential for fatigue and therefore inaccuracy (E5c) (Spetzler and Staël von Holstein 1975).

Elicitation in Ecology

In a recent review of 11 major ecological journals during 1996–2003, Ellison (2004) found 69 articles utilizing Bayesian statistical methods. Of these, 25 used informative priors (e.g., Anholt et al. 2000, Fleischman and Burwen 2003, Wintle et al. 2003), which tended to be based on information from previous studies rather than elicitation of expert opinion. When elicitation was used, the methodology for deriving expert knowledge was not necessarily reported (e.g., Taylor et al. 1996, Cheng and Chen 2005, Clark et al. 2005, Fuentes et al. 2006). More generally there are few examples of designed elicitation applied in ecology (e.g., Link and Sauer 1996, Garthwaite and O'Hagan 2000, McCarthy and Masters 2005, Hahn 2006). In this paper, we consider two examples which have been reported in some depth (Crome et al. 1996, Borsuk 2004), and describe their elicitation design using the framework described above. We then undertake more detailed assessment for three examples from our experience.

Crome et al. (1996) undertook both traditional frequentist and Bayesian analyses (E1) to assess the impact of logging on species counts based on data obtained using a Before-After-Control Impact-Pairs (BACIP) design. In the Bayesian analysis, posterior probabilities of scenarios were linked directly to management actions (E1). Stated advantages of the Bayesian approach included: scientific relevance, by avoiding questions of "Is the logging effect zero?" in favor of "Is the effect large enough to influence my scientific or personal beliefs or management aims?" (E1); a focus on reporting uncertainty in model parameters; and the ability to include subjective probabilities about individual events (E2).

The direct elicitation approach (E4a) used by Crome et al. (1996) assumed that experts familiar with interpreting BACIP results (E3) would also be able to express their ecological knowledge in terms of an effect size for logging (E2). The authors investigated the impact of various polarized beliefs embodied in four different prior distributions (E5d), with the aim of determining whether the posterior model led to consensus between prior knowledge and observed data (E1, E6b). This focus on whether data agree with experts differs subtly from the usual focus of many informative Bayesian studies, on whether the prior has added information to the data (E6b).

Fractiles of the expected impact of logging on species (E4b, E2) were elicited from 15 experts (E5a, E6a), via interview based on a standard questionnaire (E6c). Questions addressed variable intervals (E4a) for the three quartiles, e.g., the median was sought using (E4c) "Choose a level of impact (percentage increase or decrease) so that there is a 50% chance that the effect will be below this level." The possible range and outer quantiles (E5c) were also elicited (E4e). Elicitors felt that experts preferred quartiles to be elicited explicitly (e.g., asking for level exceeded by 25% of cases) rather than via bisection (i.e., conditioning on the median) (E4b).

Fractiles were analyzed using principal components (E4d) to identify three experts (E6a) considered representative of the most divergent views (E6b, E5a): pessimistic (conservationist), indifferent (lay person), and optimistic (logging industry). The fractiles for these three experts (E4b) were then used to encode the effect of logging on species counts in logged areas as a two component mixture of lognormal distributions (E3), using nonlinear least squares (E4d). Both priors and posteriors were compared arising from Bayesian analyses, using non-informative priors or informative priors from each class of opinion. Prior opinion was effectively weighted according to the uncertainty of each exemplar expert, as reflected by variance of encoded lognormal mixtures given elicited fractiles (E4e). Summary statistics based on posteriors of ranges of logging effects, based on each type of expert, defined decision rules for proscribing management actions (E1).

A second example is given by Borsuk (2004), who was motivated to use a Bayesian approach due to the difficulty in accurately characterizing, using process models (E1), "how improvements in oxygen conditions will improve the health of fish and reduce the frequency of fish kills" (E2). Modeling focused on management requirements rather than improvement of mechanistic understanding (E1). We examine the approach taken on the first of two different models reflecting theories and evidence regarding the underlying processes and causal factors for fish kills (E2). Modeling was decomposed into three main steps: modeling fish population health, incidence of fish kills depending on population health and other factors, and integrating these two models into an overall model (E3). Expert opinion was required to parameterize the first level of modeling for input into the overall model (E2).

For the first step, expert opinion was sought on (E4a) "the aggregate relationship between fish population health and the annual extent of low oxygen bottom water" in summertime for a varying number of days (E2). Elicitations followed model decomposition into a hierarchical model (E3). Categories of population response (fish population health) were carefully defined in consultation with experts and determined to depend on two defining attributes being extent of visible disease and growth rates (E3, E4b). The probability of these fish population categories (E4b), for different numbers of hypoxic days, were elicited from experts, providing an indirect approach of determining the differential effect of hypoxic days at different population levels (E4a). These were based on published thresholds of tolerance of fish to hypoxia (E6a). Questions asked were similar to: "Given a summer in which bottom water oxygen concentration (depth greater than 1.5 meters) in the mid-channel of the Neuse Estuary averages less than 2.0 mg/L for 10 out of 92 days in July, August, and September, what is the probability that fish population health at the end of the summer can be characterized as excellent? good? poor?" (E4c).

A consensus approach (E6c, E4e) to combining expert opinion (E5a) was undertaken, accounting for imprecision in both estimated value and range of expert opinions (E5). Prior estimates were obtained by fitting a cumulative logit regression model (E3) via maximum likelihood to the elicited responses, with known covariates, using the proportional odds assumption (E4d). Imprecision on estimates was encoded by using an aggregate estimate of the effective sample size relevant to the multinomial distribution of responses (E4e, E4d). Other modeling assumptions were tested, including a latent trait formulation and alternative link functions (E5d). Model predictions were then expressed in a form more closely reflecting original definitions (E6c, E5e), based on thresholds provided by experts (E2) (Borsuk 2004: Fig. 2).

Our experiences in expert elicitation cover a range of ecological topics, environmental management objectives, elicitation techniques, and statistical models. We undertake more detailed assessment (especially regarding E5 and E6) for three examples from our experience.

Case A.—Setting and evaluating vegetation condition benchmarks, by balancing misclassification rates and eliciting statistical distributions (Low Choy et al. 2005)

Case B.—Subregionalization of terrestrial bioregions via a finite mixture of multivariate Gaussian distributions, using elicitation of latent mixture allocation (Pullar et al. 2004, Accad et al. 2005).

Case C.—Modeling habitat suitability or predicting species distributions, via logistic regression, with informative priors imputed from elicitation of predicted presence/absence case-by-case (e.g., site-by-site) for specific habitats (Denham and Mengersen 2007, O'Leary et al. 2008*a*; A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*).

Vegetation benchmarking

Motivation (E1).—Vegetation management often requires identification of "benchmark" sites in "reference" or optimal natural condition (Eyre et al. 2006) (E1). Condition is usually described in terms of maturity and disturbance, with measurement of indicators such as density of large trees or fallen woody debris. Typically, however, data on these indicators are available for few of the vegetation types of interest. For these vegetation types, Bayesian regression can set appropriate benchmarks within a decision analysis framework (Low Choy et al. 2005). However, for remaining types with no available data, expert knowledge is the sole source of information (E1). Elicitation within a Bayesian framework ensures that expert-derived benchmarks can later be combined with data, when available, and also guide design of monitoring (E1, E6b).

Specification of expert knowledge (E2).—Given an indicator of vegetation condition, Y, a benchmark was desired, such that the "best" sites in reference condition would score above this benchmark, degraded sites would score below this benchmark down to an effective score of zero, and intermediate sites would score somewhere between these two levels (E1).

A naive one-step approach would involve setting the benchmark at a level considered by experts (E2) to represent the median condition $\tilde{Y}_i$ evaluated across a set of sites i in reference condition $i \in R$ (E4a, E3). There are two difficulties with this approach. First, it confounds the elicitation of ecological knowledge with the decision-making process (E2), so is open to motivational biases (E6a). Some scientists can be more conservative than others, specifying a higher benchmark to avoid mistakenly capturing sites in degraded condition. Other experts could desire a more inclusive benchmark to maximize the chance of capturing all sites in reference condition (E5c). Second, specifying a benchmark, based only on sites in reference condition, leads to implicit and potentially inappropriate conditioning biases by individual experts (E5b).

Since the benchmark is to be used to discriminate between sites in reference and degraded condition, it should explicitly account for some baseline level of degraded condition (E2). Such a baseline would enable assessment of misclassification rates, being the potential to misclassify degraded sites as being in reference condition, and vice versa (E1). The problem was therefore decomposed into two stages (E3) to allow indirect assessment of the benchmark (E4a): first describe the range of indicator values Y for different levels of condition Z (both reference and degraded) (E4b); and then assess misclassification rates (E4b).

Statistical formulation (E3).—A categorical variable $Z_i = k$ indicates known condition at site i is at level k, including the two extremes of degraded $k = 1$ and reference $k = K$ condition (E3). For a hypothetical set of sites considered to be in degraded condition $Z_i = 1$, the expert can be asked to describe the expected range of values $p(Y_i \mid Z_i = 1)$ for an indicator Y_i (E3, E2). This can be repeated for sites in reference condition $Z_i = K$, and at intermediate levels of condition $1 < Z_i < K$ (E4b). This model provides a basis for estimating the probability that sites exceed a potential benchmark T within each category k of condition $\alpha_k = p(Y_i > T \mid Z_i = k)$ (E4d). Then the true positive rate TPR $= \alpha_K$ is the correct classification rate for reference sites, and the false positive rate FPR $= \alpha_1$ is the misclassification rate of degraded sites. Of interest are the FPR and FNR, which is the false negative rate and simply calculated as FNR $= 1 -$ TPR. Depending on the decision-making context, a threshold T can be selected to minimize one or more misclassification rates (E3, E5c). For example environmental managers may desire a neutral decision (E2), which requires simultaneous minimization of both FPR and FNR, if the cost of each error is comparable (Low Choy et al. 2005) (E4d).

Encoding (E4).—Specifically, experts may be asked for a series of fractiles from which to impute the distribution of indicator values at each level of condition $p(Y_i \mid Z_i = k)$ (e.g., O'Hagan et al. 2006) (E4b). The most extreme fractiles should be elicited first to ensure the expert does not inadvertently constrain themselves too narrowly to their median or modal estimate of the indicator. This approach has been shown to minimize anchoring as a source of bias (e.g., Spetzler and Staël von Holstein 1975) (E5c). First key concepts were defined, such as "site," "study area," "reference condition," and each indicator (E6d). Questions targeting quantiles were asked in terms of frequencies rather than probabilities to improve accuracy (Kynn 2008) (E4c) "Imagine 100 sites within the study area that are in reference condition. What is the lowest possible value of the attribute you would expect at any of these sites? [You would be surprised if more than one in a hundred sites scored lower.] What attribute value would be exceeded by only two or three sites in 100? . . ."

For a normal or lognormal distribution, two quantiles can be related algebraically either on the unlogged or logged scale (E4d) to the mean and variance, via straightforward arithmetic (e.g., Garthwaite and O'Hagan 2000, Low Choy et al. 2008). Numerical approaches may be required to encode parameters for other distributional choices (e.g., Léon et al. 2003) (E4d).

Protocol (E6).—The design required a simple low-technology implementation, to enable elicitation by elicitors with limited statistical expertise, in remote areas with little computing facilities, consistently across various experts and elicitors (E6e). Thus a simple questionnaire was designed (E6c), with results that could be encoded in a spreadsheet to enable some graphical feedback (E4c). A boxplot was used for graphical representation, and some basic training in interpreting the boxplot was devised (E4c, E6d). Joint presentation of introductory and training materials could be made at annual Technical Advisory Panel meetings (E6c). Training included examination of a well-understood indicator in a wider group, with thorough discussion revising definitions of concepts, both ecological and statistical. Pilot studies validated the questions in the questionnaire (E5c). Verbal, tabular and graphical representations were provided to support elicitation as well as feedback (E5c, E4c). Consistency within and between individual experts (E5a) was improved via comparison of boxplots describing the range for a vegetation condition indicator, across different regional ecosystems (E5c).

Bioregionalization

Motivation and specification of expert knowledge (E1, E2).—Environmental management has historically been managed through identification of ecoregions or bioregions, being areas that share common broad scale geographical attributes such as geology and climate. Bioregionalization has traditionally relied on experts to integrate a large source of material to delineate acceptable boundaries around bioregions (E2). Experts refer to many sources of information, including their own experiences from fieldwork or conceptual knowledge (E6a). In synthesizing this information experts must refer to multiple indicators of landscape processes, then prioritize those driving delineation of each boundary (E2). Expert panels have become a common mechanism to achieve this (Neldner 2006) (E6c).

Two sources of available data were considered for constructing a prior in Rochester et al. (2004): subregional and bioregional boundaries agreed via consensus of an expert panel; or an allocation suggested by another regionalization method, such as hierarchical clustering, applied to a different dataset (E2, E6a). For the case study considered by Accad et al. (2005), two sources of indirect prior information were expert-defined subregional boundaries and presence/absence of two indicator vegetation species.

Statistical model formulation (E3).—In contrast to the usual data-mining approaches used for data-driven bioregionalization (Mather and Doornkamp 1970, Hall et al. 2002), a model is required to enable Bayesian analysis combining data with expert knowledge (E1). The Gaussian mixture model has been applied for data-driven bioregionalization (e.g., Stepinski and Vilalta 2005) using non-informative priors (E2) and empirical Bayes for computation (Cheeseman and Stutz 1996). Here we consider the case where expert knowledge has been incorporated as informative priors (E1, E2), with full Bayesian inference via Markov Chain Monte Carlo estimation (Rochester et al. 2004, Accad et al. 2005) (E4d). Informative priors not only provide a mechanism for incorporating expert knowledge (E2), but for mixture models they ensure proper priors to avoid numerical difficulties (Hobert and Casella 1996) (E4d).

Several covariates $y_{i1}, \ldots, y_{iJ}$ representing landscape characteristics $j = 1, \ldots, J$ can be used to classify sites $i = 1, \ldots, n$ into environmental envelopes $k = 1, \ldots, K$. Envelopes can be mapped to geographic regions where sites are allocated according to $z_i = k$ (Rochester et al. 2004). The multivariate Gaussian finite mixture model can be expressed as a two-level hierarchical model (e.g., Fraley and Raftery 1999). At the first level, each site i is allocated to region $z_i = k$, with probability $w_k = p(z_i = k)$. At the second level, a separate Gaussian distribution $\phi(\theta_k)$ describes the environmental envelope in each of the $k = 1, \ldots, K$ regions, with parameters μ_k, $\mathbf{\Sigma}_k$:

$$p(y|w, \mu, \Sigma) \sim \sum_{k=1}^{K} w_k \phi\left(y|\mu_{k,} \Sigma_k\right). \tag{2}$$

To satisfy the independence assumption, sites were selected via random sampling, with minimum distance accounting for spatial autocorrelation, within each region. For a Bayesian implementation (Lavine and West 1992, Diebolt and Robert 1994) conjugate prior distributions may be used, being a multivariate normal for μ, inverse Wishart for Σ, and Dirichlet for w.

Encoding (E4).—Prior information on regionalizations would be difficult to elicit structurally, due to the dimension and complexity of parameters μ, w, and especially Σ (E4a). However predictive elicitation is easily applied here since, by defining boundaries, experts have essentially designated a "hard" allocation of each site i to a single region k, via $\left\{z_i^{(0)} = k\right\}$. This in turn defines prior estimates:

$$w_k^{(0)} = \frac{1}{n} \sum_i I\left[z_i^{(0)} = k\right]$$

(E4a, E2). A discriminant analysis (Fraley and Raftery 1999) can be used to fit a multivariate normal distribution to each region to obtain estimates of regional means, estimated via

$$\mu_{jk}^{(0)} = w_k^{(0)} \sum_{i:z_i^{(0)}=k} y_{i,j}$$

and similarly prior variances $\Sigma_k^{(0)}$ (E4d). The degrees of freedom for priors for μ_k and Σ_k reflected the effective sample sizes or expert confidence in their opinions (E4e). The impact of changing the effective sample sizes for expert knowledge could be visualized by experts by mapping (E4c) posterior estimates of bioregions $\{z_i\}$, with a separate map showing uncertainty (Accad et al. 2005) (E5d, E6b).

Protocol (E6).—Expert opinion on bioregional boundaries is generally decided via consensus (E4c, E6b) at meetings (E6c) held a few times a year (e.g., Neldner et al. 2004), with information at hand on vegetation survey results, reports, and hardcopy maps on various themes ranging from topography to geology and climate (E6a). Experts are selected (E6a, E5a), and expert-defined boundaries verified, via a rigorous and well-accepted process (e.g., Neldner et al. 2004) (E5). In addition, experts involved in the informative Bayesian modeling exercise were asked to verify encoding of expert knowledge through visualisation of the envelopes representing the range of environmental attributes, defined by the mixture model parameters μ_k and $\mathbf{\Sigma}_k$, within each expert-defined subregion (E4c, E5e). An integrated geographic information system (GIS) and other modeling tools were developed to support modeling and provide feedback in the form of maps, in both cases (E5c, E6e).

Habitat models: interactive predictive elicitation

Motivation (E1).—Experts may be unable to undertake the difficult elicitation task of estimating regression coefficients (Martin et al. 2005). Yet knowing the habitat at a hypothetical site, experts are often able to assess ecological response such as probability of presence or relative abundance (E2). This indirect elicitation approach is well suited to complex regressions, and to cases where experts have field-based practical knowledge rather than theoretical abstract knowledge (Denham and Mengersen 2007). Here we discuss indirect elicitation for logistic regression for habitat modeling (E1). We discuss a prototype elicitation tool and its successor, which were both applied to elicitation of habitat requirements of the Australian endangered brush-tailed wallaby (*Petrogale pennicillata*; see Plate 1). The prototype elicited opinions of just two experts (Denham and Mengersen 2007, O'Leary et al. 2008*b*), while its successor undertook more thorough elicitation from nine different experts (J. V. Murray, R. W. Goldizen, R. A. O'Leary, C. A. McAlpine, H. P. Possingham, and S. J. Low Choy, *unpublished manuscript*).

Specification of expert knowledge (E2).—Denham and Mengersen (2007) elicit predicted probability at sites, using a map-based interface to show habitat covariates and other contextual information (E4a). They extend the predictive elicitation approach for normal regression (Kadane et al. 1980) and logistic regression (Chen et al. 1999), taking advantage of spatial contextual information (Craig et al. 1998). It differs from other elicitation techniques for habitat modeling via regression outlined in the description of E2. This approach has been modified (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*) to follow the conditional mean prior approach of Bedrick et al. (1996).

Statistical formulation (E3).—Let $Y_i = 1$ denote an observed presence at site i and $Y_i = 0$ denote an observed absence. Denote by x_{ij} the jth covariate (e.g., habitat factor, environmental gradient) measured at site i. The logistic regression of response y on covariates $\mathbf{x}$ involves unknown coefficients β_j corresponding to the jth covariate:

$$y_i \overset{\text{iid}}{\sim} \text{Bern}(\mu_i)$$

where

$$\text{logit}(\mu_i) = \beta_0 + \sum_j x_{ij}\beta_j. \quad (3)$$

Normal priors $\beta \sim \mathcal{N}(b, \mathbf{\Sigma})$ are typically used (Gelman et al. 2004) with full covariance structure $\mathbf{\Sigma}$ (Denham and Mengersen 2007) or independent variances $\mathbf{\Sigma}$ = diag(ε_1^2,ε_J^2) in the absence of collinearity (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*).

Similar to Bedrick et al. (1996), experts are asked to provide assessments Z_m on conditional means μ_m, here the probability of presence conditional on known habitat. Quantities comprising Z_m are elicited from experts on the probability of presence at sites $m = 1, \ldots, M$, with corresponding habitat covariates X_m. Hence,

$$Z_m \overset{\text{iid}}{\sim} \text{Beta}(\mu_m, \gamma_m),$$

where

$$\text{logit}(\mu_m) = \beta_0 + \sum_j x_{mj}\beta_j. \quad (4)$$

Here the mean μ and variance γ are related to the usual Beta shape and scale parameters a and b via $\mu = a/(a + b)$ and $\gamma = \mu(1 - \mu)\rho$ with $\rho = 1/(1 + a + b)$. We may assume the same site-specific probabilities of presence, and therefore the same coefficients β, apply in both the data model (Eq. 3) and the prior model (Eq. 4) (Chen et al. 1999) (E3). Prior estimates of coefficients β can then be obtained simply by applying the Beta regression defined in Eq. 4 (E4d). Alternatively this may be transformed to a Binomial regression (Denham and Mengersen 2007) with the same μ, but effective sample size $n_m = \mu(1 - \mu)/\gamma_m$ (E4d). The design matrix $\mathbf{X} = \{x_{mj}\}$ for the M elicitation sites can be preselected and therefore designed by the elicitor (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*) or nominated by the expert (Denham and Mengersen 2007) (E2). Elicitation at each site m implicitly conditions on knowledge of J covariates that are measured at the site and available for perusal (numerically or mapped) by the expert (E2, E5b).

Encoding (E4).—The tool links (Denham and Mengersen 2007) or interfaces (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*) with a geographic information system (GIS), that can be used to map (habitat) covariates across the study area or to query covariate values (habitat profile) for particular sites (E4c, E6e). Knowing the covariate profile at a site, the expert provides the range of possible values for predicted probability of presence at the site (E2), which following Eq. 4 can be parameterized by a beta distribution $p(Z_m \mid X_m, a_m, b_m) \sim \text{Beta}(a_m, b_m)$ (E3). To achieve this, the expert may edit numeric estimates of the Beta's shape and scale parameters (a_m, b_m) (E4d) (Denham and Mengersen 2007) or quantiles (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*) (E4d), or graphically manipulate a probability density curve via its quartiles (Denham and Mengersen 2007) or a boxplot (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*) (E4c). A moment-matching approach can be used to calculate the Beta parameters corresponding to the median and upper quartile (Denham and Mengersen 2007) (E4d), or a numerical approach used to estimate the closest fitting Beta parameters given the mode and two or four quantiles (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*) (E4b).

Feedback can be instantly provided on the expert-based prior model (E5e), showing the usual regression diagnostics (E5c), providing an opportunity for reflection to ensure coherent and accurate estimates (E6e). Elicitations or model diagnostics can be saved as "dynamic" graphical documents where interaction may continue at a later time, or as purely static graphical documents for reporting (E4c, E6e). This underlying data management also supports: sensitivity analysis to priors by supporting elicitation by the same expert in different projects containing different sites (E5d), as well as elicitation from multiple experts for the same or different sites (E5a). Provision is made for recording varying levels of confidence that an expert may have in their assessments at each site, which may be used to weight each site's contribution to the prior distribution (E4d). This hybrid approach oscillates between a variable interval approach, eliciting quantiles of predicted probability, and a fixed-interval approach, checking the mode, tails and other quantiles of the encoded beta distribution (E4b, E4d).

Protocol (E6).—The elicitation tools are designed to accompany personal interview of a single expert by an elicitor (E6c) and can be run from desktop (Denham and Mengersen 2007) or a laptop computer (A. James, S. Low Choy, and K. Mengersen, *unpublished manuscript*) (E6e). Interviews are more streamlined since the tool manages many of the tedious aspects of interaction and documentation (E5c, E6e).

Discussion

The case studies presented above demonstrate that the steps in the elicitation process (E1–6) provide a useful overall description of elicitation design. They also show that these steps cannot simply be followed in a linear fashion, since many are interrelated. We have distilled eight key issues of relevance to designing elicitation in ecology contexts:

1) A major motivation for using expert opinion is that it is an important source of information for ecological models.
2) Relevant and accurate prior models target available expert knowledge.
3) Problem decomposition, leading to hierarchical model representations, in the statistical formulation phase facilitates incorporation of expert knowledge as informative priors.
4) Indirect rather than direct elicitation can be more effective and control biases in many situations.
5) Communication styles that target various styles of thinking used by ecologists lead to more effective and accurate elicitation.
6) Technology can be harnessed to provide an interactive environment and feedback to better facilitate and streamline elicitation.
7) Assessing informativeness of priors and their impact on model performance is important for understanding and evaluating model outputs.
8) Expert panels can provide a useful mechanism for facilitating elicitation in ecology.

These points are related to the five documented by Kadane and Wolfson (1998 [their points a–e]) and overlap with four recommendations of Kynn (2008). We address each point here.

Motivation (E1) for an informative Bayesian analysis varied among the cases studied. Immediate interim results were required where expert knowledge was the only source of information (Borsuk 2004; case A) or the best source since spatial data sets were not of consistent resolution and accuracy across the study area (cases B and C). Hence in these cases expert knowledge was the "most worthwhile to elicit" (Kadane and Wolfson 1998 [their point a]). This illustrates the importance of expert knowledge for similar ecological contexts (point 1).

The process of identifying relevant and available expert knowledge (E2) generally revealed useful options that depended on model formulation (point 2). For assessing vegetation condition (case A), it was crucial to recognize that it was problematic to ask experts to directly set a benchmark, yet easier to ask them to describe the distribution of attributes under varying levels of condition. Targeting the most influential factors for modeling simplified elicitation and also allowed stochastic variation to summarize the effects of finer scale processes (Borsuk 2004). Utilizing a mixture model, together with indirect elicitation that predicted site allocation by specifying boundaries, provided a natural way to target and incorporate expert knowledge compared to the more popular data mining approaches (case B). For habitat modeling (case C), elicitation was simplified by acknowledging that ecologists, particularly those with substantial field experience, find it easy to predict probabilities of species occurrence at sites with mappable habitat characteristics. In these cases, decomposition of the model (point 3, E3) provided a more targeted, less complex and more accessible basis for elicitation (Spetzler and Staël von Holstein 1975), so that elicitation tasks would be "as 'small' and distinct as possible" (Kynn 2008; recommendation 5). See the Appendix for details on some elicitation tools.

Model decomposition (point 3) requires some thought, but naturally results in specification of hierarchical models, as demonstrated by all case studies. To simplify elicitation and encoding, however, generally required a little extra complexity in statistical formulation and therefore estimation. This complexity took different forms: a hierarchical "prior" standalone model comprising three levels of dependencies (Borsuk 2004); an additional step in the Bayesian learning cycle where priors were the result of analysis of preliminary data elicited from experts (cases C and B); a decision analysis applied to the prior model to balance misclassification error rates (case A) or applied to the posterior to guide management actions (Crome et al. 1996).

When experts were familiar with the modeling framework, expert knowledge was sought directly via a structural approach for eliciting parameters in the model (Crome et al. 1996) (E4a). However in the four other case studies, an indirect approach to elicitation was found necessary or more appropriate (point 4), supporting the fourth conclusion presented by Kadane and Wolfson (1998 [their Discussion section]). Indirect encoding enabled transformation of existing expert knowledge on boundaries between regions into suitable priors on properties of regions (case B). Expert knowledge was targeted more accurately by asking experts to assess ecological response for given covariate values (Borsuk 2004; case studies A and C). In these cases, elicitation was designed to target expert knowledge and be more easily understood, thus lowering cognitive biases and reducing the need for "mental gymnastics" (Spetzler and Staël von Holstein 1975, Kynn 2008: recommendation 5). This was achieved by asking the expert only about observable quantities (Kadane and Wolfson 1998 [their point b]), keeping questions more concrete than abstract.

Feedback helps experts maintain self-consistency, explore their own knowledge (not often expressed so precisely) and therefore greatly reduce cognitive biases (point 6). Although agreeing with Kadane and Wolfson (1998 [their point d]) that frequent feedback be provided, we extend this to recommend that a variety of communication styles be used to suit the varying learning and thinking styles of experts (point 5). Experts with aural and oral thinking styles are helped by discussion with the elicitor, and those with abstract thinking by providing concrete numeric information (all cases). Visual thinking is supported via feedback through graphs (cases A–C) and maps (B and C) (Kynn 2008: recommendation 9), particularly for validation of elicited prior models. Kinetic thinking is supported by interactive feedback, particularly with technology assistance (C).

Predictive elicitation tools for regression (case C) exemplify interactive graphical and map-based systems of elicitation, permitting hybrid approaches to encoding. These tools have many benefits for the elicitation process (point 6). They can educate the user about statistical concepts and models. For example, the tool in case C educates by dynamically showing the link between a probability density (Beta distribution for elicited response), its quantiles, and parameters (numbers); by revealing the link between data elicited site-by-site (elicited information) and the model (species response curves); and by demonstrating the use of regression diagnostics in a way that relates directly to the expert's knowledge. With modern technology, the frequent feedback found desirable by Kadane and Wolfson (1998 [their point d]) can be semi-automated and provided interactively. Moreover an interactive environment greatly assisted experts in these case studies to reflect, refine (Kynn 2008: recommendation 9) and provide coherent and accurate estimates. The ability to revisit and adjust estimates substantially reduced pressure on experts. This interaction can be supported by linking to geographic information systems (GIS) (B and C), using a software tool (C), or by spreadsheets (A). Dynamic linking of graphs representing elicited information, feedback and prior models also assist experts (C). Technology provides a useful means of undertaking tedious or repetitive tasks, therefore minimizing error (by elicitors or experts), also noted by (Kynn 2008: recommendation 5).

Little research has compared the impact or informativeness of priors (point 7) obtained using different expert elicitation techniques (e.g., Gavasakar 1988, O'Leary et al. 2008*a*). In contrast to many non- or weakly informative Bayesian analyses, in all of these case studies the observed data do not provide the baseline for comparing sensitivity of model outputs to prior assumptions. Rather, in many cases (Borsuk 2004; A–C) prior information establishes a baseline for assessing the impact of additional observed data, when they become available (e.g., Bernardo 2006, Bousquet 2008). Midway between these two extremes, one case (Crome et al. 1996) focuses on agreement between prior information and data. In case B, stakeholders found it useful to visualize how changing the weight (effective sample size) of prior information impacted on predictions (mapped boundaries).

Expert elicitation can take advantage of well-developed and existing mechanisms in ecology for more general expert consultation (point 8), an advantage since a familiar context greatly facilitates elicitation (Kynn 2008: recommendation 4). In our experience (A–C), expert panels were used either as the mechanism for implementing the elicitation (A and B) or as the mechanism for identifying and enlisting experts (C). The logistical and scientific benefits of these mechanisms should not be underestimated; in many cases ensuring the correct balance of expert opinion (E6b) is as important as any other factor for the perceived and actual validity of resulting estimates. Ecological experts are familiar with designing processes to minimize motivational biases, such as large group consultation (A–C). In many cases, expert panels, or subsets of expert panels, provide a convenient means of delivering training (E6d), providing an ideal setting for conditioning experts to potential biases (Kynn 2008) (E5b), refining definitions (B), developing shared understanding of ecological and statistical concepts, and demonstrating the elicitation process (A), including a calibration exercise (E5e) (A and C) (Hamilton et al. 2005).

This paper illustrates the value of informative Bayesian statistical analyses for ecological applications, through detailed discussion and comparison of five case studies. We have proposed and validated a six-step process for designing expert elicitation that may be applied in various ecological contexts. Case studies revealed several issues of particular importance when designing elicitation in ecology, often resonating with previous research. This paper contributes to ongoing research into the relative merits of various elicitation approaches, and revisits the principles of successful elicitation in a modern context.

Acknowledgments

We are grateful for the courage shown by the ecologists involved in case studies, in first considering and then embracing a new statistical paradigm, and the ensuing rigorous scientific debate in presenting the new approach. Ecologists for case studies included: Justine Murray, Wayne Rochester, Annie Kelly, Geoff Smith, Teresa Eyre, Arnon Accad, and John Neldner. We thank anonymous reviewers for helpful comments that have measurably improved the manuscript. The first author's work was supported by travel grants from the International Biometrics Society and the School of Mathematical Sciences at Queensland University of Technology (QUT), as well as research funding provided by a QUT Postdoctoral Fellowship. The first and third authors were also partially supported by Australian Research Council Discovery Grant DP0667168. The second author was supported by a QUT Postgraduate Student Award.

Literature Cited

Accad, A., S. Low Choy, D. Pullar, and W. Rochester. 2005. Bioregional classification via model-based clustering. Pages 1326–1332 *in* A. Zerger and R. M. Argent, editors. MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Canberra, Australia.

Al-Awadhi, S. A., and P. H. Garthwaite. 2006. Quantifying expert opinion for modelling fauna habitat distributions. Computational Statistics 21:121–140.

Alho, J. M., J. Kangas, and O. Kolehmainen. 1996. Uncertainty in expert predictions of the ecological consequences of forest plans. Applied Statistics 45:1–14.

Anholt, B. R., E. Werner, and D. K. Skelly. 2000. Effect of food and predators on the activity of four larval ranid frogs. Ecology 81:3509–3521.

Bernardo, J. 2006. Objective Bayesian statistics: an introduction to Bayesian reference analysis. Postgraduate Tutorial Seminar. Valencia 8/ISBA 2006 World Meeting on Bayesian Statistics, Valencia, Spain.

Borsuk, M. E. 2004. Predictive assessment of fish health and fish kills in the Neuse river estuary using elicited expert judgment. Human and Ecological Risk Assessment 10(2): 415–434.

Bousquet, N. 2008. Diagnostics of prior-data agreement in applied Bayesian analysis. Journal of Applied Statistics 35(9): 1011–1029.

Box, G., and G. Tiao. 1973. Bayesian inference in statistical analysis. Wiley Publishing, New York, New York, USA.

Boyce, M. S., L. L. Irwin, and R. Barker. 2005. Demographic meta-analysis: synthesizing vital rates for Spotted Owls. Journal of Applied Ecology 42:38–49.

Burgman, M. 2004. Expert frailties in conservation risk assessment and listing decisions. Pages 20–29 *in* D. L. P. Hutchings and C. Dickman, editors. Threatened species legislation: is it just an Act? Royal Zoological Society of New South Wales, Mosman, NSW, Australia.

Chao, C. T. 2003. Markov chain Monte Carlo on optimal adaptive sampling selections. Environmental and Ecological Statistics 10(1):129–151.

Cheeseman, P., and J. Stutz. 1996. Bayesian classification (autoclass): theory and results. Pages 153–180 *in* U. M. Fayyad, G. Piatetsky-Shapiro, P. Smith, and R. Uthurusamy, editors. Advances in knowledge discovery and data mining. MIT Press, Cambridge, Massachusetts, USA.

Chen, M. H., J. G. Ibrahim, and C. Yiannoutsos. 1999. Prior elicitation, variable selection and Bayesian computation for logistic regression models. Journal of the Royal Statistical Society, Series B (Methodological) 61:223–242.

Cheng, K. F., and J. H. Chen. 2005. Bayesian models for population-based case-control studies when the population is in Hardy Weinberg equilibrium. Genetic Epidemiology 28: 183–192.

Clark, J. S., G. Ferraz, N. Oguge, H. Hays, and J. DiCostanzo. 2005. Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. Ecology 86:2232–2244.

Clark, K. E., J. E. Applegate, L. J. Niles, and D. S. Dobkin. 2006. An objective means of species status assessment: adapting the Delphi technique. Wildlife Society Bulletin 34: 419–425.

Clemen, R. T., and T. Reilly. 2001. Making hard decisions with decision tools. Duxbury; Thompson Learning, Pacific Grove, California, USA.

Clemen, R. T., and R. L. Winkler. 1999. Combining probability distributions from experts in risk analysis. Risk Analysis 19(2):187–203.

Clyde, M. A. 1999. Comment on Bayesian model averaging: a tutorial by Höeting et al. Statistical Science 14:382–417.

Cox, D. R. 2000. Comments on paper by Lindley. Statistician 49(3):321–324.

Craig, P. S., M. Goldstein, A. H. Seheult, and J. A. Smith. 1998. Constructing partial prior specifications for models of complex physical systems. Statistician 47:37–53.

Crome, F. H. J., M. R. Thomas, and L. A. Moore. 1996. A novel Bayesian approach to assessing impacts of rain forest logging. Ecological Applications 6:1104–1123.

Cummings, C. L., H. M. Alexander, A. A. Snow, L. H. Rieseberg, M. J. Kim, and T. M. Culley. 2002. Fecundity selection in a sunflower crop-wild study: can ecological data predict crop allele changes? Ecological Applications 12:1661–1671.

Dawid, A. P. 1982. The well-calibrated Bayesian (with discussion). Journal of the American Statistical Association 77(379):605–610.

Dawid, A. P. 2004. Probability, causality and the empirical world: a Bayes-de Finetti-Popper-Borel synthesis. Statistical Science 19(1):44–57.

Denham, R., and K. Mengersen. 2007. Geographically assisted elicitation of expert opinion for regression models. Bayesian Analysis 2(1):99–136.

Dickey, J. M., and T. J. Jiang. 1998. Filtered-variate prior distributions for histogram smoothing. Journal of the American Statistical Association 93:651–662.

Diebolt, J., and C. P. Robert. 1994. Estimation of finite mixture distributions through Bayesian sampling. Journal of the Royal Statistical Society, Series B. Methodological 56(2):363–375.

Ellison, A. 1996. An introduction to Bayesian inference for ecological research and environmental decision making. Ecological Applications 6:1036–1046.

Ellison, A. M. 2004. Bayesian inference in ecology. Ecology Letters 7:509–520.

Eyre, T., A. Kelly, and V. Neldner. 2006. Biocondition: a terrestrial vegetation condition assessment tool for biodiversity in Queensland, field assessment manual. Version 1.5. Technical report. Environmental Protection Agency, Brisbane, Australia.

Ferrier, S., G. Watson, J. Pearce, and M. Drielsma. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodiversity and Conservation 11(12):2275–2307.

Fleischman, S. J., and D. L. Burwen. 2003. Mixture models for the species apportionment of hydroacoustic data, with echo-envelope length as the discriminatory variable. ICES Journal of Marine Science 60:592–598.

Fleishman, E., R. M. Nally, J. P. Fay, and D. D. Murphy. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the great basin. Conservation Biology 15(6):1674–1685.

Fraley, C., and A. Raftery. 1999. Mclust: software for model-based cluster analysis. Journal of Classification 16:297–306.

Fuentes, M., T. G. F. Kittel, and D. Nychka. 2006. Sensitivity of ecological models to their climate drivers: statistical ensembles for forcing. Ecological Applications 16:99–116.

Garthwaite, P. H., and J. M. Dickey. 1985. Double and single-bisection methods for subjective probability assessments in a location-scale family. Journal of Econometrics 29:149–163.

Garthwaite, P. H., and A. O'Hagan. 2000. Quantifying expert opinion in the UK water industry: an experimental study. Statistician 49:455–477.

Gavasakar, U. 1988. A comparison of two elicitation methods for a prior distribution for a binomial parameter. Management Science 34:784–790.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1(3):515–533.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis. Second edition. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Gneiting, T., and A. E. Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102(477):359–378.

Hahn, E. D. 2006. Re-examining informative prior elicitation through the lens of Markov chain Monte Carlo methods. Journal of the Royal Statistical Society, Series A 169:37–48.

Hall, M. J., A. W. Minns, and A. K. M. Ashrafuzzaman. 2002. The application of data mining techniques for the regionalisation of hydrological variables. Hydrology and Earth System Sciences 6(4):685–694.

Hamilton, G., C. Alston, T. Chiffings, E. Abal, B. Hart, and K. Mengersen. 2005. Integrating science through Bayesian belief networks: case study of *Lyngbya* in Moreton Bay. Pages 392–399 *in* A. Zerger and R. M. Argent, editors. MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Canberra, Australia.

Hobbs, N. T., and R. Hilborn. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecological Applications 16:519.

Hobert, J. P., and G. Casella. 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the American Statistical Association 91(436): 1461–1473.

Hogarth, R. M. 1975. Cognitive processes and the assessment of subjective probability distributions. Journal of the American Statistical Association 70(350):271–289.

Jeffreys, H. 1961. Theory of probability. Third edition. Oxford, Clarendon Press, UK.

Jenkinson, D. 2005. The elicitation of probabilities: a review of the statistical literature. Technical report, University of Sheffield, Sheffield, UK.

Kadane, J. B. 1980. Predictive and structural methods for eliciting prior distributions. North-Holland, Amsterdam, The Netherlands.

Kadane, J. B. 1990. The 1988 Wald Memorial Lectures: the present position in Bayesian statistics: comment. Statistical Science 5(1):80–82.

Kadane, J. B., J. M. Dickey, R. L. Winkler, W. S. Smith, and S. C. Peters. 1980. Interactive elicitation of opinion for a normal linear model. Journal of the American Statistical Association 75:845–854.

Kadane, J. B., and L. J. Wolfson. 1998. Experiences in elicitation. Statistician 47:3–19.

Kahneman, D., and A. Tversky. 1972. Subjective probability: a judgment of representativeness. Cognitive Psychology 3: 430–454.

Kuhnert, P. M., T. G. Martin, K. Mengersen, and H. P. Possingham. 2005. Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. Environmetrics 16:717–747.

Kynn, M. 2006. Designing elicitor: software to graphically elicit expert priors for logistic regression models in ecology. ⟨http://www.winbugs-development.org.uk/elicitor/files/designing.elicitor.pdf⟩

Kynn, M. 2008. The "heuristics and biases" bias in expert elicitation. Journal of the Royal Statistical Society, Series A 171(1):239–264.

Lavine, M., and M. West. 1992. A Bayesian method for classification and discrimination. Canadian Journal of Statistics 20:451–461.

Léon, C. L., F. J. Vázquez-Polo, and R. L. González. 2003. Elicitation of expert opinion in benefit transfer of environmental goods. Environmental and Resource Economics 26: 199–210.

Lindley, D. V. 2000. Philosophy of statistics. Statistician 49(3): 293–337.

Link, W. A., and J. R. Sauer. 1996. Extremes in ecology: avoiding the misleading effects of sampling variation in summary analyses. Ecology 77:1633–1640.

Low Choy, S., K. Mengersen, and J. Rousseau. 2008. Encoding expert opinion on skewed non-negative distributions. Journal of Probability and Statistics 3(1):1–21.

Low Choy, S., B. Stewart-Koster, A. Kelly, T. Eyre, and K. M. Mengersen. 2005. Identifying good condition in native vegetation: a Bayesian regression and decision theoretic approach. Pages 435–441 *in* A. Zerger and R. M. Argent, editors. MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Canberra, Australia.

Martin, T. G., P. M. Kuhnert, K. Mengersen, and H. P. Possingham. 2005. The power of expert opinion in ecological models: a Bayesian approach examining the impact of livestock grazing on birds. Ecological Applications 15: 266–280.

Mather, P. M., and J. C. Doornkamp. 1970. Clustering multivariate analysis in geography with particular reference to drainage-basin morphometry. Transactions of the Institute of British Geographers 51:163–187.

McCarthy, M. A., and P. Masters. 2005. Profiting from prior information in Bayesian analyses of ecological data. Journal of Applied Ecology 42:1012–1019.

Merritt, W. S., R. A. Letcher, and A. J. Jakeman. 2003. A review of erosion and sediment transport models. Environmental Modelling and Software 18:761–799.

Neldner, J. 2006. Why is vegetation condition important to government? A case study from Queensland. Ecological Management and Restoration 7:S5–S7.

Neldner, V. J., B. A. Wilson, E. J. Thompson, and H. A. Dillewaard. 2004. Methodology for survey and mapping of Regional Ecosystems and vegetation communities in Queensland. Version 3.0. Technical report. Queensland Herbarium, Environmental Protection Agency, Brisbane, Australia.

O'Hagan, A. 1998. Eliciting expert beliefs in substantial practical applications. Statistician 47:21–35.

O'Hagan, A., C. E. Buck, A. Daneshkhah, R. Eiser, P. Garthwaite, D. Jenkinson, J. Oakley, and T. Rakow. 2006. Uncertain judgements: eliciting experts' probabilities. John Wiley and Sons, New York, New York, USA.

O'Leary, R. A., S. J. Low Choy, K. Mengersen, M. Kynn, P. M. Kuhnert, R. Denham, T. J. Martin, J. Murray, and P. J. Jarman. 2008*a*. Comparison of expert elicitation methods for logistic regression for presence of endangered brush-tailed rock-wallaby *Petrogale penicillata*. Environmetrics. [doi: 10.1002/env.935]

O'Leary, R., J. Murray, S. Low Choy, and K. Mengersen. 2008*b*. Expert elicitation for bayesian classification trees. Journal of Applied Probability and Statistics 3(1):95–106.

Pascual, M. A., and P. Kareiva. 1996. Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. Ecology 77:337–349.

Pearce, J. L., K. Cherry, M. Drielsma, S. Ferrier, and G. Whish. 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. Journal of Applied Ecology 38:412–424.

Phillips, L. D., and S. J. Wisbey. 1993. The elicitation of judgmental probability distributions from groups of experts: a description of the methodology and records of seven formal elicitation sessions held in 1991 and 1992. Report nss/r282. Nirex UK, Didcot, UK.

Prato, T. 2005. Bayesian adaptive management of ecosystems. Ecological Modelling 183:147–156.

Pullar, D., S. Low Choy, and W. Rochester. 2004. Ecoregion classification using a Bayesian approach and model-based clustering: expert elicitation using a spatial and graphical tool. Pages 927–932 *in* C. Pahl-Wostl, S. Schmidt, A. E. Rizzoli, and A. J. Jakeman, editors. Transactions of the Second Biennial Meeting of the International Environmental Modelling and Software Society, Volume 2. International Environmental Modelling and Software Society (iEMSs), Manno, Switzerland.

Reckhow, K. H. 1990. Bayesian inference in non-replicated ecological studies. Ecology 71:2053–2059.

Renooij, S. 2001. Probability elicitation for belief networks: issues to consider. Knowledge Engineering Reviews 16: 255–269.

Rochester, W., A. Accad, S. Low Choy, V. Neldner, D. Pullar, and K. Williams. 2004. Final report UQ-EPA subregion classification project. Technical report. The University of Queensland, Brisbane, Australia.

Roloff, G. J., and B. J. Kernohan. 1999. Evaluating reliability of habitat suitability index models. Wildlife Society Bulletin 27:973–985.

Shepherd, G. G., and C. W. Kirkwood. 1994. Managing the judgmental probability elicitation process: a case study of analyst/manager interaction. IEEE Transactions on Engineering Management 41:414–425.

Siu, N., and D. Kelly. 1998. Bayesian parameter estimation in probabilistic risk assessment. Reliability Engineering and System Safety 62:89–116.

Smith, C. S., A. L. Howes, B. Price, and C. A. McAlpine. 2007. Using a Bayesian belief network to predict suitable habitat of an endangered mammal the Julia Creek dunnart (*Sminthopsis douglasi*). Biological Conservation 139:333–347.

Spetzler, C. S., and C.-A. S. Staël von Holstein. 1975. Probability encoding in decision analysis. Management Science 22(3):340–358.

Stepinski, T., and R. Vilalta. 2005. Digital topography models for Martian surfaces. IEEE GeoScience and Remote Sensing Letters 2(3):260–264.

Taylor, B. L., P. R. Wade, R. A. Stehn, and J. F. Cochrane. 1996. A Bayesian approach to classification criteria for Spectacled Eiders. Ecological Applications 6:1077–1089.

ter Braak, C. J. F., and R. S. Etienne. 2003. Improved Bayesian analysis of metapopulation data with an application to a tree frog metapopulation. Ecology 84:231–241.

Tversky, A., and D. Kahneman. 1973. Availability: a heuristic for judging frequency and probability. Cognitive Psychology 5:207–232.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: heuristics and biases. Science 185:1124–1131.

Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. Science 211:453–458.

Wade, P. R. 2001. Bayesian methods in conservation biology. Ecological Applications 6:1036–1046.

Walls, L., and J. Quigley. 2001. Building prior distributions to support Bayesian reliability growth modelling using expert judgment. Reliability Engineering and System Safety 74(2): 117–128.

Weber, M., and K. Borcherding. 1993. Behavioral influences on weight judgments in multiattribute decision making. European Journal of Operational Research 67:1–12.

Winkler, R. L. 1967. The assessment of prior distributions in Bayesian analysis. Journal of the American Statistical Association 62:776–800.

Wintle, B. A., M. A. McCarthy, C. T. Volinsky, and R. P. Kavanagh. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology 17(6):1579–1590.

APPENDIX

A table listing software for expert elicitation (*Ecological Archives* E090-017-A1).