REVIEW AND SYNTHESIS

A guide to eliciting and using expert knowledge in Bayesian ecological models

Petra M. Kuhnert,[1]***** **Tara G. Martin**[2] **and Shane P. Griffiths**[3]

[1]*CSIRO Mathematics, Informatics and Statistics, Waite Campus, Adelaide, Australia*

[2]*CSIRO Sustainable Ecosystems, 306 Carmody Road, St. Lucia 4067, Australia*

[3]*CSIRO Marine and Atmospheric Research, PO Box 120, Cleveland 4163, Australia*

Correspondence: E-mail: petra.kuhnert@csiro.au

Abstract

Expert knowledge in ecology is gaining momentum as a tool for conservation decision-making where data are lacking. Yet, little information is available to help a researcher decide whether expert opinion is useful for their model, how an elicitation should be conducted, what the most relevant method for elicitation is and how this can be translated into prior distributions for analysis in a Bayesian model. In this study, we provide guidance in using expert knowledge in a transparent and credible manner to inform ecological models and ultimately natural resource and conservation decision-making. We illustrate the decisions faced when considering the use of expert knowledge in a model with the help of two real ecological case studies. These examples are explored further to examine the impact of expert knowledge through 'priors' in Bayesian modeling and specifically how to minimize potential bias. Finally, we make recommendations on the use of expert opinion in ecology. We believe if expert knowledge is elicited and incorporated into ecological models with the same level of rigour provided in the collection and use of empirical data, expert knowledge can increase the precision of models and facilitate informed decision-making in a cost-effective manner.

Keywords

Bayesian models, bias, decision-making, expert judgement, expert opinion, prior information.

Ecology Letters (2010) **13**: 900–914

INTRODUCTION

There has been a recent surge in the use of expert knowledge in ecological models (Crome *et al.* 1996; Martin *et al.* 2005; Denham & Mengersen 2007; Griffiths *et al.* 2007; Mac Nally 2007; O'Leary *et al.* 2008; O'Neill *et al.* 2008; James *et al.* 2010). There are two reasons for this trend. First, the types of ecological questions being proposed, particularly those pertinent to formal decision-making, are characterized by uncertainty and paucity of empirical data. Even when data are available, they are invariably subject to error due to the size and complexity of ecological systems, resulting in parameter estimates with wide confidence intervals, leading to uninformative predictions. Second, decisions based on ecological studies focussing on conservation management of species and environmental risk assessments are often required urgently. In situations such as these where hard data are lacking yet management decisions are required, the use of expert knowledge may

provide a way forward. Yet for researchers wishing to use expert knowledge, questions remain regarding how to properly conduct an elicitation and use it in a model to address the ecological research question.

Although frequentist techniques are evolving to accommodate expert knowledge (e.g. Lele & Allen 2006), Bayesian methods are naturally suited to the incorporation of expert knowledge through 'priors'; probability distributions representing what is known about the variable (Gelman *et al.* 2003). In this study, we focus on using Bayesian models for incorporating expert opinion. Comprehensive summaries of Bayesian modeling have been described in the statistical literature and readers are encouraged to look at Gilks *et al.* (1996), Robert (2000) and Gelman et al. (2003), whereas Ellison (2004), McCarthy (2007), and Link & Barker (2010) provide summaries specifically for an ecology audience. In brief, Bayesian modeling consists of four key elements: a prior probability distribution capturing prior knowledge about a parameter; data on the parameters captured through

the likelihood; a model that describes the underlying process and incorporates both the likelihood and priors; and finally posterior estimates that result from combining the likelihood with the prior reflecting uncertainties captured by the model (McCarthy & Masters 2005; Cressie *et al*. 2009).

There has been considerable discussion about elicitation methods and how elicited information can be incorporated into a model as one or more priors to inform an analysis (Garthwaite & Dickey 1988; Steffey 1992; O'Hagan 1998). Garthwaite *et al*. (2005), O'Hagan *et al*. (2006) and Low-Choy *et al*. (2009) provide a comprehensive overview of elicitation approaches with the latter highlighting six key elements for conducting an elicitation: the purpose and motivation for the use of prior information, identifying relevant expert knowledge, development of the statistical model, encoding of priors and the management of uncertainty in the elicitation process. These studies have been useful from the point of view of providing example elicitation practices and general frameworks for the elicitation structure. However, there is no single publication that has provided a practical, enabling framework for conducting an elicitation that guides the researcher through the elicitation process.

In this study, we provide a comprehensive guide to conducting an elicitation to assist the researcher through a modeling process that explicitly considers expert knowledge and its impact in a model. We begin discussing whether a study would benefit from expert knowledge and then follow with a discussion around the modeling framework for incorporating expert opinion and the elicitation design. At each step of the elicitation, we discuss the choice of technique in the context of two real ecological examples. Through these examples, we explore the impact of expert knowledge through 'priors' in Bayesian models and offer practical solutions for managing bias in the elicitation. Finally, we offer some key recommendations that will empower researchers to incorporate expert information more robustly in their chosen modeling framework.

CONDUCTING AN ELICITATION

An expert is someone who has knowledge of the subject of interest gained through their life experience, education or training (Garthwaite *et al*. 2005). They play a critical role in situations that rely solely on expert data and therefore have a huge impact on the decision making process (Burgman 2005). In our experience, eliciting expert information needs careful structure, drawing on aspects of the social sciences (Gigerenzer 1996, 2002, 2007) to extract relevant information in an unbiased manner that is non-threatening to the expert. Furthermore, the process needs to align not only with the research question but with the model that will be

used to incorporate the expert information. We illustrate these concepts using two ecological case studies. The first case study investigates the impact of livestock grazing on Australian woodland birds using expert information from 20 ecologists (Kuhnert *et al*. 2005; Martin *et al*. 2005). The second case study focuses on estimating the abundance of pelagic fishes from gillnet captures where expert opinion from a single fish biologist was used to adjust estimates for net selectivity (Griffiths *et al*. 2007).

Will expert opinion be useful?

To determine whether expert opinion will add value to a model, it is necessary first to articulate the research question. The research question frames the problem, structures the model and identifies the data (expert and/or empirical) required for input into that model. Next, a stock take of available empirical data is required and the extent to which these data alone can provide an adequate solution to the research question determined. If expert knowledge could potentially inform the model, it is then necessary to ask whether sufficient resources (i.e. time and money) are available to elicit the expert knowledge. If there are insufficient resources, the elicitation could be severely compromised.

In the bird-grazing study, the research question targeted the impact of livestock grazing on 31 species of birds. Resource constraints limited the collection of data across different grazing regimes, to two seasons and it was acknowledged that this amount of data would be insufficient to draw strong inferences, particularly for less common species. As a result, bird ecology experts were identified and asked to participate in an elicitation exercise. Experts provided information on the impact of livestock grazing on bird abundance that was incorporated into a model along with the empirical data. In a second example, a single expert was approached to provide information about net selectivity for a targeted form of fishing using gillnets (Griffiths *et al*. 2007). Although data from gillnet catches in the Northern Prawn Fishery (NPF) in Australia had been collected, no data were available on adjusting the abundance estimates for net selectivity.

Setting up the modeling framework

Just as the collection of empirical data should align with the modeling framework, the use of elicited information in a model should also be given some consideration. Although it seems tempting to launch into the elicitation itself, structuring the model so it can adequately accommodate the elicited information can ensure a seamless analysis, from which inferences can be drawn and appropriate decisions made.

Martin *et al.* (2005) outline an experimental design for determining grazing impacts on birds. At each grazing level (low, moderate and high), the numbers of birds of each species was recorded. Given the nature of the data collected, a Bayesian generalized linear mixed model was structured with normally distributed random effects for each grazing regime (Kuhnert *et al.* 2005; Martin *et al.* 2005). To accommodate expert opinion, each random effect was re-parameterized to accommodate a shift in the mean and a rescaling of the precision as a measure of the impact of grazing. As experts were not comfortable expressing their beliefs directly as means, an indirect style of elicitation was designed. This involved eliciting a relative measure from 20 ecologists, where the experts responded qualitatively with an 'increase', 'decrease' or 'no change' in abundance under each grazing regime. The qualitative information could be translated quantitatively using the approach set out by Kuhnert *et al.* (2005) and Martin *et al.* (2005), providing statistical summaries that could be incorporated into the model.

In another example, Griffiths *et al.* (2007) outline a Bayesian framework for estimating the abundance of longtail tuna and other species of pelagic fishes using limited catch data in the NPF from gillnets that represent a passive form of fishing. Their approach develops an expression for the mean of a negative-binomial fixed effects model that incorporates the swim speed of fish (prior based on other published studies), net length, net soak time and the mean number of each species of pelagic fish caught. As mesh size can vary depending on the type and size of fish targeted, an estimate of the selectivity of the net was required to adjust the abundance estimate accordingly. As a result, the mean of the model was re-parameterized in terms of an observed mean, a true (unobservable) mean multiplied by a probability of capture. As the probability of capture could not be estimated from any surveys, an elicitation strategy was devised.

Epistemic uncertainty and natural variation

The subjective judgements captured from experts provide a measure of uncertainty that can be incorporated into a Bayesian model. This uncertainty may be epistemic (due to lack of knowledge), comprise natural variation or represent a combination of both. In most cases, the distinction between these two types of uncertainties is rarely addressed and is often confounded in the uncertainty elicited around each expert's response. However, future research effort could be informed by this distinction. For example if epistemic uncertainty outweighs natural variation then further research effort could reduce uncertainty, whereas if natural variation outweighs epistemic uncertainty, any research effort may be futile in reducing the uncertainty. In both the grazing-bird

and gillnet examples, the distinction between epistemic uncertainty and natural variation was not made explicit during the elicitation. Therefore, without proper consultation with the experts, we have no way of knowing what their uncertainty reflects, although the authors suggest that their models reflect epistemic uncertainty, more so than natural variation. This has sparked an ongoing discussion about the types of uncertainties and the importance of distinguishing between them in a model (Ferson & Ginzburg 1996; Winkler 1996; O'Hagan & Oakley 2004).
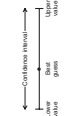
Elicitation design

We recognize that the design of an elicitation (i.e. number of experts, choice of experts and the elicitation technique) can be a daunting process. To assist, we have summarized eight commonly used elicitation approaches in ecological studies in Table 1 and provide key references that outline specific details regarding each approach. We rate the different approaches according to eight criteria that we believe are important in deciding which method to implement and discuss the pros and cons of each. These criteria attempt to evaluate: (1) the type of elicitation being implemented, i.e. whether it is a direct form of questioning or one that uses indirect techniques; (2) whether the study will benefit from the input of multiple experts; (3) whether the language used in the elicitation can be easily interpreted; (4) whether a high level of statistical expertise is required during the elicitation process; (5) whether knowledge of probability theory is required to answer the questions; (6) whether the method is appropriate for a remote style of elicitation, e.g. paper, web-based survey or telephone interview; (7) the time and ease of implementing the approach; and (8) whether uncertainty, or some measure of precision is obtained during the elicitation.

We discuss each of these criteria in the context of the number and choice of experts, their location, and the resources required to implement and conduct an elicitation. The latter in particular will often dictate the style of elicitation.

Single vs. multiple experts

The elicitation approaches summarized in Table 1 could be conducted with one or more experts. However, some approaches would benefit more from the interaction with multiple experts. For example, eliciting a probability (Method 1), a frequency (Method 2), a quantity (Method 3), a weight (Method 4), a categorical measure (Method 7) or a relative measure (Method 8) could be performed successfully using one expert but obtaining a measure of uncertainty (i.e. a precision) may be challenging. Having multiple experts available in these circumstances can be beneficial and may avoid having to explicitly ask the expert for their level of precision. For example, multiple responses

Table 1 Summary and synthesis of commonly used elicitation approaches used in ecological studies

Criteria/methods	Quantitative measure						Qualitative measure	
	(1) Probability	(2) Frequency	(3) Quantity, e.g. mean	(4) Weighting	(5) Quantitative interval	(6) Probability distribution	(7) Categorical measure	(8) Relative measure
Type of elicitation	Direct	Indirect	Direct	Indirect	Direct/indirect	Direct	Indirect	Indirect
Benefits from multiple experts*	✓	✓	✓	✓			✓	✓
Language easily interpreted	✓	✓	✓	✓			✓	✓
High level of expertise required	✓	✓	✓		✓	✓		
Knowledge of probability theory required						✓		
Can be performed remotely	✓	✓	✓	✓	✓		✓	✓
Fast and easy	✓	✓	✓	✓			✓	✓
Uncertainty explicitly specified					✓	✓		
Example	$p = 0.3$	x out of n trials	n	rank	[confidence interval graphic]	[density graphic]	low/medium/high	↑↓
Key references	Griffiths et al. (2007)	Gigerenzer (2002, 2007)	Christen & Nakamura (2000)	Mac Nally (2007)	O'Neill et al. (2008)	Kynn (2004)	Yamada et al. (2003)	Martin et al. (2005)

*Multiple experts can be beneficial in estimating a standard error. Alternatively, if one expert is used, a method for capturing the variance (5–6 above) will need to be considered.

can be aggregated by taking a mean, where the variance of the mean is a measure of uncertainty (epistemic and/or natural variation) amongst the expert responses. The benefit of using more than one expert is that the facilitator does not spend time trying to extract precision estimates from experts but the downfall is that the facilitator is then left with the task of synthesizing the expert information. To illustrate, consider the bird-grazing example. In this example, 20 bird experts were targeted because data on 31 species was required and no single expert could provide information on all species. For the gillnet study, a single fish biologist was identified and was available to participate, given the resources assigned to the project.

If only one expert is available then we recommend conducting a face-to-face style elicitation with a feedback cycle as used in the gillnet example. Although implementing a feedback process is an important part of any elicitation, it is particularly imperative in the case of the single expert to ensure consistency in their response. For the case of the single expert, we suggest engaging in Method 5 and eliciting a quantitative interval (Table 1) or if the expert has a reasonable knowledge of probability theory and is comfortable with the use of statistical jargon, engaging in Method 6, where a probability distribution is elicited. Although any of the approaches in this table could be easily implemented with one expert, Methods 5 and 6 explicitly targets the uncertainty around the quantity of interest (epistemic and/or natural variation), an important feature of any elicitation exercise.

To illustrate, in the gillnet problem, we chose to elicit two normal probability distributions that reflected the population size of longtail tuna and the distribution of fork lengths for that species caught in the net. This style of elicitation requires knowledge about probability distributions and can lead to a lengthy elicitation process if the expert does not have a sound grasp of probability theory. We also used a graphical aid (Fig. 1) as a form of feedback to the expert regarding their prior distributions about fork length size in the population (solid line) and in the catch (dashed line). In addition, the density of fish that were missed in the population (dotted line) and the resulting selectivity function (bottom plot of Fig. 1a and b) that was used to estimate the proportion of fish selected or caught by the gillnet was shown.

The expert found the graphical aid extremely useful, particularly the selectivity curve, for understanding the impact of the expert's priors as it showed the size distribution of fish caught by the net. We found Method 6 (Table 1) in conjunction with the graphical aid useful for the context of this problem, although, potentially Method 5 could also have been implemented. Note, as the modeling framework outlined in Griffiths et al. (2007) required a probability of capture, we could have elicited the probability
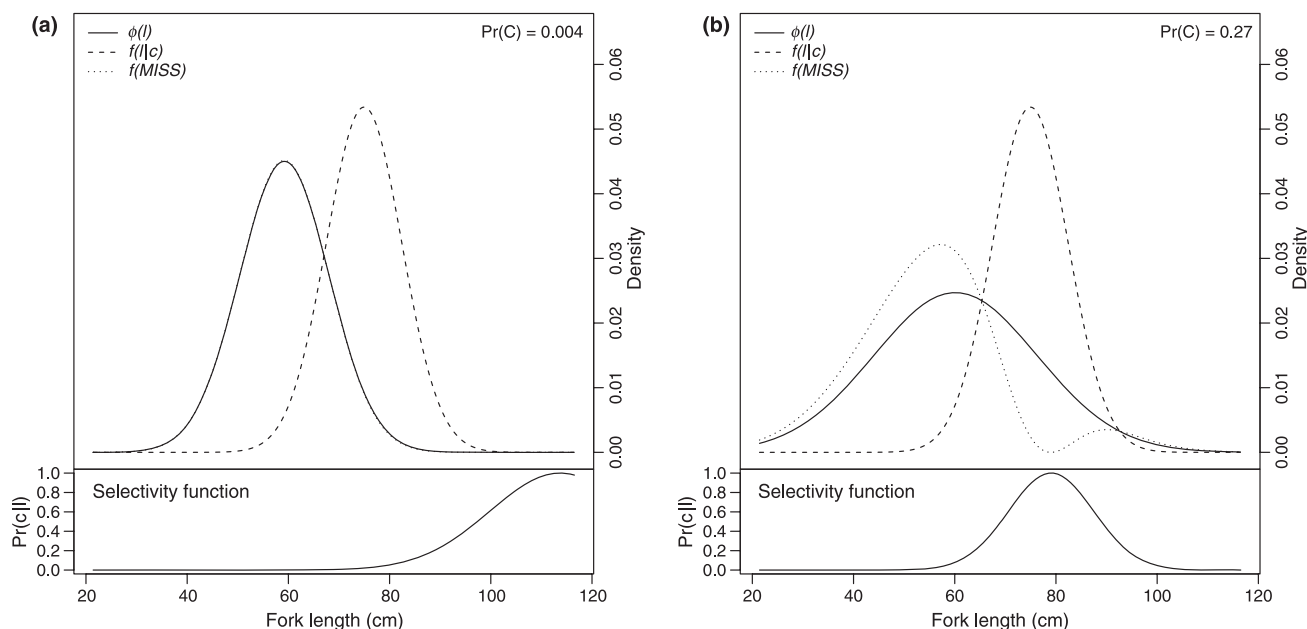
Figure 1 Prior distributions for the observed and population fork lengths of longtail tuna when (a) the expert gets it wrong and when (b) the expert gets it right. In each figure $\varphi(l)$ is the population density (solid line), $f(l\,|\,c)$ is the observed fork length density (dashed line), f(MISS) is the density of catches that were missed (dotted line) and Pr($c\,|\,$l) is the selectivity of the fish given its length. Note in (a) the distribution of $\varphi(l)$ and f(MISS) are the same. (Source: Griffiths *et al.* 2007).

directly using Method 1, Table 1. However, the biologist felt that they were not experienced enough to provide an opinion about the probability of capture directly. Nor could the expert quantify the uncertainty around this quantity with any confidence. As a result, we approached the problem using probability distributions indirectly, describing the population of pelagic fishes and the catch in the fishery, as the expert was comfortable providing estimates of uncertainty using this approach.

Choice of elicitation method
Choosing to elicit a probability, frequency, quantity or weighting/rank (Table 1, Methods 1–4) will depend largely on the type of experts chosen, the information required for the elicitation and how the prior is structured to incorporate this information in the model. If a probability is required, such as a probability of capture (Griffiths *et al.* 2007), a direct form of elicitation could be used. Although this could be easily performed remotely, the method does require the expert to have some knowledge of probabilities. An alternative approach is to avoid the use of the term 'probability' and frame the question in terms of a frequency statement (Gigerenzer & Hoffrage 1995; Gigerenzer 1996), e.g. out of *n* number of paddocks, how many (*x*) do you expect to be infested with the pest? The probability can then be formed as the ratio of these two frequencies: *x*/n. For this particular approach, the language is easily interpreted

and can be fast and easy to implement once the expert/s become comfortable with this style of elicitation. Reverting the expert back to the probability resulting from the elicitation is recommended as part of a feedback phase. Gigerenzer (1996) provides direct empirical evidence of the value of this approach to improve the accuracy and the calibration of an estimate.

If a specific quantity such as a mean (Method 3) is the focus of the elicitation, a direct form of elicitation may be used. Like a probability (Method 1), this approach can be relatively quick and easy to implement, providing the expert is comfortable expressing their opinion about the quantity of interest.

In situations where there are issues relating to language barriers, interpretability around the question being asked and the expert's experience with statistical jargon, methods that offer an indirect style of elicitation such as eliciting a weighting/rank (Method 4), a categorical response (Method 7) or asking for a relative measure (Method 8), may be more appropriate.

An example of an elicitation that relies on experts assigning weights to criteria is in marine risk assessment, where experts are asked to rank species according to criteria that assesses their susceptibility to capture and recovery, post-capture (Stobutzki *et al.* 2002). In this example, experts were asked to provide a ranking for each species (1–3) and a weighting (1–3) for each criteria in terms of the marine

indicator's importance in the overall assessment. Summaries of the expert weightings and rankings could then be incorporated into a risk assessment.

Categorical measures (Method 7) are often used to obtain a qualitative response such as 'the likely risk of a pest entering the country', or the categorization of birds into low, moderate and high abundance under a particular grazing regime (see Table 2). In most cases, this style of elicitation is chosen because it is easily interpreted by the expert.

Responses that provide a relative measure such as Method 8 also result in a qualitative response from the expert. Referring back to the bird-grazing example, experts were asked to respond about the impact of grazing on bird abundance and specifically, whether birds were likely to increase, decrease or show no change in abundance from

Table 2 Examples of indirect elicitations and how a prior probability distribution can be constructed as a result

The elicitation of a frequency

An example question: Considering n birds of species x occupying a region, how many would you observe under a highly grazed regime?

Construction of prior: The frequency can be converted to a proportion, p_i for each expert, i. Mean values and standard deviations can be calculated and a prior formed

The elicitation of a weighting/rank

An example question: Considering n birds of species x occupying a region, rank each bird from 1 to 5 according to how likely that bird would be observed under a highly grazed regime where 1 indicates not observed and 5 indicates always observed

Construction of prior: The median can be constructed from the expert rankings and a corresponding interquartile range can be calculated and a prior formed

The elicitation of a category

An example question: Considering n birds of species x occupying a region, would you expect a low, moderate or high number of that species to be observed in a highly grazed regime where low represents less than 5 birds per hectare, moderate represents between 5 and 15 birds per hectare and high represents greater than 15 birds per hectare

Construction of prior: The categories provided by the experts can be converted to numbers by taking the median of each category's quantitative representation. Summary statistics such as the mean and standard deviation can be calculated across experts and a prior can be formed

The elicitation of a relative measure

An example question: Considering n birds of species x occupying a region would you expect the bird to increase, decrease or show no change under high grazing pressure?

Construction of prior: The expert responses can be converted to a quantitative response [e.g. increase (+1), decrease (−1), no change (0)]. Values assigned will depend on the definition. Summary statistics such as the mean and standard deviation can be calculated across experts and a prior formed.

current baseline levels (Table 2). Like the categorical measure, this type of approach uses a language that is easily interpreted by the expert.

As highlighted by these examples, indirect elicitation approaches are often chosen because they avoid complicated statistical jargon. Although these approaches are useful for eliminating bias such as linguistic uncertainty (see Appendix S1) (Regan *et al.* 2002), they can be problematic when translating the qualitative response to a quantitative one for use in a Bayesian model. To illustrate the challenge, consider the grazing study of Martin *et al.* (2005). In this study, conservative values of 1, −1 and 0 respectively were assigned to the qualitative values outlined above. These values could be considered a type of expert weighting, where the larger the value assigned, the more emphasis placed on the expert information in the model. In this particular example, Martin *et al.* (2005) assigned a conservative weighting to the qualitative information. For other applications, a less conservative weighting could be applied. Ideally, we suggest assigning quantitative values to the qualitative responses prior to conducting the elicitation to avoid any potential bias resulting from the interpretation of the qualitative measure.

The Delphi approach for multiple experts

If more than one expert is required, we can consider some alternative methods for elicitation. The most commonly used elicitation method for multiple experts in ecology is the Delphi process (Delbecq *et al.* 1975; MacMillan & Marshall 2006). The Delphi approach uses feedback as a mechanism for helping experts understand the elicitation task, ensuring their response addresses the question adequately. The process begins with eliciting information from each expert independently. Results are then collated and shared amongst the group. Experts are then asked to reconsider their responses in light of the responses of others. This process of feedback and revision is continued until experts are satisfied with their respective responses.

Depending on available resources, the Delphi approach can be implemented remotely (e.g. Kangas & Leskinen 2005; O'Neill *et al.* 2008), or face-to face (e.g. Speirs-Bridge *et al.* 2010). However, if the resources are limited, then the elicitation can be conducted without any form of feedback but it has the potential to introduce bias because there is less opportunity for the expert to seek clarification or revise their response. An alternative would be to use a single expert and incorporate a feedback cycle into the elicitation to alleviate this bias. However, when using a single expert, there is the potential to select an expert that is not representative of the group of experts. The facilitator therefore needs to carefully consider the elicitation style when resources are limited. We now provide an example to illustrate this decision making process.

In the bird-grazing study, resources were available to involve multiple experts, but insufficient to conduct a proper Delphi style elicitation that incorporated a feedback cycle. As a result, the elicitation was conducted remotely through an email survey. The disadvantage of this approach is that it does not provide an opportunity for experts to review their responses in light of others and relies on experts having a good understanding of the questions being asked. Furthermore, response rates from this style of survey can be low as opposed to face-to face surveys and the facilitator needs to determine whether remote surveys are likely to yield an adequate response rate. Further discussion on response rates and pros and cons of different approaches appears in the psychology literature (Yammarino *et al.* 1991; Greenlaw & Brown-Welty 2009). For the bird-grazing study, 65% of experts targeted, responded to the survey.

The advantage of remote based survey designs is that they are cost and time effective for all involved. Depending on the length and complexity of the survey, it can be fairly straight forward and quick to complete. For example, experts participating in the bird-grazing elicitation were asked to tick a box corresponding to the most appropriate relative measure (increase, decrease, no change in abundance) for each bird species. For many ecological questions, experts will be more comfortable providing guidance on the direction of the response, as in Martin *et al.* (2005), rather than specifying a mean or frequency directly.

Forming a consensus

Forming a consensus of expert opinions is a challenging task as there is the potential to incorporate bias irrespective of whether opinions are synthesized using the Delphi approach or aggregated mathematically. A discussion of aggregation is beyond the scope of this review; however, summaries of popular methods used in opinion pooling and the implications of various approaches are discussed elsewhere (Genest & Zidek 1986; O'Hagan *et al.* 2006; Clemen & Winkler 2007; Scholz & Hansmann 2007).

The Delphi approach aims to circumvent undue influence of individual experts as opposed to using numerical aggregates. If a consensus is sought then the Delphi approach is appropriate and the facilitator conducting the elicitation needs to be experienced in dealing with human reasoning, perception, personalities and managing different forms of bias that can impinge on the group response (Burgman 2005; Kynn 2008). Often scientists may not be the best choice of facilitator in this type of setting. If the facilitator is capable of managing these types of bias then obtaining a group consensus can work well. However, in a group setting the facilitator must skilfully avoid the consensus being driven by a single convincing or influential expert.

Managing bias in an elicitation

Irrespective of the elicitation approach adopted in Table 1 or style of elicitation outlined in the previous sections, there are several heuristics, judgments and mental operations (Table 3) that the expert uses to base their response on, which have been highlighted by O'Hagan *et al.* (2006) and Kynn (2008). Of those listed, linguistic uncertainty – the uncertainty that arises because words have imprecise or different meanings (Regan *et al.* 2002; Hayes *et al.* 2007) – can contribute immensely to the uncertainty surrounding expert opinion. 'Getting the question right' is often the most difficult part of the elicitation and can lead to unwanted biases if not considered carefully. Regan *et al.* (2002) defines four types of linguistic uncertainty that can arise from an

Table 3 A summary of some of the key heuristics, judgments or mental operations that can result in bias when eliciting information from experts

Issues	Interpretation
Overconfidence/conservatism	Overestimating the accuracy of his/her beliefs or alternatively underestimating the uncertainty in a process. Conservatism relates to the process of an expert understating their belief
Representativeness	Providing opinions that are based on situations that are (wrongly or rightly) perceived to be similar
Availability	Basing a response on most recent available information and not considering past events
Anchoring and adjustment	The tendency for groups to anchor around (any) initial estimates and adjust their final estimate from this value irrespective of the initial estimates accuracy
Misunderstanding of conditional probabilities	Confusion regarding the definition of conditional probability and failure to adhere to the axioms of conditional probability
Translation	Confusion regarding the translation of a response to another scale
Affect	Expert's emotions entering into the judgment making
Hindsight bias	Expert places too much emphasis on past events and outcomes
Law of small numbers	Expert bases their opinion on small pieces of information and assumes that this extrapolates to the population
Linguistic uncertainty	Misunderstanding the question and/or applying different interpretations to the same term

elicitation: ambiguity, context dependence, underspecificity and vagueness (Appendix S1). As an illustration, consider an elicitation where the aim is to elicit a probability distribution representing the likelihood of a species becoming a weed. The word 'weed' is ambiguous and can have different meanings, e.g. exotic, undesirable, non-indigenous. The key is to be aware of the different forms of bias apparent in elicitations and where possible, try to avoid them through carefully structured elicitations.

INFLUENCE OF EXPERT INFORMATION

As highlighted in Burgman (2005), the perception of expert information is often one of honour and prestige and it is not uncommon for expert information to go unchallenged. The truth is that experts are invariably subject to bias and depending on the nature of that bias, their opinion may influence models and the decision-making process. It is therefore important to be aware of the impact that priors can have on models as this may influence the style of elicitation and choice of experts. In this section, we explore the influential nature of expert information in the context of a Bayesian model, where empirical data are captured through a *likelihood*, expert opinion is represented through a *prior* and the *posterior distribution* is the result of combining the likelihood with the prior to obtain inferences about the model parameters. Note, in all prior specifications, we work in terms of the precision, $\tau = 1/\sigma^2$ to be consistent with the notation used in standard Bayesian texts (Gelman *et al.* 2003).

There are several scenarios that can arise when combining the likelihood with priors generated from expert opinion. Both the amount of data, mean, precision and the way in which the prior mean and precision is captured and incorporated into a model can influence the posterior estimate. In situations where data are limited, the expert's opinion has the potential to drive model predictions. The facilitator therefore needs to be aware of the issues that can lead to bias and ensure that expert biases can be minimized.

As more data become available, the likelihood is moderated with the prior. However, in situations where the prior directly specifies the mean and precision, an informative prior can lead to a very informative posterior distribution, irrespective of the empirical data and how much data are collected (Lele & Allen 2006). If priors are incorporated into the model as an adjustment to an overall mean and precision, depending on their specification, the posterior estimates can be conservative. Here, the term adjustment refers to a shift in the mean or a rescaling of the precision, where the mean and precision are also considered random variables with appropriate priors attached.

To illustrate these concepts, we conducted two investigations using the ecological example described in Martin *et*

Table 4 Parameters used in the simulation study to investigate the impact of prior information in models with empirical data, where m_L and p_L are the mean and precision under low grazing respectively

Parameters	Values
Amount of data (likelihood)	Low ($n = 12$)
	Moderate ($n = 48$)
	High ($n = 192$)
Prior	m_L: −2, 2, −10, 10
	p_L: 0.5, 5, 50.
Incorporation of prior	Indirect specification
	Direct specification

al. (2005). We focus here on one bird species, the noisy miner (*Manorina melanocephala*) and consider two studies. The first investigation is a simulation study that examines the variation in the mean posterior estimates of abundance under three livestock grazing regimes when we alter the sample size (low, moderate or high), prior mean and precision and the type of prior (indirect or direct) used in the model. The second study took the raw data collected by Martin *et al.* (2005) and investigated changes to the posterior estimates for different means, precisions and sample sizes as outlined in Table 4 and compared them with the actual priors used in Martin *et al.* (2005).

Simulation study

Results of Martin *et al.* (2005) showed an increase in noisy miner abundance under high grazing and a decrease under low levels of grazing. In general, the expert data confirmed these predictions and tightened up credible intervals around estimates for both the moderate and high grazing regimes. Under low grazing, expert information did not alter the posterior estimate substantially.

We used the empirical data collected for the noisy miner to simulate scenarios of abundance (y_i) under low ($graze_{Li}$), moderate ($graze_{Mi}$) and high ($graze_{Hi}$) grazing levels collected at site i in a eucalypt woodland. Data were generated using a negative binomial (NB) distribution with a mean, θ_i and overdispersion parameter, φ that reflected estimates from Martin *et al.* (2005). The negative binomial density can be expressed as

$$f(y_i; \theta_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)\Gamma(y_i + 1)} \left(\frac{\phi}{\theta_i + \phi}\right)^{\phi} \left(\frac{\theta_i}{\theta_i + \phi}\right)^{y}, y = 0, 1, \ldots$$

where the overdispersion parameter, φ allows the variance of the distribution to exceed the mean.

We fit a Bayesian generalized linear model (GLM) (eqn 1) to 100 simulated datasets and estimated the grazing parameters, β_L, β_M and β_H for low (L), moderate (M) and high (H) grazing respectively. Prior information for the low

grazing parameter for both models was based on scenarios in Table 4. Prior information for the moderate and high grazing parameters (i.e m_M, p_M and m_H, p_H respectively) were taken from Martin *et al.* (2005) and represent expert summaries from a subset of the 20 experts (19 moderate and 18 high). We chose a gamma (Ga) prior for the overdispersion parameter and Normal (N) priors for the grazing parameters (eqn 2).

$$y_i \sim \text{NB}(\theta_i, \phi)$$
$$\log\theta_i = \beta_L \text{graze}_{Li} + \beta_M \text{graze}_{Mi} + \beta_H \text{graze}_{Hi} \qquad (1)$$
$$\phi \sim \text{Ga}(0.1, 0.1)$$

Depending on the specification of the prior information for the mean (m) and precision (p) for each grazing level (e.g. m_L, m_M, m_H, p_L, p_M, p_H), the priors placed on the parameters may take one of two forms. The first form, as shown in eqn 2 is the indirect specification (Method 8, Table 1), a relative measure that shifts the overall grazing mean, μ and rescales the precision, τ.

Prior for overall effect of grazing:
$$\mu \sim N(0, 0.1)$$
$$\tau \sim \text{Ga}(0.1, 0.1)$$
Incorporating expert information: $\qquad (2)$
$$\beta_L \sim N(\mu + m_L, \tau \times p_L)$$
$$\beta_M \sim N(\mu + m_M, \tau \times p_M)$$
$$\beta_H \sim N(\mu + m_H, \tau \times p_H)$$

The second form is shown in eqn 3 and represents a direct specification where the experts provide information about the mean and precision for each grazing parameter (Method 3, Table 1). Note, although the specification outlined in eqn 2 can be populated using information elicited directly, we used an indirect style of elicitation adopted by Martin *et al.* (2005) for the purpose of this simulation study.

Incorporating expert information:
$$\beta_L \sim N(m_L, p_L)$$
$$\beta_M \sim N(m_M, p_M) \qquad (3)$$
$$\beta_H \sim N(m_H, p_H)$$

Both models were fit using the R2WinBUGS package (Sturtz *et al.* 2005) in R (Ihaka & Gentleman 1996) and each simulation was run with a 10 000 burn-in followed by 10 000 monitored iterations (determined using standard diagnostic measures). We examined changes in the mean ($m_L = \pm 2, \pm 10$), precision ($p_L = 0.5, 5, 50$), sample size (low: $n = 12$, moderate: $n = 48$ and high: $n = 192$) and prior specification (indirect or direct) (Table 4) for each simulation and stored the parameter estimates for each level of grazing, the precision, standard deviation and 95% credible intervals. We also investigated whether there were

significant differences between low and moderate, and low and high grazed sites. Significance was concluded if the credible interval for the *difference* (log-scale) between the respective grazing regimes did not include zero.

Figures 2 and 3 show respectively changes in the standard deviation and proportion of times a significant difference is observed between estimates of low and moderate grazing. We observe how the variability in the estimate of β_L decreases as the sample size and precision increases across the 100 simulations when an indirect prior is assumed and a mean of ± 2 is elicited (Figs 2a and 3a). The proportion of significant results also increased, but only slightly. Results for the direct prior (mean of ± 2) (Figs 2b and 3b) are much more dramatic. Figure 2b shows that even with moderate or high amounts of data, the direct specification of the prior can be quite informative, even at low precision. This highlights the importance of the elicitation process, ensuring the accuracy in the prior specification.

We also repeated this simulation using prior means of ± 10 and found more striking results (not shown). In these scenarios, we found that the variation around the mean was much higher as the prior mean was chosen well outside the range of the data. Of more concern was the proportion of times a significant result was achieved. Virtually every scenario produced significant estimates nearly 100% of the time, indicating the strength of the prior, irrespective of the amount of data.

Exploration of the real data and prior information

In addition to the simulation study, we explored changes in the results from introducing different prior scenarios for the noisy miner (Table 4). Datasets were generated of varying size: low ($n = 12$), moderate ($n = 48$) and high ($n = 192$) by taking samples (with replacement) from the empirical data. Sampling was stratified to ensure balance across all three grazing regimes. The model was then fit to all three datasets using the same prior mean and precisions that were explored in the simulation study and specified in Table 4.

The results from the resampling exercise provided similar conclusions to those presented in Simulation study section. With the introduction of an indirect prior we observed that as the sample size increased from 12 to 192, the credible intervals around the posterior estimates for the low grazing parameter did not change considerably (Fig. 4) and were more in line with the estimate produced in Martin *et al.* (2005). Credible intervals narrowed with the incorporation of a direct prior suggesting that the estimates are strongly influenced by a directly specified prior. This is exacerbated further when a mean of ± 10 was explored with varying precision of 0.5, 5 and 50 (Fig. 5).
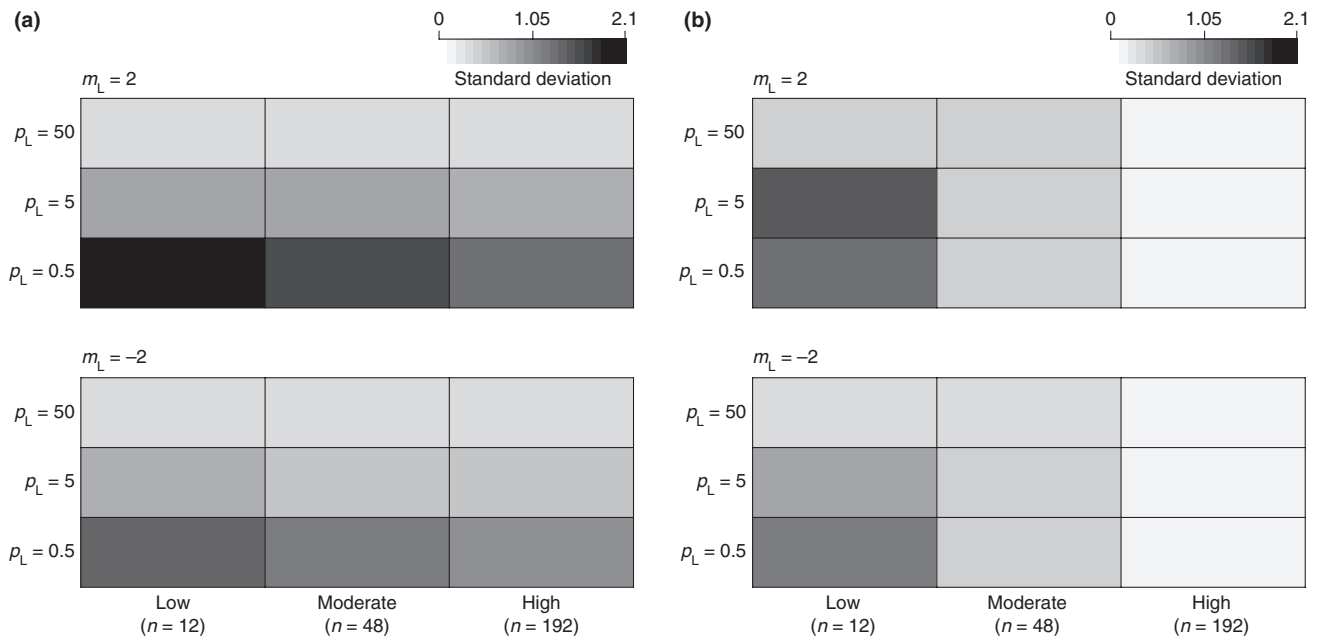
Figure 2 Plots showing the variation in the mean estimate of abundance under low grazing across different prior scenarios for (a) an indirect prior and (b) a direct prior specification. Here, the standard deviation is used as a measure of variation, m_L and p_L represent the prior mean and precision respectively at low grazing and n represents the number of observations. The darker the shading, the higher the standard deviation.
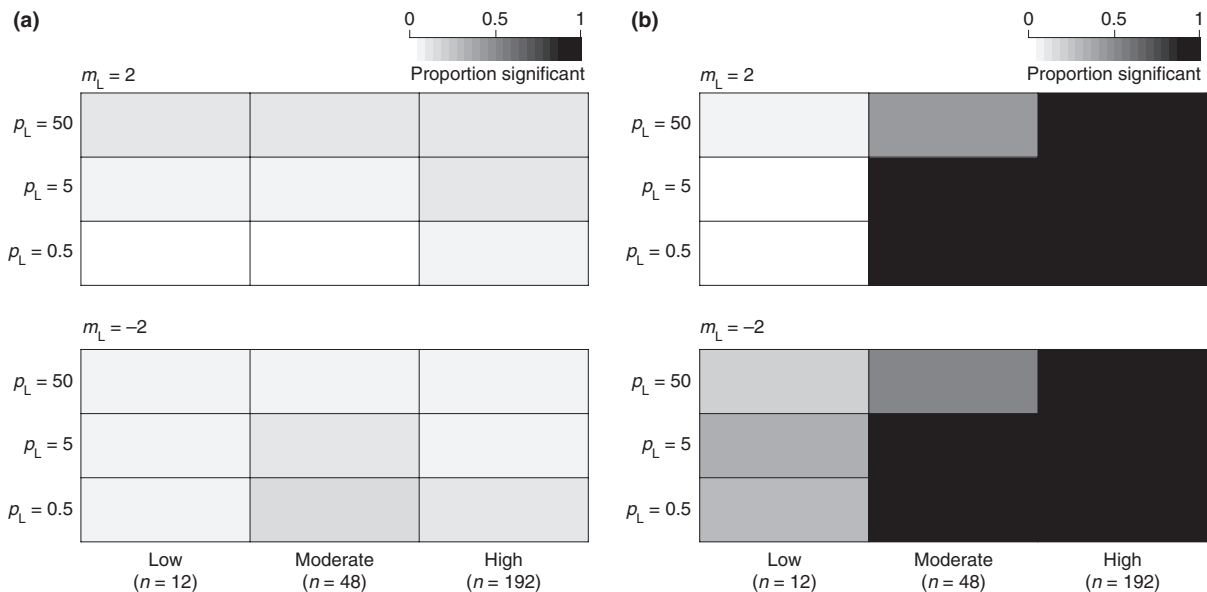

Figure 3 Plots showing the proportion of times the estimate of abundance under low grazing across different prior scenarios was significantly different from moderately grazed areas for (a) indirect prior and (b) a direct prior specification. An estimate was 'significant' if the 95% credible intervals did not include zero for the difference between coefficients. The darker the shading, the higher the proportion of significant results.

MANAGING BIAS IN AN EXPERT SETTING

One of the key criticisms of using expert knowledge in Bayesian ecological models concerns the possibility of an expert providing biased information. We have illustrated

that irrespective of the amount of experimental data and depending on the way in which the prior was elicited and incorporated into a model, the prior can have substantial impact. To illustrate the impact bias can have on the posterior estimates of a model, we describe a fisheries
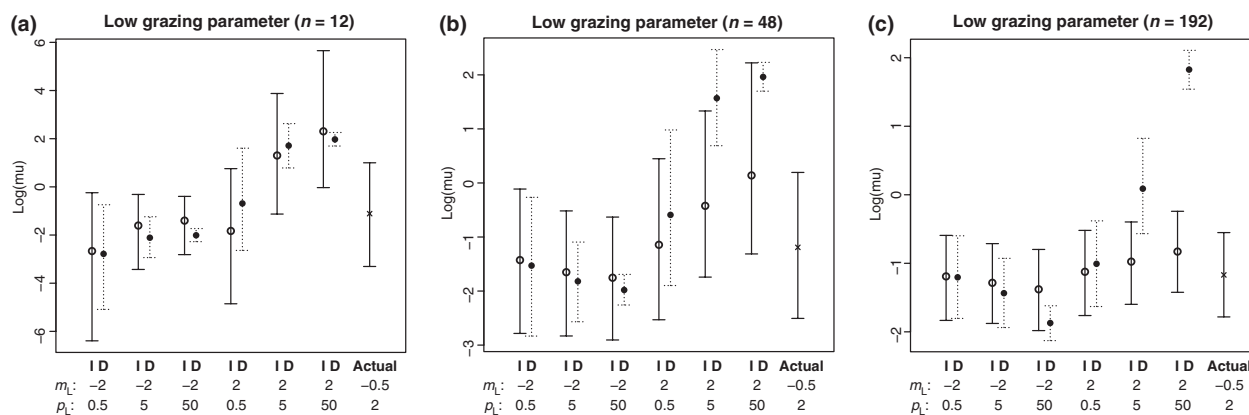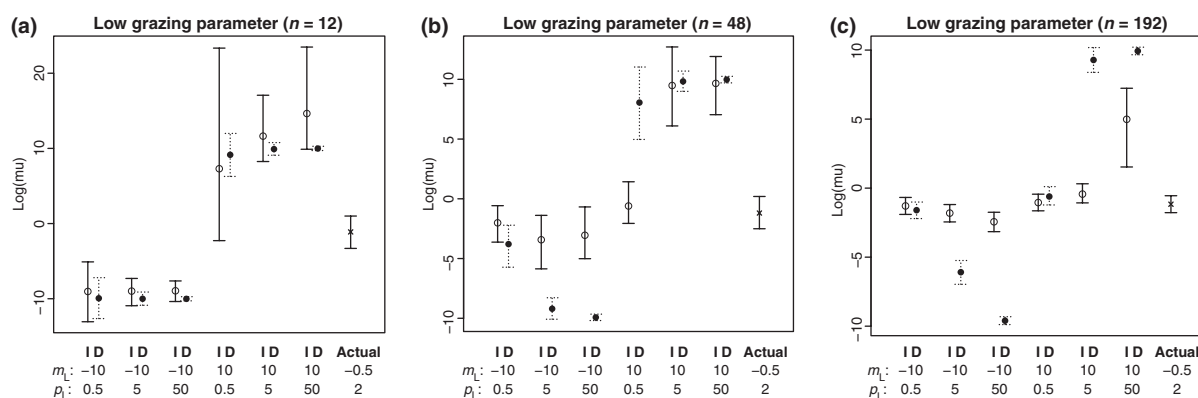
Figure 4 Posterior estimates from fitting the Bayesian model to the re-sampled empirical dataset and corresponding 95% credible intervals for the low grazing parameter in the Bayesian model of bird abundance for a sample size of (a) $n = 12$, (b) $n = 48$ and (c) $n = 192$. A prior mean of $\pm\,2$ is investigated, where the prior is either indirect (solid line) or direct (dotted line), where I = indirect and D = direct method of incorporating prior into model and Actual is the estimate calculated by Martin *et al.* (2005). Note, the scale on the *y*-axis on each plot is different.

Figure 5 Posterior estimates from fitting the Bayesian model to the re-sampled empirical dataset and corresponding 95% credible intervals for the low grazing parameter in the Bayesian model of bird abundance for a sample size of (a) $n = 12$, (b) $n = 48$ and (c) $n = 192$. A prior mean of $\pm\,10$ is investigated, where the prior is either indirect (solid line) or direct (dotted line), where I = indirect and D = direct method of incorporating prior into model and Actual is the estimate calculated by Martin et al (2005). Note, the scale on the *y*-axis on each plot is different.

problem where a direct specification of a prior by the expert produced an incorrect estimate of abundance (Griffiths *et al.* 2007).

The fishery problem

Biomass estimates of pelagic fishes in northern Australia were required for an ecosystem model intended to support ecosystem-based management of the NPF, one of Australia's largest and most valuable fisheries. However, catches of pelagic fishes in state-regulated shark gillnet fisheries are generally not recorded since they are of little or no value to the fisheries and collecting such information can be expensive and logistically difficult. One species that is

caught particularly well by these nets is one of the region's most abundant pelagic apex predators, longtail tuna (*Thunnus tonggol*).

In developing a model for abundance that takes into account the fishing method, Griffiths *et al.* (2007) considered an approach that combines geometry, Bayesian statistical modeling using a GLM framework and prior information to derive density estimates for longtail tuna in the NPF. A simplified version of the model is presented in eqn 4. The model is based on a domain of interaction that represents the region a fish resides in to have some interaction with the net. In the model formulation below, y_i represents the number of longtail tuna caught in the ith shot, ψ_i represents the observed mean, P_c represents the

proportion caught, μ_i represents the true mean and θ is the overdispersion parameter. The mean μ_i represents an average abundance where γ is the density and $A(h_i, d_i)$ is the area of the domain of interaction, where the mean number of fish represent the average number caught in the net. The area calculation comprises $p(d_i/h_i)$ representing the proportion of fish caught in the domain area, where h_i is the length of the net and d_i represents the radius, extending from the gillnet out to the edge of the domain. This radius is dictated by the length of the net and the swimming speed of the fish $(d_i = t_i s)$.

$$y_i \sim NegBin(\psi_i = P_c \mu_i, \theta)$$
$$\mu_i = \gamma \left\{ p(d_i/h_i)(2h_i d_i + \pi d_i^2) \right\} \qquad (4)$$
$$\quad = \gamma A(h_i, d_i)$$

The prior information for the probability of capture, P_c, is of interest here along with its role in moderating the likelihood in the model. For a more in-depth discussion of the model, readers are referred to Griffiths *et al.* (2007).

Priors and elicitation

The prior information for the model was captured in two forms. The first involved forming a prior for the swimming speed of fish using published data that provided information on the prior shape and structure. The second prior involved an intensive elicitation exercise to elicit the probability of capture, P_c. Gillnets are size-selective, which means their mesh size determines the size distribution of fish caught by the net. Due to limited resources and availability of experts, we enlisted the expertise of a single fish biologist and used Method 6 of Table 1 to elicit a response.

Griffiths *et al.* (2007) show that to evaluate P_c, we require information about the inverse of the ratio of the observed fork length density, $f(l|c)$ and the population density, $\varphi(l)$ of fork lengths at the maximum attainable fork length for longtail tuna. The latter prior is of interest here and needed to be elicited from the expert as no empirical data existed. A graphical tool was enlisted to help with the elicitation process.

When the expert gets it wrong

Catch data from the Queensland state N9 gillnet fishery (Griffiths *et al.* 2007) were used as an anchoring point to form a prior distribution for the observed fork length, $f(l|c)$ in collaboration with the expert. This resulted in a mean fork length for longtail tuna that was *c.* 75 cm with a variance of *c.* 56 cm.

The prior for the population density however, suggested quite a different distribution, with an elicited mean and variance of 59 and 79 cm respectively, indicating that a large proportion of longtail tuna were not being captured by gillnets in the N9 fishery. When the expert was approached about the elicited distribution for $\varphi(l)$, we found that the elicited response was based on the experts knowledge of historic Taiwanese catch data that were believed to be representative of the population due to the amount of information collected ($n = 10\ 224$ gillnet sets).

A summary of the initial prior distributions elicited from the expert are displayed in Fig. 1a along with the resulting selectivity function, $Pr(c|l)$, density of catches that were missed, $f(MISS)$ and the probability of capture, P_c. In this figure, the population density $\varphi(l)$ superimposes $f(MISS)$, which indicates that a large proportion of the population – representing smaller sized fish – is missed due to the specification of this prior. This is also reflected by the selectivity function shown in the bottom plot, which indicates that the probability of capture is highest for larger sized fish. It is obvious from this diagram that the expert miss-specified the prior for $\varphi(l)$ and was clearly driven by past events and outcomes (*hindsight bias*, Table 3) and exhibited *overconfidence* (Table 3) due to the fact that there was a large volume of Taiwanese catch data that the opinion was based on.

If we chose to accept this prior, the resulting probability of capture is 0.004, which is extremely low for any gillnet set in this fishery. This would lead to a density estimate of nearly 140 fish km^{-2}, an unrealistic estimate.

When the expert gets it right

The visual aid shown in Fig. 1a was a useful form of feedback for understanding the low probability of capture and subsequent large density estimate obtained from using the expert's initial approximation at $\varphi(l)$. Clearly this estimate was wrong and the prior distribution for $\varphi(l)$ needed to be revised.

Although further work is required to understand the biology and ecology of longtail tuna, Serventy (1956) indicated that juveniles exist in the northwest region of the NPF and undertake an ontogenetic migration to the Gulf of Carpentaria by the age of 2 years, where they are then captured by the N9 fishery. The prior for $\varphi(l)$ was therefore altered by the expert to reflect this information and the probability of capture was re-estimated as 0.2705 as shown in Fig. 1b. This provided a more realistic density estimate of 1.81 fish km^{-2} (Olson & Watters 2003).

DISCUSSION

Expert information can be a powerful tool in ecological models, particularly one that is limited by data availability. We have shown that the manner in which information is elicited and incorporated into models can have a large

impact on resulting estimates, generating controversy over the use of priors in Bayesian modeling (e.g. Dennis 1996; Lele & Allen 2006).

We have outlined a framework for conducting an elicitation that provides the researcher with a range of approaches for elicitation that can be translated into priors and incorporated into Bayesian models to address the research question. Our investigations have highlighted three main topics that warrant further evaluation. The first is around the use of multiple expert responses and their synthesis. Our research has shown the value of using multiple experts in an elicitation exercise as the aggregation of multiple responses leads to an estimate of the uncertainty around the elicited quantity. It also represents a natural mechanism for feedback through the discussion and revision of opinion amongst experts. One of the challenges however, is determining how to best aggregate multiple, independent responses. We believe this area is underdeveloped and although there are publications which present methods for expert synthesis, there is a need to evaluate and compare these methods to enable researchers to implement approaches easily and effectively.

A second area of future work is around uncertainty. Specifically, disentangling the different sources of uncertainty (linguistic, epistemic and variability) in an elicited response. To some extent, linguistic uncertainty can be tackled through the framing of the question, eliciting from more than one expert and the use of feedback and visual aids. However, epistemic uncertainty, which sometimes arises out of ignorance, subjectivity, incorrect structuring of a model or incorrect parameterization of the model is difficult to disentangle from variability, whether this is caused by natural variation or anthropogenically induced. Although recognized as important (Ferson & Ginzburg 1996; Winkler 1996; O'Hagan & Oakley 2004), further work is required to develop methodologies that enable different types of uncertainties to be captured, incorporated and interpreted from models.

The third area of investigation is the use of expert opinion for decision-making. Insufficient empirical data are often used as an excuse for lack of decision-making in the conservation and management of natural resources (Mangel et al. 2001). However expert knowledge is increasingly being used to inform the management of natural resources via formal decision models (e.g. Martin et al. in press). There is a need to evaluate whether decisions informed from expert-based models, models that use a combination of expert and empirical data or taking no decision as a result of insufficient empirical data, result in the best outcomes for natural resource management.

Our guide to eliciting and incorporating expert information into Bayesian models leads us to recommend the following key processes in an elicitation exercise:

(1) Clearly articulate the research question. This question will frame the study, collection of data (including expert data) and structure the model.

(2) Consider the resources available to help address the research question. Is there sufficient time and money to collect empirical data and/or conduct an elicitation? Are experts available and how many?

(3) Consider carefully the modeling framework for the environmental process under investigation. Identify the data, whether elicitation is required and how it will be structured into the Bayesian model.

(4) Identify what types of expert(s) are available to determine the form of elicitation required and whether a direct or indirect approach should be used and whether the use of statistical jargon is appropriate.

(5) Structure the elicitation such that the information supplied by experts can be translated into something (e.g. prior probabilities, prior distributions) that can be used for the model. Avoid doing this post-elicitation.

(6) In the case of the single expert, explore suitable methods for eliciting the uncertainty around their response and ensure a feedback process is implemented. For multiple experts, explore methods for synthesizing their responses and generating the uncertainty around the estimate.

(7) Incorporate a feedback mechanism with some form of graphical aid so the expert(s) can discuss what led to their opinion and give them the opportunity to revise their opinion as needed.

(8) Ensure a structured sensitivity analysis is conducted to investigate the impact of priors. Where empirical data are available, running the models with and without the influence of informative prior information is advised (Martin et al. 2005).

Although there is no guarantee the inclusion of expert opinion in a model will improve model predictions, in our experience when dealing with ecological models that inform conservation and natural resource management, in the absence of empirical data, basing decisions on models founded on expert opinion alone is preferable than delaying decisions until empirical data become available (e.g. Martin et al. in press). Following these recommendations will improve the use and efficacy of expert knowledge in ecological models and facilitate informed decision-making in a cost-effective manner.

ACKNOWLEDGEMENTS

REFERENCES

Burgman, M.A. (2005). *Risks and Decisions for Conservation and Environmental Management*. Cambridge University Press, New York.

Carey, J. & Burgman, M.A. (2008). Linguistic uncertainty in qualitative risk analysis and how to minimize it. *Ann. N. Y. Acad. Sci.*, 1128, 13–17.

Christen, J.A. & Nakamura, M. (2000). On the analysis of accumulation curves. *Biometrics*, 56, 748–754.

Clemen, R.T. & Winkler, R.L. (2007). Aggregating probability distributions. In: *Advances in Decision Analysis: From Foundations to Applications* (eds Edwards, W., Miles, R.F. & von Winterfeldt, D.). Cambridge University Press, Cambridge, pp. 154–176.

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

Crome, F.H.J., Thomas, M.R. & Moore, L.A. (1996). A novel Bayesian approach to assessing impacts of rain forest logging. *Ecol. Appl.*, 6, 1104–1123.

Delbecq, A.L., Van de Ven, A.H. & Gustafson, D.H. (1975). *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Scoot, Foresman and Company, Glenview, IL.

Denham, R. & Mengersen, K. (2007). Geographically assisted elicitation of expert opinion for regression models. *Bayesian Anal.*, 2, 99–136.

Dennis, B. (1996). Discussion: should ecologists become Bayesians? *Ecol. Appl.*, 6, 1095–1103.

Ellison, A.M. (2004). Bayesian inference in ecology. *Ecol. Lett.*, 7, 509–520.

Ferson, S. & Ginzburg, L.R. (1996). Different methods are needed to propagate ignorance and variability. *Reliab. Eng. Syst. Saf.*, 54, 133–144.

Garthwaite, P.H. & Dickey, J.M. (1988). Quantifying expert opinion in linear regression problems. *J. R. Stat. Soc. B Methodol.*, 50, 462–474.

Garthwaite, P.H., Kadane, J.B. & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *J. Am. Stat. Assoc.*, 100, 680–700.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Genest, C. & Zidek, J.V. (1986). Combining probability distributions: a critique and an annotated bibliography. *Stat. Sci.*, 1, 114–148.

Gigerenzer, G. (1996). The psychology of good judgment: frequency formats and simple algorithms. *Med. Decis. Making*, 16, 273–280.

Gigerenzer, G. (2002). *Calculated Risks: How to Know When Numbers Deceive You*. Simon and Schuster, New York.

Gigerenzer, G. (2007). *Gut feelings: The Intelligence of the Unconscious*. Viking, New York.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychol. Rev.*, 102, 684–704.

Gilks, W., Richardson, S. & Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Greenlaw, C. & Brown-Welty, S. (2009). A comparison of web-based and paper-based survey methods. *Eval. Rev.*, 33, 464–480.

Griffiths, S.P., Kuhnert, P.M., Venables, W.N. & Blaber, S.J.M. (2007). Estimating abundance of pelagic fishes using gillnet catch data in data-limited fisheries: a Bayesian approach. *Can. J. Fish. Aquat. Sci.*, 64, 1019–1033.

Hayes, K.R., Regan, H.M. & Burgman, M.A. (2007). Introduction to the concepts and methods of uncertainty analysis. In: *Environmental Risk Assessment of Genetically Modified Organisms, Volume 3: Methodologies for Transgenic Fish* (eds Kapuschinski, A.R., Hayes, K.R., Li, S. & Dana, G.). CABI Publishing, Oxfordshire, pp. 188–208.

Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Comput. Graph. Stat.*, 5, 299–314.

James, A., Choy, S.L. & Mengersen, K. (2010). Elicitator: an expert elicitation tool for regression in ecology. *Environ. Model. softw.*, 25, 129–145.

Kangas, J. & Leskinen, P. (2005). Modelling ecological expertise for forest planning calculations – rationale, examples and pitfalls. *J. Environ. Manage.*, 76, 125–133.

Kuhnert, P.M., Martin, T.G., Mengersen, K. & Possingham, H.P. (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. (2004). *Eliciting expert knowledge for Bayesian logistic regression in species habitat modelling*. Department of Statistics, Queensland University of Technology, Brisbane.

Kynn, M. (2008). The 'heuristics and biases' bias in expert elicitation. *J. R. Stat. Soc. Ser. A Stat. Soc.*, 171, 239–264.

Lele, S.H. & Allen, K.L. (2006). On using expert opinion in ecological analyses: a frequentist approach. *Environmetrics*, 17, 683–704.

Link, W.A. & Barker, R.J. (2010). *Bayesian Inference with Ecological Applications*. Academic Press, London.

Low-Choy, S., O'Leary, R. & Mengersen, K. (2009). Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology*, 90, 265–277.

Mac Nally, R. (2007). Consensus weightings of evidence for inferring breeding success in broad-scale bird studies. *Austral Ecol.*, 32, 479–484.

MacMillan, D.C. & Marshall, K. (2006). The Delphi process – an expert-based approach to ecological modelling in data-poor environments. *Anim. Conserv.*, 9, 11–19.

Mangel, M., Talbot, L.M., Meffe, G.K., Agardy, M.T., Alverson, D. & Barlow, J. (2001). Principles for the conservation of wild living resources. *Environ. Dev. Econ.*, 2, 39–110.

Martin, J., Nicols, J.D., Runge, M.C., McIntyre, C., Lubow, B.L., McCluskie, M.C. & Schmutz, J.A. (in press). An adaptive management framework for optimal control of recreational activities in Denali national park. *Conserv. Biol.*.

Martin, T.G., Kuhnert, P.M., Mengersen, K. & Possingham, H.P. (2005). The power of expert opinion in ecological models using

Bayesian methods: Impact of grazing on birds. *Ecol. Appl.*, 15, 266–280.

McCarthy, M.A. (2007). *Bayesian Methods for Ecology*. Cambridge University Press, New York.

McCarthy, M.A. & Masters, P. (2005). Profiting from prior information in Bayesian analyses of ecological data. *J. Appl. Ecol.*, 42, 1012–1019.

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

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E. & Rakow, T. (2006). *Uncertain judgements: Eliciting Expert Probabilities*. Wiley, Chichester, UK.

O'Hagan, A. & Oakley, J.E. (2004). Probability is perfect, but we can't elicit it perfectly. *Reliab. Eng. Syst. Saf.*, 85, 239–248.

O'Leary, R.A., Low Choy, S., Murray, J.V., Kynn, M., Denham, R., Martin, T.G. & Mengersen, K. (2008). Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata. *Envirometrics* (in press). Available at: http://www3.interscience.wiley.com/cgi-bin/fulltext/120748019/PDFSTART.

O'Neill, S., Osborn, T., Hulme, M., Lorenzoni, I. & Watkinson, A. (2008). Using expert knowledge to assess uncertainties in future polar bear populations under climate change. *J. Appl. Ecol.*, 45, 1649–1659.

Olson, R.J. & Watters, G.M. (2003). A model of the pelagic ecosystem in the eastern tropical Pacific Ocean. *Inter Am. Trop. Tuna Comm. Bull.*, 22, 133–219.

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

Robert, C. (2000). *The Bayesian Choice*, 2nd edn. Springer, New York.

Scholz, R.W. & Hansmann, R. (2007). Combining experts' risk judgments on technology performance of phytoremediation: self-confidence ratings, averaging procedures, and formative consensus building. *Risk Anal.*, 27, 225–240.

Serventy, D.L. (1956). Additional observations on the biology of the northern bluefin tuna, Kishinoella tonggol (Bleeker) in Australia. *Aust. J. Mar. Freshw. Res.*, 7, 44–63.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G. & Burgman, M.A. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Anal.*, 30, 512–523.

Steffey, D. (1992). Hierarchical Bayesian modeling with elicited prior information. *Commun. Stat. Theory Methods*, 21, 799–821.

Stobutzki, I.C., Miller, M.J., Heales, D.S. & Brewer, D.T. (2002). Sustainability of elasmobranchs caught as bycatch in a tropical prawn (shrimp) trawl fishery. *Fish. Bull.*, 100, 800–821.

Sturtz, S., Ligges, U. & Gelmen, A. (2005). R2WinBUGS: a package for running WinBUGS from R. *J. Stat. Softw.*, 12, 1–16.

Winkler, R.L. (1996). Uncertainty in probabilistic risk assessment. *Reliab. Eng. Syst. Saf.*, 54, 127–132.

Yamada, K., Elith, J., McCarthy, M. & Zerger, A. (2003). Eliciting and integrating expert knowledge for wildlife habitat modelling. *Ecol. Model.*, 165, 251–264.

Yammarino, F.J., Skinner, S.J. & Childers, T.L. (1991). Understanding mail survey response behavior a meta-analysis. *Public Opin. Q.*, 55, 613–639.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Linguistic uncertainty can lead to bias when eliciting information from experts (adapted from Regan *et al.* 2002; Carey & Burgman 2008).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.