Best practices for the provision of prior information for Bayesian stock assessment

ICES CIEM

International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

ICES Cooperative Research Report

Best practices for the provision of prior information for Bayesian stock assessment

Editor

Atso Romakkaniemi

Authors

Charis Apostolidis • Guillaume Bal • Rainer Froese • Juho Kopra
Sakari Kuikka • Adrian Leach • Polina Levontin • Samu Mäntyniemi
Niall Ó Maoiléidigh • John Mumford • Henni Pulkkinen
Etienne Rivot • Atso Romakkaniemi• Vaishav Soni
Konstantinos Stergiou • Jonathan White • Rebecca Whitlock

ICES CIEM

International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

This document is the product of an Expert Group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the view of the Council.

Contents

Executive summary

This manual represents a review of the potential sources and methods to be applied when providing prior information to Bayesian stock assessments and marine risk analysis. The manual is compiled as a product of the EC Framework 7 ECOKNOWS project (www.ecoknows.eu).

The manual begins by introducing the basic concepts of Bayesian inference and the role of prior information in the inference. Bayesian analysis is a mathematical formalization of a sequential learning process in a probabilistic rationale. Prior information (also called "prior knowledge", "prior belief", or simply a "prior") refers to any existing relevant knowledge available before the analysis of the newest observations (data) and the information included in them. Prior information is input to a Bayesian statistical analysis in the form of a probability distribution (a prior distribution) that summarizes beliefs about the parameter concerned in terms of relative support for different values. Apart from specifying probable parameter values, prior information also defines how the data are related to the phenomenon being studied, i.e. the model structure. Prior information should reflect the different degrees of knowledge about different parameters and the interrelationships among them.

Different sources of prior information are described as well as the particularities important for their successful utilization. The sources of prior information are classified into four main categories: (i) primary data, (ii) literature, (iii) online databases, and (iv) experts. This categorization is somewhat synthetic, but is useful for structuring the process of deriving a prior and for acknowledging different aspects of it.

A hierarchy is proposed in which sources of prior information are ranked according to their proximity to the primary observations, so that use of raw data is preferred where possible. This hierarchy is reflected in the types of methods that might be suitable – for example, hierarchical analysis and meta-analysis approaches are powerful, but typically require larger numbers of observations than other methods. In establishing an informative prior distribution for a variable or parameter from ancillary raw data, several steps should be followed. These include the choice of the frequency distribution of observations which also determines the shape of prior distribution, the choice of the way in which a dataset is used to construct a prior, and the consideration related to whether one or several datasets are used. Explicitly modelling correlations between parameters in a hierarchical model can allow more effective use of the available information or more knowledge with the same data. Checking the literature is advised as the next approach. Stock assessment would gain much from the inclusion of prior information derived from the literature and from literature compilers such as FishBase (www.fishbase.org), especially in data-limited situations. The reader is guided through the process of obtaining priors for length–weight, growth, and mortality parameters from FishBase. Expert opinion lends itself to data-limited situations and can be used even in cases where observations are not available. Several expert elicitation tools are introduced for guiding experts through the process of expressing their beliefs and for extracting numerical priors about variables of interest, such as stock–recruitment dynamics, natural mortality, maturation, and the selectivity of fishing gears. Elicitation of parameter values is not the only task where experts play an important role; they also can describe the process to be modelled as a whole.

Information sources and methods are not mutually exclusive, so some combination may be used in deriving a prior distribution. Whichever source(s) and method(s) are chosen, it is important to remember that the same data should not be used twice. If the

plan is to use the data in the analysis for which the prior distribution is needed, then the same data cannot be used in formulating the prior.

The techniques studied and proposed in this manual can be further elaborated and fine-tuned. New developments in technology can potentially be explored to find novel ways of forming prior distributions from different sources of information. Future research efforts should also be targeted at the philosophy and practices of model building based on existing prior information. Stock assessments that explicitly account for model uncertainty are still rare, and improving the methodology in this direction is an important avenue for future research. More research is also needed to make Bayesian analysis of non-parametric models more accessible in practice. Since Bayesian stock assessment models (like all other assessment models) are made from existing knowledge held by human beings, prior distributions for parameters and model structures may play a key role in the processes of collectively building and reviewing those models with stakeholders. Research on the theory and practice of these processes will be needed in the future.

1 Introduction: why priors are logically necessary

Samu Mäntyniemi and Atso Romakkaniemi

1.1 Scientific reasoning in stock assessment for fishery management

ICES advises competent authorities on marine policy and management issues related to the sustainable use of living marine resources and the impacts of human activities on marine ecosystems. In order to achieve this for fishery management, experts for a fish stock carry out a scientific analysis to estimate, where possible, historic and current fishing mortality, recruitment, and population size. For most stocks with population size estimates, experts can forecast future stock size as a function of a management action (e.g. catch) and calculate what action would lead to a desired management objective (e.g. fishing mortality not exceeding that corresponding to maximum sustainable yield). The estimate of current population size is the starting point for forecasting; thus, it has a central role in the process (ICES, 2013).

Given suitable technology, current population size could, in principle, be observed and known without error. In practice, it cannot be directly observed and needs to be inferred from indirect and incomplete information. This inference (estimation) is based on what scientists making the assessment believe about the relationship between true population size and observable information.

In contrast to current population size, future consequences of management actions cannot be observed, but can only be imagined (predicted). The result of this prediction depends, again, on the understanding of the scientists making the assessment.

The key point to acknowledge is that estimation and prediction require human interpretation of available information. This interpretation obviously differs between scientists and their amount of experience and knowledge. Differences in interpretation are often linked to (even minor) differences in the exact areas of specialization and experience, which lead scientists to approach research questions from slightly different perspectives. Because of the importance of this expertise, stock assessment tasks are typically entrusted to groups of experts who comprise the membership of ICES stock assessment working groups.

Broadly standardized approaches across stocks may either fail or be suboptimal because they cannot sufficiently recognize the specificity of each fish stock, and information available on individual stocks also varies (ICES, 2013).

In other words, stock assessment is always, regardless of the analytical methods used, a subjective interpretation of observed data made by a group of experts. In this context, "subjective" refers to things that exist only in the human mind, and "objective" refers to the actual state of the physical world. This means that only the raw observations can be regarded as objective facts. Any inferences about unobservable phenomena, such as stock size, require a human assumption about the relationship between stock size and observed data.

In this context, subjectivity does not entail or imply that experts would be invited to consider their personal preferences and valuation of a stock's status or their political views about how a stock should be managed. It is only their personal (i.e. subjective) scientific experience and understanding that are sought in synthesizing their existing knowledge and interpretation of observed data. Experts are expected to place themselves into a neutral and independent position regarding stock assessments and to use their personal knowledge (i.e. expertise) honestly. In everyday language, experts are

expected to be objective when assessing the status of a stock using their own subjective scientific knowledge.

In principle, a stock assessment working group could "eyeball" the data spreadsheets, discuss with their colleagues, and use their collective wisdom to derive an honest and independent assessment of the stock without using any analytical methods. However, most assessment problems are too complex for the human mind to grasp as a whole. Some form of predefined logic, organized analysis, and synthesis is usually necessary to perform the task itself and also for the transparency of the expert group's work.

Use of mathematical models has a long history in fishery stock assessment. They have been used to organize and structure the thinking of expert groups. The types of models used have ranged from yield-per-recruit analysis to virtual population analysis and further towards integrated state–space models. Modelling practices have been influenced by developments in theoretical population dynamics and statistical data analysis. The purpose of statistical data analysis is to describe a set of observed data with the fewest parameters without losing too much of the structure of the original data. Transferring this approach to stock assessment has led to the idea that the parameters of biological population dynamics models should all be statistically estimable from stock assessment data. However, biologically realistic population dynamics models tend to have so many parameters that their estimation from most stock assessment datasets is impossible. A common attempt to resolve this problem has been to reduce the number of estimated parameters by simplifying the model structure and then assume that some of the parameters are not uncertain, but are instead known exactly. The parameters still treated as unknown after this process can be estimated from the data, but at the cost of having decreased the biological credibility of the population dynamics model and the unwarranted exclusion of part of the uncertainty. Knowledge used to finalize model parameters usually comes from the expertise of the stock assessment working group: group members use their biological knowledge to produce best guesses for parameters, such as the natural mortality rate (M) and the maximum size of fish, by interpreting data and conclusions presented in scientific papers, reports, and biological databases. In this way, information from outside the stock assessment database is treated as if it were exactly known, with the consequence that the information contained in the assessment data about these parameters cannot contribute to their estimation. Conversely, similar external information for the parameters estimated within the stock assessment is often not used at all even when it is available. Interestingly, it has been common practice to treat M as fully known and fishing mortality (F) as completely unknown, even though the latter is under human control and the former is not.

The Bayesian approach to scientific reasoning has been suggested as a remedy for this dichotomy in the use of existing knowledge. Instead of having to choose whether, based on existing information, a parameter is treated as precisely known or completely unknown, the expert group has the option to describe and quantify their uncertainty about each parameter. Some of the model parameters may be reasonably well, but not exactly, known, while other parameters are less precisely understood, but not completely unknown either. Probability distributions are used to reflect the different degrees of knowledge about different parameters and the interrelationships among them. At the heart of the Bayesian approach is the idea of presenting everything that is not known exactly as a probability distribution.

Bayesian inference has been successfully applied in a growing number of stock assessments. For example, since mid-2000, Atlantic salmon (*Salmo salar*) stocks in the Baltic Sea have been assessed by an ICES working group, and the consequent advice has been

provided by using sequential Bayesian analyses (Michielsens *et al.*, 2008; ICES, 2014). Several stocks of Pacific salmon have been assessed by Bayesian assessment models (C. Michielsens, Pacific Salmon Commission, pers. comm.). A Bayesian version of a surplus production model has been applied for the assessments of northern shrimp (*Pandalus borealis*) in the Barents Sea and Greenland halibut (*Reinhardtius hippoglossoides*) in Iceland and East Greenland (ICES, 2012a,b).

The Bayesian approach allows rigorous specification and utilization of relevant knowledge existing outside the primary input data used in stock assessment. This knowledge goes by various names: "prior information", "prior knowledge", "prior belief", or simply "prior". The next section presents the basic concepts and terminology of Bayesian statistical inference. Readers not familiar with the Bayesian approach should acquaint themselves with this information in order to be able to fully absorb the contents of this report.

1.2 The basic concepts of Bayesian inference and the role of prior information

In Bayesian analysis, the concept of probability is used to measure how strongly a person believes a particular hypothesis to be true. These hypotheses can consider past and future size of the fish stock, parameters of the population dynamics model, and different causal relationships represented by different model structures. A fundamental assumption is that one true stock size exists, and the Bayesian probability is used to express what a person thinks about this value. In other words, stock status exists objectively, but knowledge about it is inherently subjective without an objectively true value. For causal models and their parameters, such as F and M, the distinction between subjectivity and objectivity is not equally clear. Causal models and their parameters are also constructions of the human mind, but can be seen to represent an accepted version of reality; accordingly, they can be thought to have one true value which the person may not know exactly.

Bayesian probability is personal and, therefore, requires specification about whose probability is being used. In the case of ICES stock assessment, the most natural unit is the group of experts. The outcome of Bayesian stock assessment made by an expert group is a set of probability distributions describing their beliefs about the future status of the fish stock under each alternative management action. Ideally, these distributions summarize everything that the group knows, including the group's interpretation of the stock assessment data. The group has typically used a variety of modelling techniques to organize their thinking and to keep the inferences logical and transparent.

Probabilities given by the expert group cannot be challenged; they represent what this particular group thinks, and different probabilities would be expected from another group. This is the key feature of the assessment; it is the honest view of this particular group of trusted experts. If the expertise was not expected to affect the assessment, Chinese berry pickers could be hired to compile the assessment of European fish stocks.

Bayesian inference can perhaps be most easily understood as a learning or inductive process. Let us consider a phenomenon which, at first, we "know" nothing about (we have no prior beliefs whatsoever about the matter, which leads us to consider that all possible hypotheses are equally probable). When making our very first observation of the phenomenon, we intuitively realize that:

- Before making the first observation, any state of the nature of the phenomenon is considered to be equally probable;

- Given the observation, we then want to consider whether a particular state or states of the nature of the phenomenon can be considered to be more probable than others;

- Based on only one observation, we are still very uncertain about the phenomenon.

By the above process, we update our knowledge about the phenomenon. This learning process can be formalized using the Bayes' theorem (Berger, 1985; Gelman *et al.*, 2004):

$$P(H = h_i|E = e) = \frac{P(E=e|H=h_i) \times P(H=h_i)}{\sum_j P(E=e|H=h_j) \times P(H=h_j)} \qquad (1)$$

The denominator in equation (1) is a sum over all the possible hypotheses and, therefore, only depends on the data. The Bayes theorem is, therefore, often written in its simpler and unnormalized form:

$$P(H = h_i|E = e) \propto P(E = e|H = h_i) \times P(H = h_i) \qquad (2)$$

Here, h_i stands for a particular hypothesis in a set of hypotheses $(h_1, \dots, h_J)$ to be examined and e stands for evidence (observation or data). $P(H = h_i|E = e)$ is the probability that the hypothesis h_i is true given the evidence, i.e. the posterior probability. $P(E = e|H = h_i)$ is the probability of observing the data e given the hypothesis h_i is true (i.e. the *likelihood*). $P(H = h_i)$ is the prior distribution, representing prior knowledge about the hypothesis. In the situation described above, our prior knowledge is "uninformative", i.e. we set the same probability for this hypothesis to be true as for any other plausible ones. When we acquired the first observation, we could not avoid going through the inductive process described above, be it consciously or subconsciously. The principle of Bayesian conditionalization (Howson and Urbach, 1993) states that as soon as we have new evidence, we update our belief (knowledge) about the hypothesis.

Consider now that after updating our knowledge with the first observation, we obtain a new observation. Again, we intuitively go through the previous steps of induction. This time, we realize that we want to update our knowledge based on the new observation without forgetting the information included in the first observation. **Hence, our prior distribution must logically be informative.** That is, the prior $P(H = h_i)$ corresponds to the knowledge that we had after the first observation. After updating with the new observation, the resulting posterior distribution contains the information included in both observations. Although this example of Bayesian inference is, of course, naive, it helps us to appreciate that very seldom is there no prior knowledge about the matter that we want to examine. Therefore, it is often clearly reasonable to use an informative prior distribution in the process instead of trying to ignore or pretend that it does not exist.

A practical problem arises, however, in specifying and quantifying prior information (i.e. how to do this?). Researchers usually base most of their knowledge on information from published literature, earlier studies, pilot experiments, etc. One way to quantify prior information is to conduct an analysis of the existing, documented observations. This may be performed by using primary data, if available, or by conducting a meta-analysis of research results published in the literature. Online databases may help in the compilation of relevant information from the literature, and they may also include processed information not published elsewhere (e.g. detailed results of analyses carried out in ICES assessment working groups). If information relevant for quantifying prior knowledge is diverse and/or non-commensurate as such, elicitation of expert opinions may be utilized.

1.3 Worked example of a Bayesian stock assessment

In this section, we introduce the concepts and logic of Bayesian inference in the context of a simple stock assessment example. We walk through the steps of solving the problem using Bayesian inference and also compare the approach to classical (or frequentist) statistical approaches based on maximum likelihood that might have been used with this problem.

Our problem is to infer the size of a fish population (N) using the method of removal sampling: fish are removed from the population in successive passes, with knowledge about the fishing effort used at each pass (Mäntyniemi *et al.*, 2005; Rivot *et al.*, 2008).

The Bayesian approach starts by problem framing: what are the relevant variables? Obviously the population size (N) must be involved and also the number of fish caught (c_i) at each pass i. Clearly, fishing effort (E) and efficiency of the fishing technique (q) need to be taken into account. To keep the example simple, we assume that there are no births or deaths occurring during the sampling experiment.

The next step is to establish the causal connections between the relevant variables. It is a useful practice to draw a picture where all the variables are connected using arrows that point from causes to effects. The resulting pictures are called directed acyclic graphs (DAGs) (Spiegelhalter *et al.*, 1996). The DAG of this problem (Figure 1.1) shows that the first catch (c_1) is caused by the initial abundance, fishing effort, and the efficiency factor. The catch obtained on the second pass depends on the same variables, but also the catch removed from the population on the first pass is affecting it.
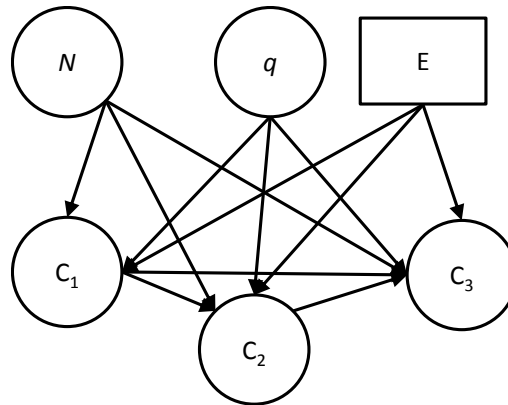
Figure 1.1. Directed acyclic graph of the removal fishing problem. Round shapes denote uncertain quantities, boxes represent known or controllable quantities, and arrows denote the direction of the causal relationship.

So far, we have used our existing understanding of the problem and created a mind map about how we think the problem works. The next phase is to start defining the mind map in more detail. How well do we already know what the values of these variables are? The Bayesian approach is to use probability statements for this purpose; nearly impossible values should be given smaller probabilities than values that we regard more credibly. Another way of thinking about the existing knowledge of a particular value, e.g. population size, is to consider the degree of disbelief; how surprised would we be if the true population size were actually 1000?

Information by which we would judge the degree of belief (or disbelief) can stem from multiple sources. Our minds could interpret data observed somewhere else in a similar situation and process the information we have adopted from the scientific literature or from discussions with colleagues. Our theoretical understanding of the subject area is likely working as "glue" by which we would synthesize past experience into a set of

probability statements about population size and other parameters of interest. The purpose of this document is to discuss the processes and practices for deriving and formulating existing knowledge as a probability distribution. These distributions are called prior distributions or simply "priors".

In practice, priors typically have a parametric form, which means that the shape and location of the distribution is determined by one or more parameters. Based on those parameters, probabilities for all potential values of the variables can be calculated. Thus, it is commonplace to use well known distributions such as the normal and log-normal distributions to describe existing knowledge.

In this example, we use a log-normal distribution for population size (Figure 1.2, upper-left caption). The most credible population size is 170 individuals. Population size is believed to lie between 106 and 376, with probability 0.8.

Fishing efficiency is also thought to affect catches. But, unlike population size, the meaning of "fishing efficiency" is not that clear and requires a more rigorous definition. For this example, we can think of the chance of capturing a fish as something that increases when fishing effort (E) increases. Following the long tradition in stock assessment modelling, we define

$$\phi = 1 - e^{-Eq} \tag{3}$$

and consequently regard Eq as the instantaneous fishing mortality rate and q as catchability.

Now, we go back to defining the bits and pieces of our graphical model. We assume that, before collecting the data, effort is under our control and is not an uncertain quantity. Compared to N and q, catches are different; there are arrows pointing to these variables. This means that their values are believed to depend on the values of the other variables. Conditional probability distributions are used to describe what we think about this dependency. Catching fish is a random process. Even if we knew exactly the number of fish and the efficiency of our fishing method, we would still be uncertain about how many fish we will catch. This type of uncertainty is often called "aleatory" because it arises from the random-looking variation in the process of which we are thinking. This is opposed to "epistemic" uncertainty that we have about N and q, which do not vary, but whose fixed value we do not know. In both cases, the uncertainty is personal and can be quantified using degree of belief-probability. If we believe that the fish react independently to fishing (i.e. there is no schooling behaviour or environmental batchiness), a natural conditional prior distribution (or "sampling model") for the catches would be the binomial distribution, which describes the number of successes in N independent trials, where each trial has the chance of success ϕ.

Finally, our full model specification can be written using mathematical notation

$$log(N) \sim N(5.3, 0.5)$$
$$log(q) \sim N(-2.3, 0.5)$$
$$\phi = 1 - e^{-qE}$$
$$c_1 \mid N, \phi \sim Bin(N, \phi)$$
$$c_2 \mid N, c_1, \phi \sim Bin(N - c_1, \phi)$$
$$c_3 \mid N, c_1, c_2, \phi \sim Bin(N - c_1 - c_2, \phi) \tag{4}$$

where bar "|" indicates that the variable has a conditional prior distribution that depends on the uncertain parameters listed after the bar. The model includes five random variables, which means that the full Bayesian model is a five-dimensional probability distribution that encodes our current knowledge about the problem.

At this stage, it is worth noting that we have not observed the catches. Ideally, the model would be formulated before collecting the data, so that the model can be utilized in planning the removal process. The joint distribution can be visualized and examined by using computer simulation and taking a large sample of values from the distribution and examining the pairwise correlation plots. For example, Figure 1.2 shows that our current knowledge about the combinations of N and q is that they look independent. This means that if we now had more knowledge about N, it would not change what we think about q. However, N and c_1 are correlated, and q and c_1 are also correlated. This gives a hint that observing c_1 will help us to learn about both q and N. We can also see that c_1 and c_2 are correlated. Thus, if we lost the data on c_1, we could use c_2 to infer the size of the first catch.
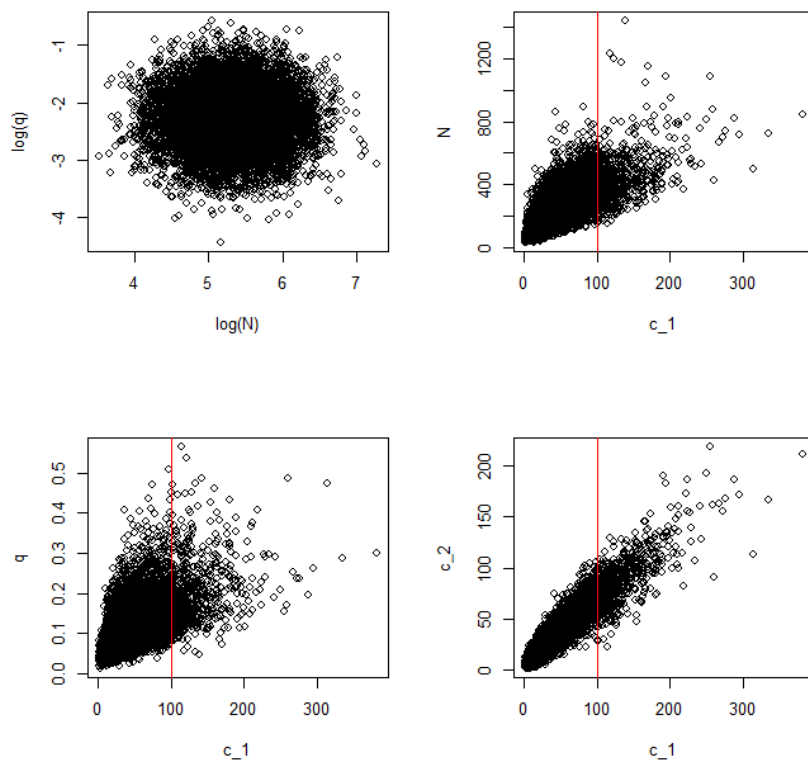
Figure 1.2. Prior correlations between some of the parameters in the removal sampling problem. Red line illustrates the location of values that would be compatible with hypothetical observation $c_1 = 100$.

Suppose then that we start collecting the data removal by removal and examine how our thinking starts to change about population size and catchability. The first catch happens to be $c_1 = 100$; what do we now know about population size? Now, the original uncertainty about the first catch has completely disappeared. In terms of the large sample of numbers that we drew from the joint prior distribution of all parameters, this means that we should only look at those combinations where $c_1 = 100$ and should remove all other values. The result is shown in Figure 1.3. The red values predicted the first catch correctly and remain within the limits of possibility. We can see that, compared to the prior knowledge, a much smaller set of combinations is now realistic. It is also noteworthy that unlike before, now N and q are correlated in our minds. High N

and small q, and high q and small N, are supported by our interpretation of the observation. The joint distribution of the combination of values that are consistent with the data is called the joint posterior distribution. Looking at the variables one by one gives the marginal posterior distribution, which describes all we know about that variable. Thus, the marginal distribution of N is of most interest in this problem. We can see that the most probable population size has moved from ca. 170 to ca. 320, while the whole posterior distribution has moved to support higher population sizes. However, the probability for a very high population size has also clearly decreased. The 80% probability interval is now 237–531; i.e. we believe that the true population is between these two values with probability 0.8.

The process of learning from observations can be intuitively described by using the simulation analogy and the idea of removing the values that are not consistent with the observations. Figure 1.3 shows how the set of parameter values gets smaller and smaller as new data are obtained, and the probability distribution describing our knowledge about N and q gets narrower. However, in addition to a simulation experiment, the way in which the probability distribution changes with new knowledge can also be presented using the well-known result of probability theory:

$$p(N, q | c_1 = 100) \propto p(c_1 = 100 | N, q) p(N, q) \tag{5}$$

where $p(N, q | c_1 = 100)$ denotes the joint posterior distribution of N and q (red dots in Figure 1.3), $p(N, q)$ is the joint prior of N and q (grey dots in Figure 1.3) and $p(c_1 = 100 | N, q)$ is the probability of observing $c_1 = 100$ for each combination of N and q. In other words, the Bayes' rule above states that the posterior distribution of unobserved variables, given the observed values, is proportional to the product of the prior distribution and the conditional probability of observing the data at hand. In this example, the conditional probability of observations was defined using our belief that fish behave independently; consequently, the binomial distribution was used. Now that the observation is fixed at $c_1 = 100$, the weight given to each combination of N and q is the probability by which this data would be observed if the combination were true. The higher the probability of data, the more realistic the (N,q) combination seems.

Using the conditional distribution of data in this way gives rise to the so-called likelihood function, which can be seen as our personal interpretation of the objective data that were observed. By first expressing our knowledge about the observation process as a conditional distribution of potential data, we create a predefined logic by which we later interpret any specific data that we happen to obtain.

The likelihood function for a set of parameters consists of probabilities, but it does not form a probability distribution for these parameters. To make this difference clear, the values or weights given to each parameter combination are termed a "likelihood". The widely used maximum likelihood estimation (MLE) seeks to find the combination with the highest likelihood, i.e. the values that make the observed data look most probable. However, this is not usually the same combination that is most probable, given the observed data. For example, after observing that $c_1 = 100$, the ML estimate is that the population size was $N = 100$ and that the catchability q is infinite so that the capture probability $\phi = 1$. Thus, it is our prior belief that catchability should be around 0.1 that keeps the posterior distribution in a sensible range.
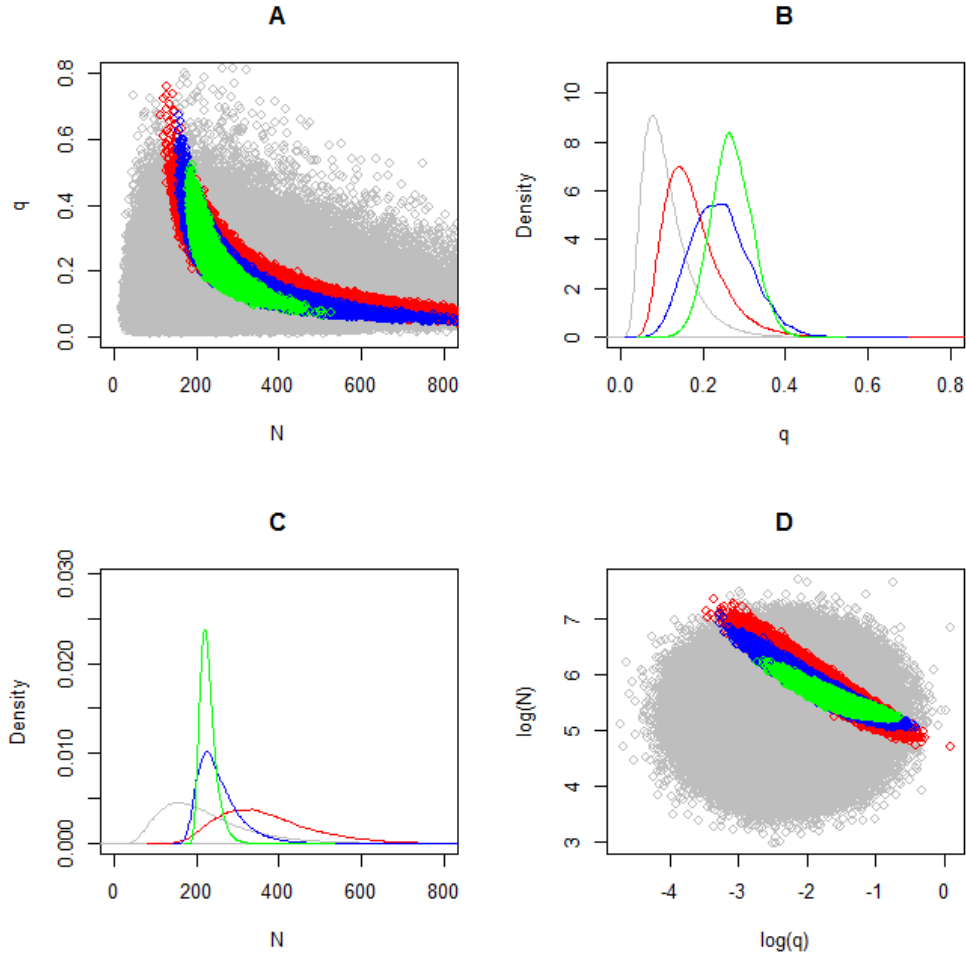
Figure 1.3. Joint (A and D) and marginal distributions (B and C) of the population size _N_ and catchability coefficient _q_ after observing different amount of data. Grey distribution shows the prior when no observations have been made. Red corresponds to $c_1 = 100$, blue shows the knowledge after knowing that $c_1 = 100$ and $c_2 = 50$, and green is based on $c_1 = 100$, $c_2 = 50$, and $c_3 = 30$.

Frequency approach

The marginal posterior of N now shows how uncertain we are about the population size after the first catch. The approach to quantifying uncertainty in classical statistics is different. Because N represents the state of nature which is assumed constant when collecting the data, it does not have a frequency probability distribution. However, it is possible to think about other estimates that could potentially realize. But this scenario requires hypothetical values to be assumed for N and q; otherwise, the frequency distribution for other potential data does not exist. It is common practice to use the ML estimates obtained from real data and use them as true values; it is then possible to calculate the distribution of potential data and also examine how the potential estimates would vary. Our example case is interesting: now that $\hat{N} = 100$ and $\hat{q} = \infty$, the expected value of potential catches is $\hat{N}[1 - exp(-\hat{q})] = 100$ and the variance is $\hat{N}[1 - exp(-\hat{q})]exp(-\hat{q}) = 0$, so we think that the data that we observed are actually the only data we could have observed and, consequently, that our potential maximum likelihood estimates would also not vary. This problematic situation arises from the idea of maximum likelihood estimation and from the practice to use that estimate as if it were the true state of nature. This situation reveals a major difference between Bayesian and classical statistical inference. Bayesian inference is measuring uncertainty

about the parameters themselves, whereas the frequency approach is bound to assess how potential new estimates would vary when some known values for the parameters have first been assumed.

When the second fishing pass was also made, $c_2 = 50$ was observed, and the question is what we now think about population size? The answer is again a posterior distribution for N, which we can obtain in the simulation example by accepting only those combinations of red dots, for which $c_2 = 50$. These values and the resulting marginal distributions are shown in blue. Alternatively, we could start from the sample from prior (grey) and choose values that predicted $c_1 = 100$ and $c_2 = 50$. In analytical terms, the Bayes' rule can be used sequentially by taking the posterior based on the first catch as the prior for the second catch

$$p(N, q | c_1 = 100, c_2 = 50) \propto p(c_2 = 50 | N, q, c_1 = 100) p(N, q | c_1 = 100) \tag{6}$$

or by starting from the original prior and analyzing the whole data set at once

$$p(N, q | c_1 = 100, c_2 = 50) \propto p(c_2 = 50 | N, q, c_1 = 100) p(c_1 = 100 | N, q) p(N, q). \tag{7}$$

The posterior distribution of the population size is now more peaked and covers a narrower range of values, indicating that we now have a clearer idea about population size. The most probable value is ca. 230 and almost all of the probability mass is concentrated between 180 and 400.

Frequency approach

The probability of obtaining these two catches would be highest if N = 195 and q = 0.36, so this pair of values is the maximum likelihood estimate. Again, probability statements about population size and catchability are only possible within the Bayesian approach. At this point, however, other potential maximum likelihood estimates would vary if we assumed that the observed estimate was the true state of nature. This variation can be examined by using parametric bootstrapping; by assuming that the true values are N = 195 and q = 0.36, we can generate a large number of potential datasets and calculate the ML estimate from each dataset. Figure 1.4 shows the results of such an analysis. If population size were really 195, the potential ML estimates would vary around that value so that the true value would be the mean of all estimates. The joint distribution of ML estimates for N and q has a banana shape that resembles the joint posterior distribution of N and q in Bayesian analysis. It is worth noting that if N = 195, the most frequently occurring ML estimate would be ca. 180 (not 195), and that ca. 60% of the ML estimates would be smaller than the assumed true population size.
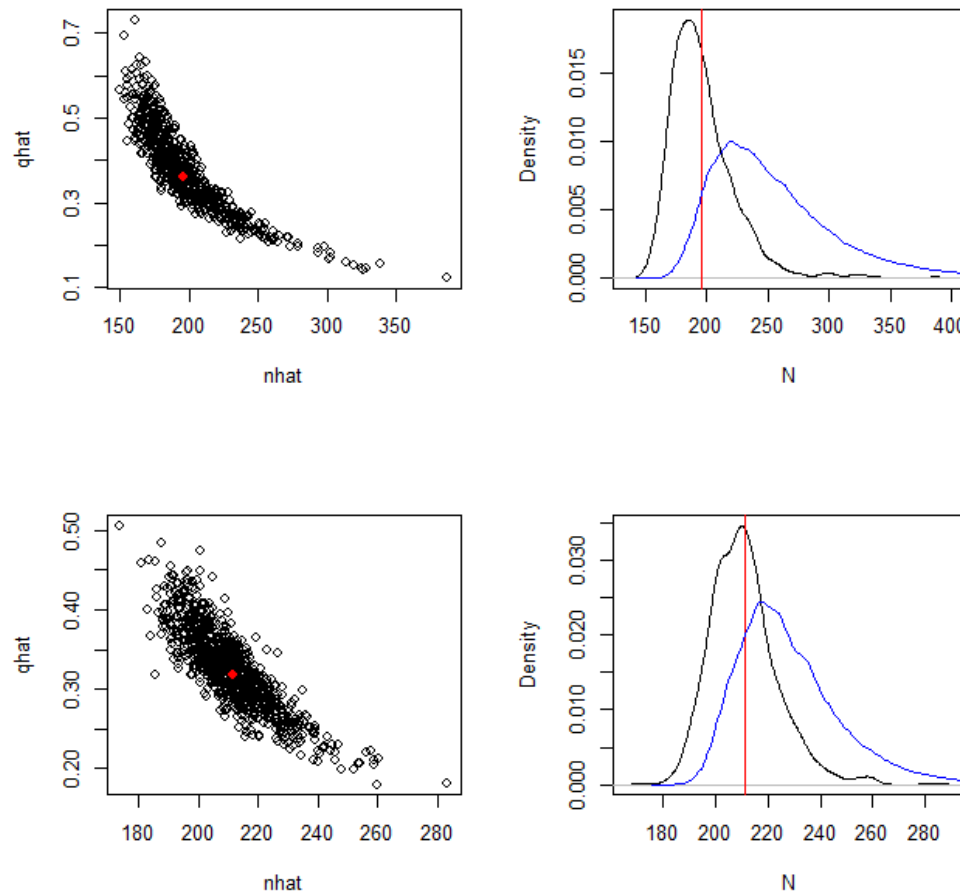
Figure 1.4. Frequency distributions of potential maximum likelihood estimates (black dots and lines) based on the observed maximum likelihood estimates (red dots and lines). Upper row includes first two catches and the lower row shows the case with all three observations. Posterior distributions of the population size (blue) are overlaid for comparison.

After making the last observation $c_3 = 30$, the marginal posterior distribution of N is concentrated to an even narrower range (190–280) than before (Figure 1.3), but the most probable population size has not changed from 220. The marginal distribution is the final result of our analysis after observing the three sequential catches.

Frequency approach

After these three observations, the maximum likelihood estimate of the population size is N = 211, q = 0.32. If these were the true values, then other potential maximum likelihood estimates would vary less in comparison to the case where only two sequential catches were hypothesized. The potential estimates would still be correlated, but less than with only two catches. With three catches and assumed true values of N = 211, q = 0.32, the most frequently occurring ML estimate for N would be 210. The frequency distributions of potential point estimates are quite often misinterpreted as if they were posterior distributions of the actual parameters based on the observed data. Figure 1.4 shows the marginal posterior distribution of N from the Bayesian analysis overlaid with the distribution of potential ML estimates obtained assuming that the observed ML estimates were the true values. It can be immediately seen that the distributions do not have the same shape and location, but it is even more important to realize that

the interpretation of the x- and y-axes are very different. For the posterior distribution of N, the x-axis shows all the alternative true population sizes, and the y-axis shows the degree of belief assigned to each of these values based on prior knowledge and three successive catches. For the frequency distribution, the x–axis does not represent the alternative true population sizes, but alternative maximum likelihood estimates that could be potentially observed if the observed maximum likelihood estimate (red line) were the true value. The y-axis shows the relative frequencies of these estimates, after repeating the fishing process for a very large number of times. In other words, the frequency distribution is not trying to assess the true population size, but considers a large sample of new point estimates.

The interpretations can be bridged in the following way. The blue line shows the probability of each true population size and, therefore, also shows how credible the observed ML estimate is compared to other alternative true population sizes. In this case, the credibility is quite high, but not the highest of all.

The joint distribution of *q* and *N* still shows some correlation, which hints that a further reduction in uncertainty about *N* would still be possible if more information about catchability could be obtained. Whether the current uncertainty is too large or not does not belong to the field of Bayesian inference; the uncertainty is what it is and does not possess any value in itself.

Whether attempts should be made to further reduce uncertainty always depends on the context. The most rigorous way to analyze the need for collecting more data is to adopt the Bayesian decision-analysis approach (Raiffa and Schlaifer, 1961) where the costs of data collection can be contrasted with expected gains of managing with and without the potential new data (McDonald and Smith, 1997; Mäntyniemi *et al.*, 2009). Such a value-of-information (VoI) concept is tightly linked to the honest use of prior information in fishery stock assessment and management. Whenever the expert group intentionally leaves some information unutilized in the Bayesian stock assessment (uses too flat prior distributions compared to actual knowledge), the value of any new information will become overestimated. Whenever the group pretends to know parameters that are not well known (e.g. by fixing natural mortality), the value of any new information will become underestimated (Mäntyniemi *et al.*, 2009).

Terminology

Belief. Knowledge that includes uncertainty. The word "belief" is used to underline the inherent subjectivity of any uncertain knowledge.

Probability distribution. Also called probability density function (pdf) when it is specified as a function. A parameter the value of which is not exactly known has a probability distribution. The probability distribution encapsulates the current belief/knowledge about what values the parameter may have and how probable each value is. Probability distribution may be continuous or discrete.

Subjectivity. Knowledge is necessarily a personal thing and is subjective by its very nature. Our knowledge and the lack thereof are inside our mind (belief) and, therefore, are subjective. Collective knowledge is also (collectively) subjective. Subjectivity is equally present both in the Bayesian and non-Bayesian approaches. Subjectivity does not exclude neutrality, honesty, or "unbiasness", but rather stresses the importance to strive for them.

Objectivity. The real world outside the human mind is objective. Human knowledge is not objective, and there is no objective thinking. Therefore, objectivity only relates to "the truth out there".

Prior (belief/knowledge/information). Also simply called "prior". Existing relevant knowledge available before the analysis of the newest observations (data) and the information included in them. Prior knowledge is not fed into a statistical analysis in a form of observations, but in a form of probability distribution, which summarizes what is believed about values of the parameter of concern. Apart from specifying probable parameter values, prior knowledge is also specifying processes, i.e. defining how data are related to the studied phenomenon (model structure and sampling model).

Posterior (belief/knowledge/information). Formal synthesis of prior knowledge and new observations (data) by using Bayes' rule results in probability distribution, which is the posterior knowledge of the parameter of concern.

Updating. The process of applying Bayes' rule to combine prior knowledge and data.

Bayes' rule/Bayes' theorem. $P(h|e) \propto P(e|h)P(h)$. See page 6 for more details.

Uncertainty about uncertainty

The chance parameter ϕ is interesting because it essentially represents a certain kind of probability. However, it is not a Bayesian degree of belief, but rather a parameter that describes the behaviour of the real-world system. While Bayesian probability measures what we know about the system, this chance parameter represents the randomness in the system. If we imagine trying to catch an infinite number of fish, ϕ would represent the proportion of successes in such a situation. On the other hand, ϕ does not really exist in the same way as the true population size N, but exists in our minds as a property of the system. In other words, we can never observe ϕ directly; we can only observe counts of fish. This implies that we will always be uncertain about the (imaginary) true value of ϕ, which breaks down to uncertainty about catchability (q), effort (E), or both.

Now, the Bayesian approach is to express uncertainty about these parameters using a prior distribution. This is an important case because we are using the degree of belief – probability to express what we know about a "physical" probability. This is where the Bayesian inference encapsulates the frequency probability and allows us to measure the uncertainty about it. This is not possible in classical statistics, which uses only the concept of frequency probability. On the other hand, the Bayesian approach assumes that everyone knows their own degree of belief and does not possess the concept of a person's degree of belief about his or her own degree of belief. Yet another concept of probability would be needed for assessing such uncertainty.

1.4 The aim of the report

This report is a manual representing a review of the methods to be applied when providing prior information to Bayesian stock assessments and marine risk analysis. It is compiled by the ECOKNOWS project, where a critical summary of the existing methods to formulate prior information and further development of the methods is one of the essential deliverables to be provided. The manual also facilitates a better communication of scientific information expressed in the form of probabilities.

Due to the application of an ecosystem approach to fishery management, there is an increasing need to expand risk-related advice to new species. Often, little or no traditional stock assessment data exist about these species; hence, effective utilization of various background data and other sources of information is essential. Our report is a comprehensive handbook guiding the use of this type of information in assessment and advice, which is relevant both in data-rich target fisheries and especially in data-poor cases related to, for example, bycatch species.

As already pointed out, much research is built on more or less conscious (subjective) reasoning leading to choices or preferences regarding scientific questions to be studied, data, analyses, and conclusions. This also holds in the guidelines we present here, which are our "posterior beliefs" based on our experience thus far and which we are offering as "priors" for the readers. Thus, in order to be consistent with our message, we want to stress that updates to our guidelines are expected in the future[1]. We also welcome any feedback on how well stated and transparent our reasoning is, or how consistent it is with readers' own experiences.

---

[1] The updates may arise from our own or from the readers' future experiences and they may either confirm or refute our current understanding.

2   The nature of different information sources

Jonathan White, Guillaume Bal, Niall Ó Maoiléidigh, Konstantinos Stergiou, Samu Mäntyniemi, Atso Romakkaniemi, Rebecca Whitlock, Rainer Froese, Vaishav Soni, Polina Levontin, Adrian Leach, and John Mumford

2.1   Data

Jonathan White, Guillaume Bal, and Niall Ó Maoiléidigh

Setting an informative prior from data should be considered if only a small amount of data describing the target variable exist, or if the data in question are not believed to be representative of the true variable. This may be the case if sampling bias is suspected, e.g. where data are believed to give a sample estimate of the distribution rather than the true population distribution. For example, the size of fish in a commercial catch may be subject to sampling bias either from restrictions on the fishing net mesh size, hook size, or an implemented minimum take size. If the size of fish in the entire population is the target of the variable being modeled, then an informative prior of the true range of fish size in the population would be desirable. This could be determined from scientific sampling, samples taken over a longer time-series where the time-series incorporates individuals from the full population, or from the literature.

The aim is to develop a frequency distribution reflecting the true range of the variable or parameter of interest, with the weight of the distribution function concentrated around its midpoint. To this end, if the data to be applied in a model are believed to exhibit some sampling bias away from the true population (such as measurements of fish size from catches with fishing restrictions), then they should not be used in constructing the prior. Reliance on the same data would give rise to a higher degree of certainty in the posterior than should be accepted.

In establishing an informative prior distribution for a variable or parameter from other ancillary data, several steps should be followed. The expected frequency distribution needs to be chosen that will determine the parameters needed to define the distribution within the mathematical syntax of the chosen software. For example, a normal distribution is determined by values describing the mean and standard deviation; a binomial distribution is determined by the probability of success and the number of trials; and a negative binomial distribution is determined by the probability of success and the number of successes. The choice of the frequency distribution determines the (i) shape of the distribution, (ii) range of values around the midpoint (typically chosen as the mean, median, or mode of the distribution) and also their symmetry or asymmetry, and (iii) density of values (their probability) relative to distance from the midpoint.

2.1.1   Choice of prior probability distribution

The aim of setting an informative prior distribution from data is to describe the situation – the variable of interest – as well as possible. Two main factors will determine this choice:

- prior understanding of the frequency distribution from other examples or knowledge and;
- the size of the dataset being used to set the prior, and its apparent frequency distribution.

Prior knowledge of the variable in question and scientific knowledge of general types of data are important in *a priori* choosing the density function for an informative prior. The sample size of the data, however, will have the greatest influence on the choice.

2.1.2   Influence of past studies and scientific understanding

General knowledge of probability distributions, datasets of a similar nature, and experience with sets of the same type of data from different sources are likely to influence the choice of the prior distribution function. Examples of generally expected probability distributions include:

- Weight and length measurements of individuals in an age class tend to be normally or log-normally distributed.

- Numbers of males to females in a population of known total size tend to follow a binomial distribution.

- Organisms in the environment tend to be clumped or contagious in their spatial distribution; therefore, probability distributions of their counts from spatial sampling techniques (such as sweep net or quadrate surveys) tend to be strongly right skewed, following negative binomial or log-normal distributions.

- Organism distributions are sometimes regularly distributed within contagious clusters and, at this smaller scale, may display a uniform frequency distribution.

- Certain events occur with a regular temporal distribution, such as temperature relative to the time of day or year, or the eruptions of a geyser. In such cases, this should be reflected by the chosen probability model.

- Some weather events have a contagious temporal distribution, such as the frequency and strength of tornadoes and winds, for which the Weibull distribution has often been used to model probabilities.

Previously collected datasets of the same variable from a different source can also give valuable insight into the expected frequency distribution form. For example, length measurements of a population of a species, different from the population of interest, may be informative. While care needs to be taken doing this, it can be a useful approach. An "information donor" population should be chosen that is believed to be close to the population of interest. In this sense, the term "close" applies not only to the spatial and temporal localities of "donor" and "recipient" population estimates, but also to the value of the variable under scrutiny.

Furthermore, the general form of a frequency distribution is more robust than the actual range of its elements (or measurements). For example, while a population of the wood mouse (*Apodemus sylvaticus*) from a productive habitat may have a mean weight of 26 g and a population from a less productive habitat may be significantly smaller with a mean weight of 20 g, the frequency distribution (probability function) of weights of the two populations may be expected to exhibit the same shapes and forms. This will hold true for many measurement types and their classes and justifies the application of a probability distribution type based upon experience.

2.1.3   Sample size

The way in which a dataset is used to construct a prior needs consideration. Sample size is most important, as this will influence how the frequency descriptors of the data

are implemented, either using the full dataset directly or alternatively calculating or estimating descriptors of the dataset.

Large sample sets clearly are the most reliable for setting informative priors. Frequency distributions of dataset values can be used to fit expected probability distribution functions; then, goodness-of-fit tests are performed to choose the most appropriate form and its values to define the prior. This can lead to strongly informative priors that can be appropriate when the variable in question is very closely associated with the data used to create the prior. If, however, the data used in setting the prior represent a different population, time, or location, care should to taken to ensure they are not overly influential, as they could limit development of the posterior by being too strong and thus limit the model and its data in influencing the posterior estimate.

For smaller datasets still large enough to exhibit a discernable probability distribution type or shape, descriptors can be calculated and the frequency distribution can be examined and compared against known, expected frequency distributions, and the mean, median, mode, standard deviation, variance, etc. calculated directly.

For small datasets, comprising anything below ca. 50 samples, choosing an appropriate probability distribution becomes difficult. For small datasets, the shape of the distribution function is not always apparent. Figure 2.1 shows frequency plots of values randomly drawn from a log-normal distribution (mean = 5; s.d. = 0.447) and an increasing number of drawn samples (n = 10, 50, 100, 1000, and 10 000). For the first two series of draws (n = 10 and 50) and, to an extent, the third (n = 100), the shape of the distribution is not clear, and no distinct pattern could be reliably proposed. In such cases, the basic details from the data (mean, median, standard deviation) can be calculated and applied to a prior. However, the choice of the distribution type would be left to expert opinion or prior knowledge (see above).

If small sample sizes are to be used, misspecification of the prior could result in two ways: (i) inaccuracy in the direct estimate of the distribution descriptors (mean, median, standard deviation) and (ii) from application in a model of an incorrect probability distribution type. In these cases, a balance must be reached between information arising from the data and that from expert opinion. In such instances, it is important to ensure that the informative prior incorporates the expected variability in the variable and does not force estimates to a very narrow and potentially falsely precise value.
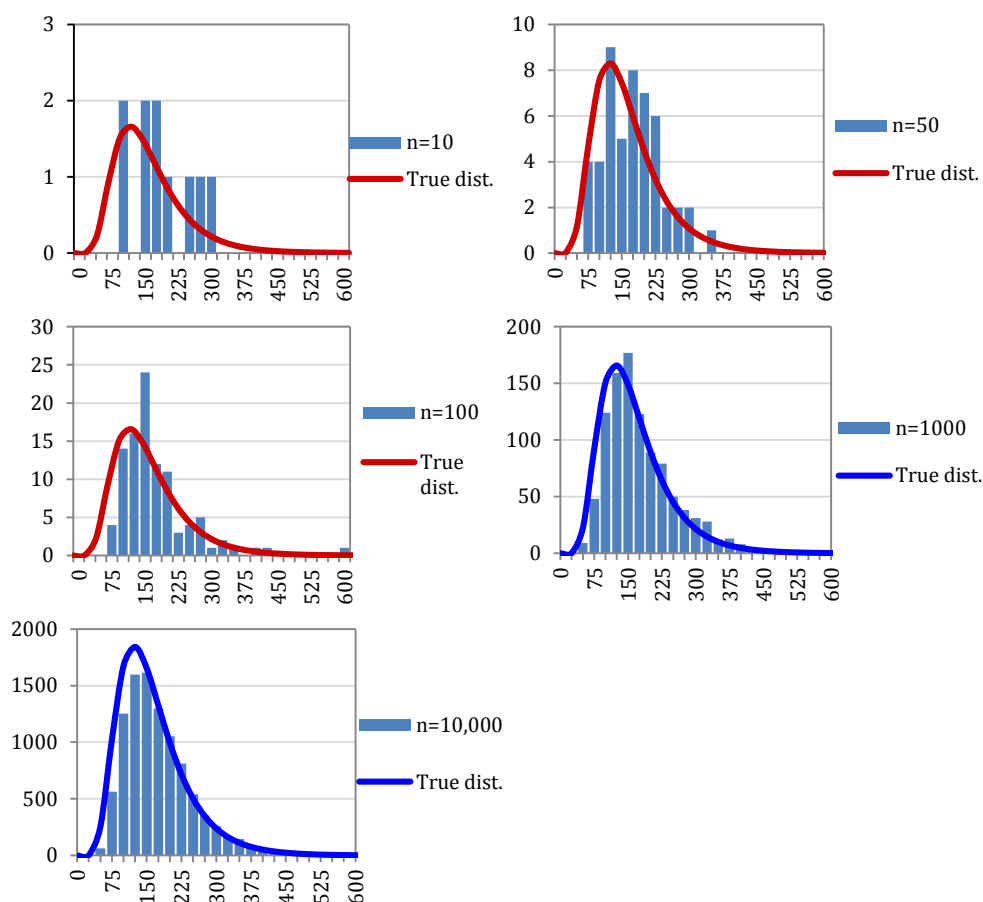
Figure 2.1. Frequency distributions of n = 10, 50, 100, 1000, and 10 000 randomly drawn samples from a log-normal distribution with a predetermined mean of 5 and standard deviation of 0.447. The sample means and standard deviations given in Table 2.1 are included for comparison (note that for n = 10, 50, and 100, the expected distribution plots in red emphasize their difference to the observed sample frequencies).

Table 2.1. Sample size and associated means and standard deviations from randomly drawn samples of a log-normal distribution with a predetermined mean of 5 and standard deviation of 0.447. Frequency distributions are plotted in Figure 2.1.

Distribution	n	10	50	100	1000	10 000	True values	
Log scale	Mean	128.03	150.29	161.83	163.05	163.62	164.02	
	s.d.	49.285	73.334	67.344	76.687	78.026	77.194	
Transformed	Mean	4.765	4.904	5.001	4.993	4.994	5.000	Used to establish
	s.d.	0.4747	0.4674	0.4201	0.4492	0.4551	0.4472	distributions

2.1.4   Priors for unobserved values and parameters.

Unobserved model components can fall into two categories:

- Variables that are representative of a true value or count. For instance, the number of fish or their size at a specific life stage in a specific location that are not observed. These variables are true values even if not measured (i.e. not observed).

- Parameters or notional (conceptual) values that are used in mathematical modelling, but are not measurable (e.g. parameters within stock–recruitment models).

For unobserved values, informative priors can be implemented from alternative datasets if samples of the population of interest are not available. These may be scientific samples or fishing samples. Priors may also be derived from combinations of datasets if there is an expected reliable relationship among the different datasets. The criteria described above for using data in the derivation of priors apply here as well.

For parameters, the situation is similar. While the target value and its probability distribution may not be observed directly from a dataset, the combination of datasets through a relationship or model can indicate the expected probability distribution form and range [e.g. a parameter in a growth relationship, such as the growth rate (r) and carrying capacity (k) of the Beverton–Holt (1957) growth model].

If prior estimates of a value are going to be derived from two or more datasets, the conditions detailed above, namely the sample size of the dataset and the expected probability distribution form, need to be considered for each dataset. The methods presented in Section 3.2 include an appropriate toolbox for deriving priors from several datasets.

2.2 Literature

Konstantinos Stergiou, Samu Mäntyniemi, Atso Romakkaniemi, and Rebecca Whitlock

2.2.1 Introduction

Literature is an important source of information for various parameters related to the assessment of fish stocks. For instance, in a review of length-at-first-maturity of fish in the Mediterranean Sea, a region traditionally considered data-poor, information was found for 565 marine fish stocks representing 150 species (Tsikliras and Stergiou, 2014). This is especially true of the so-called grey literature (i.e. theses, proceedings, local journals, journals that publish in languages other than English) which was not widely available before the Internet era and, until recently, was generally not included in various online bibliographic databases (e.g. Scopus, Web of Science). For instance, >60% of the articles cited in four reviews on various biological aspects of Mediterranean marine and freshwater fish were in local, grey literature (Stergiou and Tsikliras, 2006). Despite the existence of large databases that accommodate the available literature and from which relevant information can be extracted (Section 2.3), the percentage of the published primary and grey literature that is not incorporated into such databases can be high (e.g. Stergiou and Moutopoulos, 2001; Stergiou and Karpouzi, 2002; Apostolidis and Stergiou, 2008; Tsikliras and Stergiou, 2014). For instance, within the ECOKNOWS project (www.ecoknows.eu), for the five Mediterranean case study species and the nine parameters examined, the percentage of records derived from both primary and grey literature not included in FishBase (Froese and Pauly, 2014, www.fishbase.org, version 8/2011) ranged from 0 to 100%, depending on the species and parameter (Tables 2.2 and 2.3) (Stergiou *et al.,* 2012).

Table 2.2. Information already existing in FishBase (FB; www.fishbase.org; Froese and Pauly, 2014) and new information to be added to FishBase in the framework of the ECOKNOWS project (ECO) for the five Mediterranean case study species. Underlined values indicate cases where the new information collected exceeds in number the records already existing in FishBase. VBGP = von Bertalanffy growth parameters, W–L = parameters of the weight–length relationship, L-freq = length frequency distributions, L_m = length at first maturity, diet = quantitative description of feeding habits (expressed as mass and/or volume of prey items), predator = number of recorded cases of predation of the species by other organisms (data from Stergiou *et al.*, 2012).

Parameter		*D. annularis*		*M. merluccius*		*M. surmuletus*		*S. porcus*		*S. cabrilla*	
		ECO	FB	ECO	FB	ECO	FB	ECO	FB	ECO	FB
VBGP	K	5	15	23	66	10	27	1	20	4	5
	L_∞	5	15	23	66	10	27	1	20	4	5
	t_0	5	11	15	27	8	18	1	13	4	5
W–L	a	16	54	25	61	13	51	8	32	7	25
	b	18	54	25	61	21	51	8	32	7	25
L-freq		<u>42</u>	2	<u>30</u>	7	<u>35</u>	2	<u>10</u>	2	<u>14</u>	2
Maturity	L_m	5	6	2	9	3	10	-	4	-	2
Fecundity		<u>1</u>	-	5	7	-	-	<u>1</u>	-	-	-
Spawning season		10	10	12	14	<u>17</u>	10	<u>8</u>	1	<u>10</u>	5
Diet		<u>24</u>	3	<u>39</u>	23	8	15	<u>11</u>	10	1	3
Predator		<u>6</u>	3	17	17	7	8	<u>3</u>	2	<u>1</u>	-

Table 2.3. Percentage of records that are not being included in FishBase (FB; www.fishbase.org; Froese and Pauly, 2014), based on data collected in the ECOKNOWS project for the five Mediterranean case study species, as presented in Stergiou *et al.* (2012) (Table 2.2). For abbreviations, see Table 2.2.

Parameter		*D. annularis*	*M. merluccius*	*M. surmuletus*	*S. porcus*	*S. cabrilla*
VBGP	K	25	26	27	5	44
	L_∞	25	26	27	5	44
	t_0	31	36	31	7	44
W–L	a	23	29	20	20	22
	b	25	29	29	20	22
L-freq		96	81	95	83	88
Maturity	L_m	46	18	23	0	0
Fecundity		100	42		100	
Spawning season		50	46	63	89	67
Diet		89	63	35	52	25
Predator		67	50	47	60	100

2.2.2 What information to collect

Compiling information from the literature for the purposes of prior formulation is a rather straightforward process when the parameter of interest is available in a common form (i.e. with the same interpretation and units) across publications (as an example of the opposite situation, many parameterizations of the commonly-used Beverton–Holt (Beverton and Holt, 1957) and Ricker (1954) stock–recruitment functions can be found in the literature, so substantial care must be taken when compiling this information). In this case, one can use the published parameter estimates straightforwardly in a

Bayesian meta-analysis that combines the information across studies (Section 3.2). Each point estimate can be treated as a proxy for the mode of a hypothetical posterior distribution for the estimate (i.e. the posterior distribution that would have been obtained if the analysis were a Bayesian one), while the standard deviation of the estimate may be treated as the posterior standard deviation. However, care should be taken to understand what the reported values are and if they differ between studies, so that they can be used appropriately. For instance, if the published standard deviation were obtained using classical statistical methods, it would not be equivalent to a standard deviation from a Bayesian analysis (according to the theory, uncertainty will tend to be underestimated). Reported posterior distributions from appropriate Bayesian analyses can be directly used in meta-analyses.

In addition to the results of data analyses, the literature also contains auxiliary information about the subject and methods of a particular study. Description of the study area, study design (data collection and analysis), and other aspects of the study may be relevant when considering whether and how to use published results to formulate a prior. A careful review of the auxiliary information becomes more important when published results available do not share a common definition across studies and as the complexity of the models used to obtain those results increases. In such cases, auxiliary information may be needed to transform published parameter values in order to reach a common interpretation and scale. If the published results were obtained using different modelling approaches, the validity (e.g. underlying assumptions) of those approaches may merit consideration, but in general, use of all results obtained using appropriate methods is recommended to fully account for structural (i.e. model) uncertainty. Auxiliary information may also play a central role when the parameter of interest shares certain features across studies, but covariates are needed to explain some of the variation across studies (partial exchangeability, see Section 3.2.3) or if correlation between parameters is considered (Section 3.2.4).

2.2.3 Publication bias

Ideally, the scientific literature would include all knowledge accumulated by the scientific community. However, it is well known that research and publication practices are biased towards showing only "positive" results. In other words, research efforts that yield "negative" results (i.e. results that are not statistically significant) are very rarely published. As Browman (1999) states: "The issue of negative results remains complex. It reflects our training, our thoughtfulness about what we do as scientists (and how we do it), and our humanity, with all its inherent biases." (see also Gould, 1993).

Use of statistical-significance testing as the criterion for publication leads to over-representation of studies in which the null hypothesis was rejected. This is predicted to happen by chance alone in a proportion of all studies even if their null hypotheses were all true. The proportion is mainly determined by an arbitrary threshold (α, the probability of a type-I error or false positive) which is chosen to act as a cut-off for statistically significant vs. non-significant results. The smaller the threshold, the smaller the proportion of studies with rejected null hypothesis that will be published. Typically, the null hypothesis of the study states that the value of a particular model parameter equals zero, and this is taken to represent "no effect". In many cases, these are so-called "silly null hypotheses" where it is already known from other sources of knowledge that the null hypothesis cannot be true (Johnson, 1999). The true parameter value may be so close to zero that, from any practical point of view, there is no effect. However, hypothesis testing considers an exact hypothesis (e.g. equality to 0), so that the p-value

is a direct function of the amount of data collected. If the true parameter value deviates even slightly from the value under the null hypothesis, the *p*-value can be made small enough for publication by collecting a sufficiently large amount of data. Thus, studies with silly nulls and large datasets will get published more easily, even though these are not desirable criteria for publication.

In order to make a point of the problem, the above description is somewhat exaggerated and can be seen as the worst-case scenario; there are other mechanisms that dilute publication bias. Whether the bias has severe consequences for the formulation of priors for Bayesian stock assessment can be difficult to assess and will require a certain degree of expert judgment. It clearly diminishes the amount of published information available for Bayesian meta-analysis, where one would ideally use the results from all relevant studies, whether statistically significant or not.

Statistical hypothesis testing usually considers differences between groups or relationships between two or more variables. The publication bias probably has less impact on auxiliary information, which has not been directly subject to selection through testing procedures. An example might be a case where length–weight data have been collected in order to estimate weight based on length in a larger sample, and the estimated weight for the larger sample has been compared to other populations or used in further regression analyses using hypothesis tests. In this example, the length–weight relationship might be affected by the publication bias to a lesser extent. A reduced effect of the publication bias can also be expected in the grey literature, which falls short of the standards required for publication in peer-reviewed international journals (e.g. due to lack of statistical significance or lack of a wider relevance for the scientific community).

2.3 Online databases

Rainer Froese, Konstantinos Stergiou, and Vaishav Soni

2.3.1 The nature of databases in the context of this work

For the purpose of this subsection, databases are understood to be information systems containing information on fish or fisheries in well-defined database fields that are organized in interlinked tables. We are referring here mainly to information other than primary data (typically, they are estimates derived from the primary data, i.e. results collected from literature, etc.); therefore, it would be more justified to name these information systems as "knowledge bases". However, the use of the word "database" is so common in this context that we use it.

The free online encyclopedia Wikipedia is a wonderful information system, but it is basically a collection of text files that cannot be used to extract standardized information to create meaningful graphs about e.g. catches in the North Atlantic in 1970. An example of a useful database is the FAO Capture database, which contains global catches by species, area, country, and year from 1950 onward (www.fao.org/fishery/statistics/software/fishstatj/en). Another useful database with time-series of catches, biomass, and recruitment is the RAM Legacy Stock Assessment Database maintained at Dalhousie University in Canada (http://ramlegacy.marinebiodiversity.ca/). Two useful databases produced by ICES are the stock assessment database (http://www.ices.dk/marine-data/tools/Pages/stock-assessment-graphs.aspx), which contains time-series of biomass, recruitment, and fishing mortality as well as reference points for the fully assessed ICES stocks, and DATRAS (http://www.ices.dk/marine-data/data-portals/Pages/DATRAS.aspx), which contains, for about 200 species, standardized catch data from scientific trawl surveys

conducted in the Northeast Atlantic. A global database of relevance to fisheries research is FishBase (www.fishbase.org) which has compiled key parameters from the scientific literature on life history traits, such as length–weight, growth, maturity, fecundity, and mortality. SealifeBase (www.sealifebase.org) is a similar database with information on invertebrates and marine mammals.

2.3.2 What makes a database a suitable source of information

A widely used database type is the relational database model, where all information is stored in interlinked tables. For such a database to be useful for scientific work, the content of data fields has to be clearly described (units, limits, methods used), and the quality of the data has to be indicated whenever possible. Such "metadata" are needed for the database as a whole, but also the source of every piece of information has to be indicated, such that it can be traced back to its origin. For scientific data, this typically requires a link to a publication.

Whenever there are established standards for information, such as ISO standards, these shall be used to enable interlinking of information with other databases that use the same standards. Important examples of standards are scientific names of species, which should follow FishBase (www.fishbase.org) for fishes and SeaLifeBase (www.sealifebase.org) for other aquatic organisms, as both databases are contributors to the Catalogue of Life authority file for scientific names (www.catalogueoflife.org). Another example is the ISO standard for official names of countries and territories. For spatial data, geographic coordinates are the standard of choice.

Typical data types are numbers (e.g. 123, 0.123), Boolean fields (e.g., TRUE, FALSE or YES, NO), attributes selected from a limited list (e.g. "low", "medium", "high"), restricted text fields (e.g. "*Gadus*" or "myfile.xls"), and free text fields, such as for comments of any kind. The strength of the relational database model lies in the ease of creating queries that connect previously unconnected information. One famous example was the connection of annual catches from the FAO global catch database with the trophic level of the respective species from FishBase. This information could be connected because both systems followed the same standard for scientific names of the respective species. The resulting analysis (Pauly *et al.*, 1998) concluded that fisheries were impacting global foodwebs.

One often overlooked aspect of databases is their level of completeness, i.e. whether all relevant information (e.g. all species, areas, years, studies) are included. If instead the database contains a biased subsample of the true distribution of the respective data, then any analysis of this database will have biased results.

2.3.3 Biological information from FishBase

FishBase is a global online database with key information on all 32 000 known fish species. Of special relevance to fisheries research are the compilations of thousands of published studies on fish biology and population dynamics. The respective estimates for parameters such as length–weight, growth, mortality, length- and age-at-maturity, fecundity, and diet composition were extracted by specialists, standardized if needed, and quality-controlled as to their probability when compared with known limits and with other studies for similar species. Doubtful parameter estimates are marked as such, with a short justification. Every number can be traced to the respective publication, to the specialist who entered it into the database, and to whom may have marked it as questionable.

An overview of the topics covered by FishBase can be gained from Figure 2.2.

Information by Topic

Trophic ecology	Life history	Uses
○ Diet	○ Growth	○ Aquaculture
○ Food items	○ L-W relationship	○ Aquaculture profiles
○ Food consumption	○ Length frequencies	○ Introductions
○ Ration	○ Recruitment	○ Diseases
○ Predators	○ Reproduction	○ Ciguatera
Physiology/Behavior	○ Maturity	○ Processing
○ Metabolism	○ Spawning	○ Ecotoxicology
○ Gill area	○ Fecundity	○ Genetics
○ Brains	○ Eggs	○ Allele frequencies
○ Vision	○ Egg dev.	○ Heritability
○ Fish sounds	○ Larvae	○ Otoliths
○ Swim. speed	○ Larval dynamics	○ Mass conversion
	○ Abundance	

Figure 2.2. Topics covered by the FishBase information system. A click on one of the radio buttons results in a list of all species for which the respective information is available.

FishBase is available free on the Internet. A copy of the complete database in MySQL format can be obtained for a processing fee of USD 50. Users can then construct their own specific SQL queries against the database. Figure 2.3 shows the key tables relating to nomenclature and how they are interlinked.
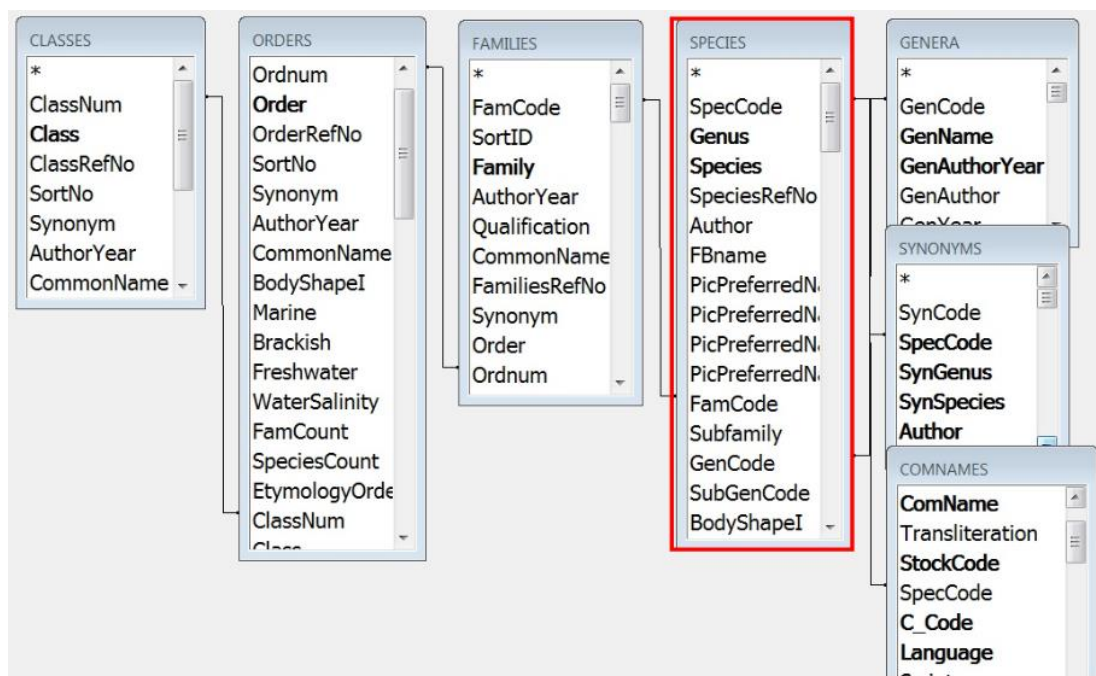
Figure 2.3. FishBase tables dealing with nomenclature, i.e. currently valid scientific names and their classification, as well as common names by language and country. The lines connect the fields on which the tables are linked.

Figure 2.4 shows how fisheries-relevant information, such as trophic ecology and population dynamics, are structured within FishBase. Note that information can be specific to a population, stock, or cultured strain so that all information that does not apply to the species in general is linked through the STOCKS table.
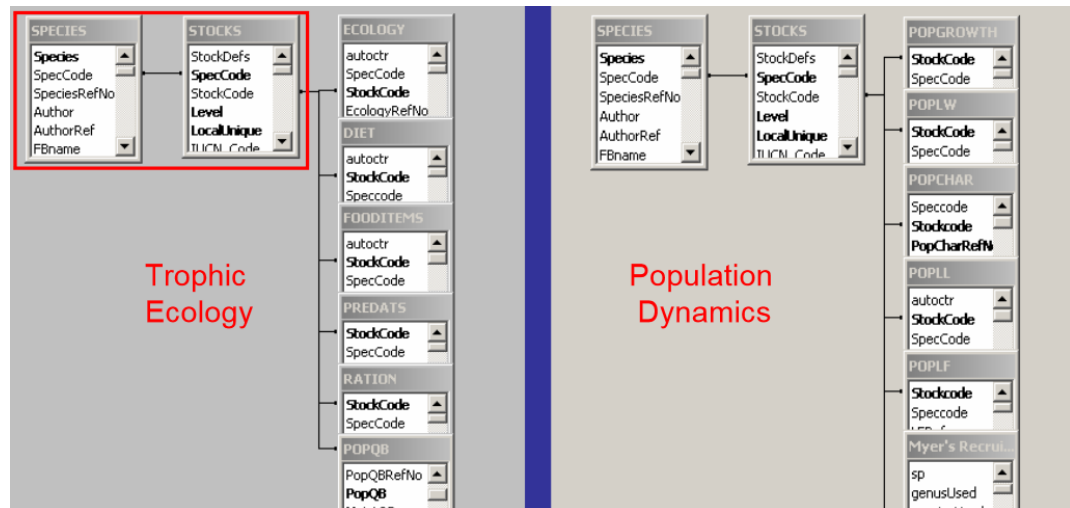
Figure 2.4. FishBase data structure for information on trophic ecology and for population dynamics.

Figure 2.5 shows an example, e.g. the FAO Catch and Aquaculture production databases, of how an external database is linked to FishBase.
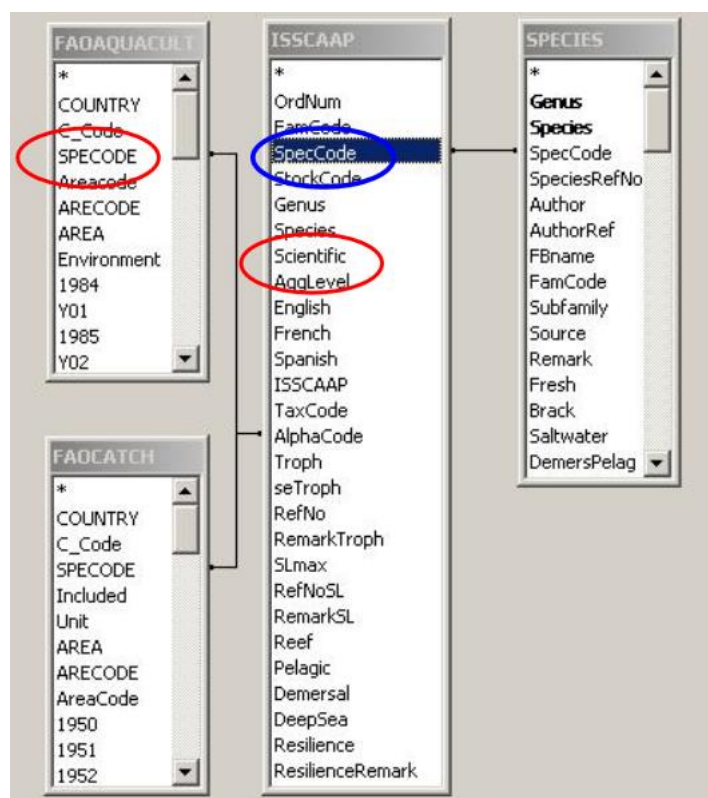
Figure 2.5. FishBase data structure for linking external databases, here the FAO databases for global catches and for aquaculture production, through the ISSCAAP table which contains scientific names as well as the fields used by FAO to identify species items (Alphacode and SPECODE).

Properly attributing scientific information to a source is not a trivial matter because several different scientific names may have been used for the same species, and many references treat more than one species. Figure 2.6 shows the respective data structure in FishBase. The REFRENS table contains all of the relevant information about a publication. The BIBLIO table establishes the link between a species and a reference and gives the page number from which the information has been retrieved. The SYNONYMS table shows the scientific name that has been used for the species in the

publication, which is then linked to the currently valid name and the SPECIES table that holds the general account for that species and links to all other information through the STOCKS table, as shown in Figure 2.4.
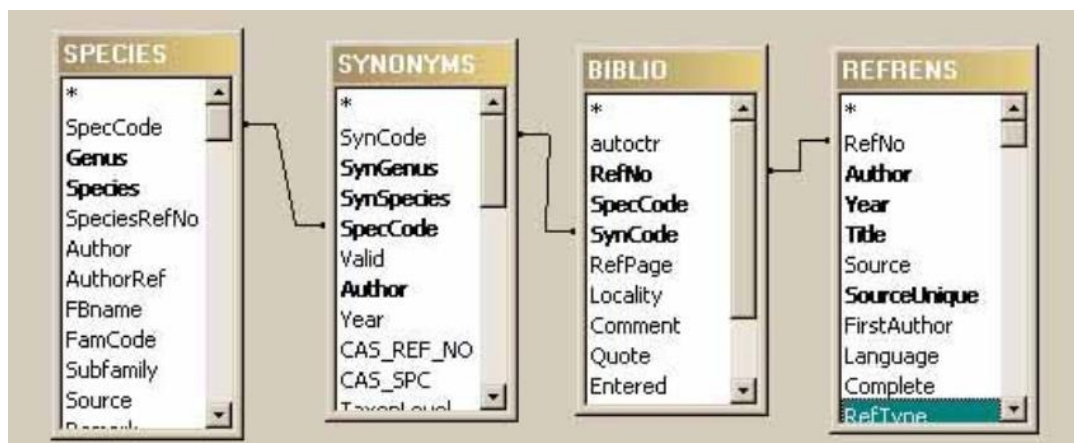
Figure 2.6. Data structure for proper linking of publications and species. The BIBLIO table gives the page number from which the information has been extracted, and the SYNONYMS table links the scientific name used in the publication to the currently valid name of the species.

As the above examples make clear, writing a query against a complex database requires good understanding of not only the data structure, but also the content of the fields and possible related fields with quality information. Incomplete understanding is likely to give misleading results. The FishBase team, therefore, prefers to be involved in scientific studies based on the database, e.g. by early involvement in prospective projects. Alternatively, especially when dealing with only a few species, users are encouraged to glean the required information from the FishBase portal where the respective data presentations have been thoroughly tested and are accompanied by relevant comments. Still, if users insist on doing the data mining by themselves, the whole database is available upon request, as indicated above. Full descriptions of all fields and tables are available in the FishBase manual at www.fishbase.org/manual/english/contents.htm.

FishBase comes with a variety of tools of relevance to fisheries research. For example, researchers can plot their own parameter estimate of length–weight, maturity, or growth parameters against the background of all other estimates for the species and family, thus obtaining an idea of how their new estimate compares with existing estimates.

Following a rule-based process, information on growth, maturity, maximum age, and fecundity has been translated in species-specific estimates of resilience in the categories high, medium, low, and very low. These categories have been used to assign priors to estimates of the intrinsic rate of population increase (Martell and Froese, 2013), as shown in Table 2.4.

Table 2.4. Default values used for batch processing of stocks based on resilience assignments in FishBase, where *r* is the maximum intrinsic rate of population increase.

Resilience	High	Medium	Low	Very low
r(year^{-1})	0.6–1.5	0.2–1	0.05–0.5	0.015 –0.1

Recently, a full hierarchical Bayesian analysis has been implemented in FishBase for the simple problem of estimating weight from length (Froese *et al.*, 2014). Whenever new length–weight relationships are added to FishBase, the Bayesian estimates of many more other species are updated automatically. Figure 2.7 shows the section from the FishBase species summary page that contains this and other information based on models.

Estimates of some properties based on empirical models

Phylogenetic diversity index (Ref. 82805): $PD_{50} = 0.5625$ [Uniqueness, from $0.5 =$ low to $2.0 =$ high].

Bayesian length-weight: a=0.00708 (0.00605 - 0.00829), b=3.03 (2.98 - 3.08), based on LWR estimates for this species (Ref. 93245).

Trophic Level (Ref. 69278): 3.2 ±0.4 se; Based on diet studies.

Resilience (Ref. 69278): Medium, minimum population doubling time 1.4 - 4.4 years (rm=0.5-1.2; K=0.2-0.6; tm=2-5; tmax=25; Fec=17,300-301,000).

Vulnerability (Ref. 59153): Moderate vulnerability (39 of 100) .

Price category (Ref. 80766): Low.

Figure 2.7. A screenshot example of information in FishBase not extracted from the published literature, but summarized by appropriated models based on the available data.

Users can also run their own Bayesian analysis of length–weight studies by selecting and weighing the studies they want to incorporate for their purposes.

2.3.4   Biological information from SeaLifeBase

SeaLifeBase is a FishBase-like information system for all aquatic living organisms (marine and freshwater) other than fish, with information for over 100 000 species of invertebrates, marine mammals (complete), marine reptiles (complete), seabirds (complete), and algae. Because SeaLifeBase started only a few years ago and because of the large number of species covered, SeaLifeBase is less complete than FishBase. But for commercially important species such as Norway lobster (*Nephrops norvegicus*), it contains numerous studies on maximum size, growth, length–weight, and maturity (Figure 2.8). Similar pages can be found for other commercially important invertebrates such as brown shrimp (*Crangon crangon*) or blue mussel (*Mytilus edulis*).

Nephrops norvegicus (Linnaeus, 1758)
Norway lobster

Figure 2.8. Screenshot of parts of the species summary page for Norway lobster in the online database SeaLifeBase. Note that the "More information" section is reasonably complete with regard to key life history traits. Behind the links are tables with the respective parameters extracted from numerous studies.

2.3.5 Description of the ICES DATRAS database

DATRAS (the Database of Trawl Surveys) is an online database of trawl surveys that is publicly available from http://datras.ices.dk and resides at ICES headquarters in Copenhagen, Denmark. The survey data cover the Baltic Sea, Skagerrak, Kattegat, North Sea, English Channel, Celtic Sea, Irish Sea, Bay of Biscay, and the eastern Atlantic from the Shetlands to Gibraltar. At present, there are more than 48 years of time-series data for about 200 species in DATRAS. The surveys are primarily designed to provide information on abundance of commercially important species. However, information on all other fish species taken in survey catches are collected on a routine basis by national institutions and uploaded to DATRAS.

System component

The following components are part of the DATRAS system:

DATSU: Data screening utility. All data are run through the data-screening utility before they are uploaded to the DATRAS database. The utility consists of:

- an SQL database where the exchange format and all check descriptions are defined;
- a Microsoft.Net program which performs all the checks;
- an Internet front-end where the data submitter can perform the data screening and see the results; and
- an internal Access front-end to the SQL database where exchange formats and checks can be set up by ICES staff.

RECO: Data storage for ICES codes with a web front-end for searches in the database.

DATRAS database: Data storage for raw trawl survey data consisting of an SQL database that can be accessed and downloaded through the Internet.

DATRAS data warehouse: Data storage for processed trawl survey data, with an Internet front-end from which data can be downloaded.

DATRAS has an integrated quality-check utility. All data have to pass an extensive quality check before being entered into the database. It has been developed to collate and document the survey data, ensure data quality, standardize data formats and calculations, and ease data availability. Data from DATRAS are used for stock assessments and fish community studies by the ICES community and public users. DATRAS offers raw data and data products (cpue-based products and indices) for free download according to the ICES data policy, and it can be used for production of custom trawl data products as requested by ICES expert groups or as a part of external projects.

As an example, extensive use of information in DATRAS is documented by Froese and Sampang (2013). They use the SMALK data tables to estimate growth, length–weight, and maturity data. Subsequently, they use the cpue-per-length-per-area table to derive proxy indices for spawning–stock biomass, number of recruits, stock–recruitment relationship, biomass reference points, and total mortality (Figure 2.9). Their proxy stock–recruitment relationship is shown in Figure 2.10.
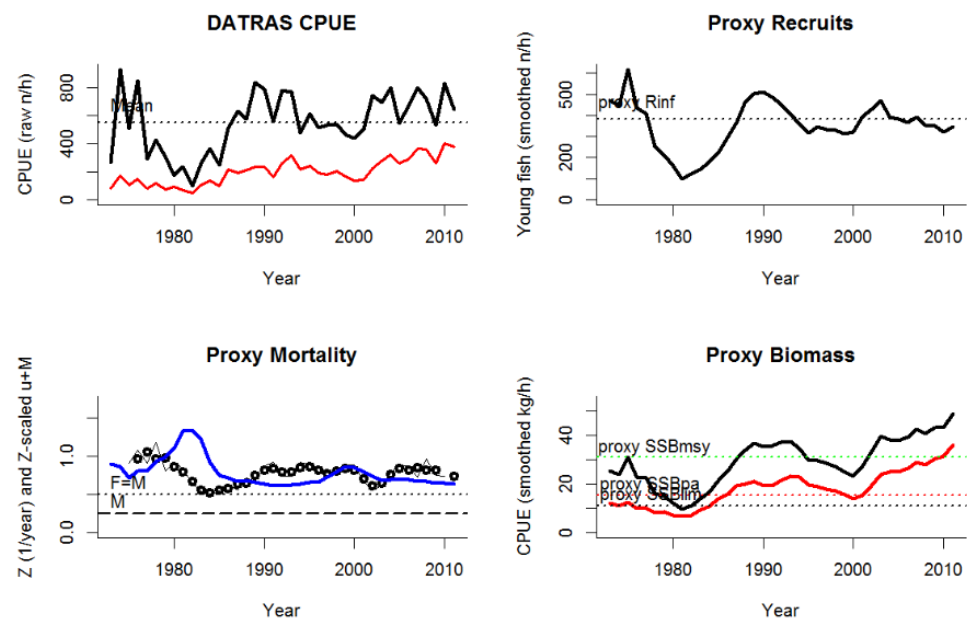
Figure 2.9. Evaluation of a data-limited stock, here North Sea dab (*Limanda limanda*). The black line in the upper left graph shows the raw data obtained from the ICES DATRAS cpue-per-length-per-area database, as numbers of dab caught on average per year by one hour of standardized research trawling. The dotted line indicates the mean of the time-series. The red line indicates the number of individuals larger than the length at 50% female maturity. The upper right graph shows the number of "youngest fish in the survey", as proxy for recruits, with the dotted line indicating the geometric mean of recruitment at large stock sizes. In the lower right graph, length of individuals was converted to weight and summed to show biomass of mature (red) and all fish (black). The dotted horizontal lines are proxy reference points for spawning–stock biomass. The lower left graph shows total mortality (*Z*) experienced over the respective previous two years (black circles) and scaled exploitation rate (blue line), i.e. the ratio *u* between commercial catches and total survey biomass plus the rate of natural mortality, both as proxies for fishing mortality, with indication of natural mortality (*M*, dashed line) and total mortality (*Z*) if *F* = *M* (dotted line) as reference points.
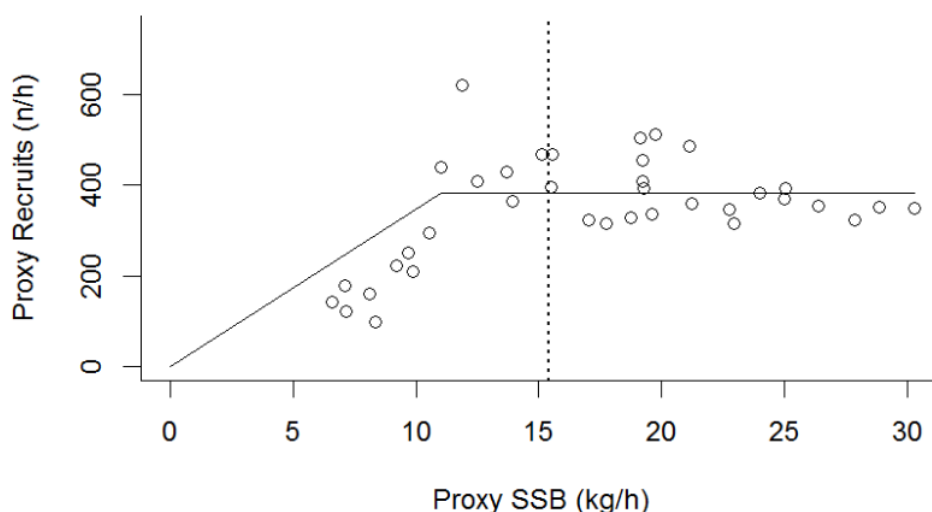
Figure 1.10. A preliminary stock–recruitment relationship based on cpue in weight as proxy biomass of mature fish and number of young fish (= proxy-recruits) two years later, derived from length composition in standardized survey catches. A hockey stick was fitted by a rule-based procedure. The dotted vertical line indicates proxy SSBpa as the precautionary borderline to potentially compromised recruitment.

2.3.6. Description of the ICES stock assessment database

When delivering the results of analytic fish stock assessments, ICES needs to produce graphs and plots that are consistent for all fish stocks. Prior to 2013, this was done using a local Access database with an Excel plotting option. This setup resulted in annual databases containing a basic summary of the assessment, reference points, and yield-per-recruit information for some stocks, but the graphs were only available in the ICES advice. In 2013, ICES developed a web-based interface with a database containing all assessment results since 2001 (Table 2.5). The database is extended with options to add catches split between landings/discards/recreational catches and adding confidence intervals. The database also contains the editorial settings for the graphs, which means that from 2013 onwards, comparison of assessment results over time is possible, and if an assessment is accepted for trends rather than absolute, this information is also available.

The new stock assessment database was developed to be able to present plots based on analytic fish stock assessments in a standardized way containing all assessment information, reference points, and (where available) yield-per-recruit data. The database has a web-based password-protected interface where only stock experts can upload data and edit the graphs.

The database can be accessed (http://standardgraphs.ices.dk/webservices.aspx) and used by outsiders, who can view and download data and plots from published advice.

Table 2.5. The stock assessment database contains the summary of analytic assessments from 2000 onwards.

Parameters		Mean	Confidence intervals	Before 2013
Recruitment		X	If any	
Biomass	Spawning	X	If any	
	Total	If any	If any	
Fishing mortality	Catch	If any	If any	
	Landings	X	If any	
	Discards	If any	If any	
	Industrial	If any	If any	
	Recreational*	If any	If any	
SOP		If any		
Yield/SSB		If any		
Catches	Catch	If any		
	Landings	X		
	Discards	If any		
	Industrial	If any		
	Recreational*	If any		
Stock settings	Recruitment age	X		
	Fishing mortality ages	X		
	Reference points	If any		
Yield-per-recruit information	.sen or .ypr	If any		

Note: "X" = must be available, "if any" = may be inserted if available. For years before 2013, only the highlighted rows could be filled.

The predecessor of this database was used intensively by, e.g. Froese and Proelss (2010) to estimate fisheries reference points (*MSY, Fmsy, Bmsy*) for all ICES stocks with suitable data.

2.4 Experts

Rebecca Whitlock, Polina Levontin, Adrian Leach, and John Mumford

2.4.1 Introduction

Expert opinion is a valuable aid for analysis and decision-making when the available data are very limited, of mixed quality (e.g. noisy), or not directly relevant (O'Hagan, 2012). In some cases, expert judgment may be the best available source of information; however, it is more commonly used in support of available evidence, adding interpretation from different perspectives, e.g. to observed relationships. Expert opinion may also be used to inform policy when urgent measures are required before sufficient scientific evidence is available to serve as the basis for action (Knol *et al.,* 2010).

Formal expert elicitation is a systematic approach for consulting experts on uncertain issues (Knol *et al.,* 2010). Formal elicitation methods have increasingly been developed and applied for incorporating expert knowledge in ecology and fish stock assessment and management (Burgman, 2005; Fletcher, 2005; Martin *et al.,* 2005, 2012; Uusitalo *et al.,* 2005; Kuhnert *et al.,* 2010). The rise of expert knowledge in ecological modelling can be attributed, in part, to the fact that expert judgment is generally sought in situations characterized by high levels of uncertainty, which is a feature of many ecological problems. In this context, experts can bring the benefit of broad experience and the ability

to assimilate and interpret complex and possibly equivocal information (O'Hagan, 2012).

Elicitation of stakeholder views can also contribute to a participatory approach to fish stock assessment and management (Mäntyniemi *et al.,* 2012). It is increasingly acknowledged that management approaches that account for uncertainty and diversity of knowledge and objectives tend to produce improved outcomes for both stocks and fishers (Mackinson *et al.,* 2011), e.g. through more transparent models and greater acceptance of assessment results among stakeholders (Mäntyniemi *et al.,* 2012). Treating the views of different stakeholders as alternative hypotheses and expressing them as probability statements (e.g. Mäntyniemi *et al.,* 2012) allows their incorporation in a quantitative and objective manner.

What constitutes expertise?

In-depth knowledge of the subject matter is usually key to being an expert. Fazey *et al.* (2006) suggest a definition based on experience, whereby the relevance, breadth, and depth of experience pertaining to the subject are an expert's defining features. Expertise also relates to how an expert is able to organize and draw on his/her knowledge and experience. Representation of problems in terms of formal principles and use of known strategies in problem solving have been suggested as attributes of an expert approach (Wood and Ford, 1993).

2.4.2 Selection of experts

The choice of experts is an important part of the elicitation process (O'Hagan *et al.,* 2006). In general, it is desirable to select a group of experts that is broad enough to encompass the complete range of scientific thought on a particular topic (Bedford and Cooke, 2001), so that the results of elicitation will not just reflect one particular opinion. Seeking a set of experts with diverse backgrounds can also help to reduce the potential for "groupthink", i.e. an increased desire for consensus, or group-reinforced bias that can occur when members share a common background (Janis, 1971; Esser, 1998). In addition to knowledge and representativeness, experts' interpersonal skills and communicative ability also merit consideration.

A further important (and difficult to assess) criterion is impartiality; is there a direction in which the expert's personal, professional, or cultural values may influence the elicitation? While it can be argued that all experts are stakeholders in some sense because they can affect the outcomes of models and, through them, policy (Krueger *et al.,* 2012), experts should not have a direct economic or personal stake in the potential findings (Hora and von Winterfeldt, 1997). In case the recruitment of impartial experts is not possible, recording any potential conflict of interest is recommended (O'Hagan *et al.,* 2006).

Several sources may be helpful in identifying experts:

- published literature
- expert networks and expert databases
- records of participants in conferences and symposia
- lists of recipients of academic awards and honours
- professional social media (LinkedIn, ResearchGate, etc.)

2.4.3   Potential sources of bias

The facilitator should be aware that judgment under uncertainty is subject to a number of well-documented biases. The use of heuristics (or short-cuts) to make judgments under uncertainty can lead to systematic biases in experts' evaluations (Tversky and Kahneman, 1974). Thorough discussion of these biases and their implications for elicitation practice can be found in Gilovich *et al.* (2002), while reviews are provided in Otway and von Winterfeld (1992) and O'Hagan *et al.* (2006). In the following section, several potential sources of bias that are particularly relevant to elicitation of expert opinion are reviewed.

Anchoring

In judging the value of a quantity under uncertainty, people have a natural tendency to associate their answer with some (possibly entirely random) anchor. Thus, they start with an initial estimate (the anchor) and adjust up or down. The bias arises because people tend not to adjust sufficiently far away from the initial estimate. This effect persists when anchors are entirely arbitrary and even when the units of the anchor are not the same as the quantity to be elicited (Wong and Kwong, 2000). This is clearly problematic in the context of eliciting expert opinion, since we want the results of the elicitation to reflect only the expert's knowledge. Some effects of anchoring include insufficient revision of prior beliefs (which act as an anchor) in the face of new evidence (Edwards, 1982) and anchoring on a value corresponding to equal probability for a set of mutually exclusive categories, such that the probability assigned to one identical category will differ depending on the number of categories (Fischhoff *et al.*, 1978). In order to reduce the effect of anchoring, provision of quantitative examples should be avoided during elicitation. It might also be possible to reduce anchoring via the ordering of questions (e.g. avoiding serial questions on a related quantity where anchoring may lead to assessments that are too close together).

Availability

The ease with which instances of a certain event (or instances of a particular class of events) come to mind informs people's judgments about the probability of that event (or the size of that class) – this is known as the availability heuristic (Tversky and Kahneman, 1974). While the availability heuristic is likely to be effective in general, it may also introduce bias into expert opinion when factors other than frequency or probability influence the ease with which examples come to the expert's mind. Personal experience, recent events, and disproportionate media coverage of infrequent events are factors that can lead to overly high probability assessments.

Representativeness

The representativeness heuristic uses an assessment of the degree of correspondence between an outcome (e.g. a sample or instance) and a model (e.g. a population or category) as a proxy for a probability assessment (Tversky and Kahneman, 1973; O'Hagan *et al.*, 2006). This heuristic is particularly relevant to elicitation in that it applies to the assessment of single-event probabilities for unique events (O'Hagan *et al.*, 2006). Bias can be introduced through the fact that representativeness and probability are imperfectly correlated (e.g. an outcome may be representative, but improbable). A number of biases have been attributed to the representativeness heuristic:

- Conjunction fallacy

People's tendency to assess the conjunction of two events as being more probable than one of the events separately is termed the conjunction fallacy (Tversky and Kahneman, 1983). This can occur if one of the events in the conjunction is deemed unlikely because of being unrepresentative. Framing questions in terms of frequencies (e.g. "4 out of 15" rather than "27%") can help to reduce bias associated with the conjunction fallacy (e.g. Jones *et al.*, 1995).

- Base-rate effect

People's tendency to ignore the relative frequency of an event within some defined class is termed the base-rate effect (O'Hagan *et al.*, 2006). In practice, this means that people assign too much weight to an individual observation or event, neglecting the available background information. To illustrate this by example, consider that a patient receives a diagnosis that he/she has tested positive for the human immunodeficiency virus (HIV), and that the test gives a low rate (1%) of both false positive and false negative results. Suppose also that HIV is very rare among the general population, with an incidence of only 1 in 10 000. In this example, the patient's conclusion (based on the test result) that he/she is, in fact, HIV positive would constitute the base-rate effect, since it neglects the information about the rareness of HIV. This can be seen by noting that out of 1 million people tested at random, 100 are expected to be HIV positive, whereas the test will give a positive diagnosis for about 1%, or 10 000 people. Formulating questions in terms of frequencies may lessen the base-rate effect.

- Insensitivity to sample size

Even experts may fail to fully appreciate the fact that extreme results are less likely in large samples than in small samples. As a result, they may show a tendency to be overconfident relative to what can be concluded about population parameters from a small sample (Tversky and Kahneman, 1971).

Overconfidence

Overconfidence refers to the tendency of experts to give probability distributions that are too narrow. This has been demonstrated using experiments that assess experts' performance at gauging credible intervals (the x% credible interval is the range of values for y such that there is an x% probability that the true value of y lies in this range). For example, if experts' credible intervals corresponded well with reality, the proportion of 50% credible intervals that contain the true value should be about 50% (O'Hagan *et al.*, 2006). In fact, assessors' 90, 95, and 98% credible intervals have been found to contain the correct answer between 40 and 70% of the time (Alpert and Raiffa, 1982; Lichtenstein *et al.*, 1982). It has been suggested that overconfidence may result from people's concerns about being so uninformative as to be unhelpful (Yaniv and Foster, 1997) or from a type of anchoring where the lower and upper limits are anchored to the central or best estimate (Slovic, 1972). In order to reduce the effects of overconfidence, training of experts in probability and calibration exercises with feedback have been recommended (e.g. Lichtenstein and Fischhoff, 1980). O'Hagan *et al.* (2006) suggest that asking experts to provide credible intervals for moderate levels of uncertainty (e.g. 50%) is preferable to asking for credible intervals for high levels of certainty (e.g. ≥95%) in this respect.

Is expert opinion independent from data used in the model?

When expert knowledge comprises only part of the total information about a parameter or model, care is needed to ensure that the same data are not used twice, leading to

overly precise assessments. This can occur if some of the observations that will be used to fit a model are available to experts and used by them to formulate priors. Using the same data that will appear in a model's likelihood function to inform priors violates the likelihood principle, which requires that all interpretation of the data should enter inference via the likelihood function (Berger and Wolpert, 1988; Mäntyniemi *et al.*, 2012). In practice, it may not be possible to restrict experts' access to data that will be used to update their priors; for example, when eliciting a prior for initial stock size for a new stock assessment model, experts' responses might reflect the results of other assessments that have used the same data. However, experts should be made conscious of this issue and asked to try to exclude interpretations of data that will be used in the likelihood function from their judgments to the extent possible.

Motivational bias

Motivational biases include impartiality and groupthink, as well as misinterpretation (inadequate translation of knowledge into response), wishful thinking (expert's hopes influence their judgments), and impression management (response in a way that will be perceived as politically correct) (Booker and McNamara, 2004). An asymmetrical incentive structure can also lead to biased probability assessments. For example, if experts are held accountable when an undesirable event occurs, but otherwise face no consequences, their assessment of the probability of that event occurring may be biased high (Burgman, 2005). In order to minimize this type of bias, it is necessary to carefully consider the incentive structure for a particular problem. Building a trusting relationship between expert and facilitator may help to reduce some forms of motivational bias (Oliver *et al.*, 2012).

2.4.4 Training of experts

Some understanding of probability is required for effective elicitation of prior probability distributions; therefore, experts may require training in this regard. O'Hagan *et al.* (2006) write:

"Substantive expertise in a specialist area is no guarantee of normative expertise in providing coherent probability assessments. Careful thought needs to be given to the training in probability and statistics that the expert should receive at the beginning of the elicitation exercise".

Establishing the extent of experts' understanding of probability and distributions can help to give an idea of the amount of training that may be needed (O'Hagan *et al.,* 2006). Training about probability distributions might include highlighting their key features, e.g. the fact that values around the mode are judged more likely than those in the tails, or that values far out in the tails can be thought of as nearly impossible. The concept that the area under the curve in a given region bounded by x_1 and x_2 corresponds to the probability of $x_1 < X < x_2$ is also important. The "chips in bins" or "Roulette" method of elicitation (Oakley and O'Hagan, 2010) may be a useful tool for experts who lack training in probability distributions. This involves giving the expert n "chips" to distribute among m bins, so that the proportion of chips allocated to a particular bin corresponds to the expert's probability of X lying in that bin. The Roulette method uses a probability format that is equivalent to the frequency format, which may initially be more accessible to experts.

Experts might feel uncomfortable about providing a personal probability (e.g. if they believe that the subjectivity of personal probability carries a negative connotation). In such cases, familiarization with the idea of probability as a degree of belief – and the notion that personal knowledge is inherently subjective, varying from person to person

– may be required. Presenting experts with published examples of elicitation that use subjective (personal) probabilities may also help legitimize the process in their eyes. Experts who are concerned that their opinions may not be "correct" should be reminded that the objective of the elicitation is to capture their knowledge in a form that expresses neither too much nor too little uncertainty (O'Hagan *et al.,* 2006).

O'Hagan *et al.* (2006) go on to suggest that, in addition to covering probability and probability distributions, training should give experts an awareness of the most common biases that arise from the use of heuristics (discussed above) and how to overcome them. For example, to counter overconfidence, experts could be encouraged to consider alternative models for the processes affecting the variable of interest.

Another key component of training is practice elicitations in which experts are asked to answer questions whose answers are known to the facilitator, but not to the expert. Performing several such exercises with feedback to the experts can help experts give better-calibrated assessments (i.e. assessments of probability that correspond more closely to reality). For example, this process might alert the expert to his/her tendency towards overconfidence. An example (eliciting the distance between two towns) can be found in O'Hagan (1998). Training questions with feedback and tests in which experts are asked to judge the same variable in two or more ways are additional methods of expert calibration that can help to counter bias (Krueger *et al.,* 2012).

3   Methods

Samu Mäntyniemi, Atso Romakkaniemi, Etienne Rivot, Rebecca Whitlock, Henni Pulkkinen, Rainer Froese, Konstantinos Stergiou, Polina Levontin, Adrian Leach, Juho Kopra, Charis Apostolidis, John Mumford, and Sakari Kuikka

3.1   Introduction to methods

Samu Mäntyniemi, Atso Romakkaniemi, Etienne Rivot, and Rebecca Whitlock

The methodologies to aid in deriving priors must fulfill the criteria of transparency and thorough documentation. Guidelines for choosing the exact method to use in a particular situation can increase consistency and transparency, not only through establishing a common understanding for the methodological choices, but also by providing a common ground for arguing for or against any particular methodological choice. Formal methods for prior formulation naturally enhance the consistency and rigorousness of the process.

Section 2 describes various information sources that can be used to formulate priors and classifies them into four main categories (primary data, literature, online databases, and experts). As pointed out earlier, online databases typically contain processed/analyzed data taken from the literature; thus, this category is clearly a derivative of the others. In addition, some categories are closely linked (e.g. experts may have attained their expertise by collecting and/or analyzing primary data and by assimilating information in the literature). Consequently, throughout compilation and processing of any type of prior information, it is worth keeping in mind that:

- Categorization is useful mainly for structuring the process and acknowledging different aspects of it.
- All the categories are ultimately grounded in observations/data.
- Data (i.e. observations) also have subjective features, e.g. the choices of what type of data is collected, how it is collected, and the criteria for inclusion in a database are man-made decisions.

Depending on the type and quantity of information available, a number of methods can be used to formulate priors, ranging from methods based on processed information, such as hierarchical meta-analysis (e.g. Sutton and Abrams, 2001; Royle and Dorazio, 2008), empirically-based methods (e.g. McAllister *et al.*, 2001), and methods to extract and codify expert opinion (e.g. O'Hagan *et al.*, 2006). While specific examples of the above approaches can be found, the methodological literature focusing on deriving priors is somewhat fragmented and, as such, offers limited support regarding selection of the most suitable method from the available options.

Here, we attempt to build a "decision tree", indicating the most appropriate choice(s) of information sources and methodology in various situations (Figure 3.1).
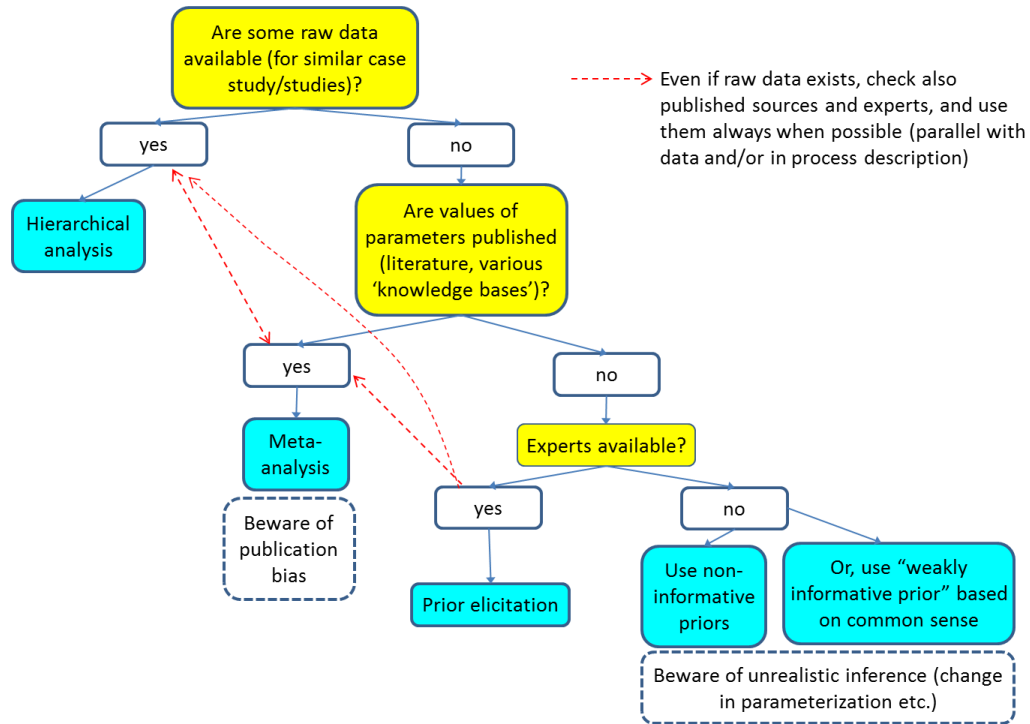
Figure 3.1. A schematic "decision tree" aiding in the process of deriving prior information.

Given that all existing information has its grounding in observations, we propose a hierarchy in which sources of prior information are ranked according to their proximity to the primary observations (i.e. raw data), so that use of raw data is preferred where possible. This hierarchy is reflected in the types of methods that might be suitable, e.g. hierarchical analysis and meta-analysis approaches are powerful, but typically require larger numbers of observations (either primary observations or some derived quantities) than the other methods. As a next option (after looking for raw data), we advise checking the literature. Meta-analyses of results from the literature are widely undertaken in research, and the Bayesian methods for meta-analysis are reviewed by Sutton and Abrams (2001) as well as later in this section. Expert opinion lends itself to being used in extremely data-limited situations and can be used even in cases where no observations (either primary or from the literature) are available. It should be noted that elicitation of parameter values is not the only task where experts play an important role; they also have expertise in describing the process to be modelled as a whole.

These information sources and methods are not mutually exclusive, so that some combination may be used in deriving a prior distribution. Information from the literature could be used on its own (in the absence of raw data), but it can also be used in conjunction with raw data. Similarly, expert knowledge could be utilized together with other sources of information. For example, prior information from the literature or from experts could be used in an analysis of raw data to obtain the prior distribution for a further analysis (i.e. a sequential Bayesian approach, Michielsens *et al.,* 2008). Whichever method is chosen, it is important to remember that the same data should not be used twice. If the plan is to use the data in the analysis for which the prior distribution is needed, then the same data cannot be used in formulating the prior. This should also be considered when using expert opinion.

We advocate the use of existing knowledge and experience in the formulation of priors wherever possible. In case, this is not feasible (bottom-right corner of Figure 3.1); meth-

ods for deriving non-informative priors (i.e. priors that contain little information relative to the data) can be found in Jeffreys (1961), Box and Tiao (1973), and Bernardo (1979), although they should be applied with care (Punt and Hilborn, 1997). Priors that are uniform on some scale are often used as non-informative priors in fish stock assessment. However, owing to sensitivity to the measurement scale used, a uniform prior will not necessarily be non-informative for all the quantities of interest in a model. We recommend that users take special care when using parameter transformations. For instance, it is classical to use the logit() transformation on some parameters that must take values in the range [0,1], such as a survival rate, θ. Using a very flat (large variance) normal distribution (or even a uniform distribution over a wide range) on the logit scale turns out to be an inappropriate choice as it corresponds to a prior distribution on θ that gives very high weights to values of θ near 0 and 1 and may thus have an unwanted impact on the final inference. To avoid such kinds of misspecification, we strongly recommend that users always simulate the model based on priors only (i.e. with no data) to check whether the priors really correspond to the amount of information they want to introduce *a priori* into the model.

Assuming that prior information exists and can be used, the next subsections specify the contents of the toolbox of methods presented in the above "selection tree" and provide examples of each method. Meta-analysis is an established term to be used when the results of various studies are combined, i.e. when literature is used as the source of information. Hierarchical modelling can be used to analyze both primary data and literature (thus also databases) and encompasses the Bayesian alternative for the meta-analysis. We start with a detailed description of hierarchical modelling methods that can be applied in a meta-analysis or hierarchical modelling context as defined above. Then, we introduce the online toolbox developed in the ECOKNOWS project to analyze FishBase information. Finally, we provide illustrative examples and discuss methods for deriving priors using expert knowledge.

3.2 Hierarchical modelling

Henni Pulkkinen, Etienne Rivot, and Rebecca Whitlock

In population modelling, the data available from the stock of interest are very valuable. Those data are thus usually reserved for model fitting (i.e. for use in observation models). In a full Bayesian analysis, external sources of information can be used to derive informative prior distributions for model parameters.

Let us denote θ the vector of parameters of interest and y the data available to estimate θ. In a Bayesian analysis, the prior distribution $p(\theta)$ synthesizes the knowledge about the value of θ before seeing the data, and the prior is updated by the data into the posterior distribution $p(\theta|y)$. In its simplified unnormalized form (see Gelman *et al.*, 2004 for more details), the posterior distribution is defined as the product of the prior and the likelihood, which is the sampling distribution of the data in the model that conveys the information brought by the data:

$$p(\theta|y) \propto p(\theta) \cdot p(y|\theta) \qquad (8)$$

The posterior distribution in (8) then synthesizes the *a priori* knowledge and the information brought by the data.

However, in data-poor situations, the data available for the stock of interest may be scarce or even unavailable. In this context, hierarchical modelling approaches are ex-

tremely useful as they allow learning via synthesis of information from other conspecific or related stocks. Often, data are available from other stocks of the same species or from other related species, and these data can be used to build informative prior distributions for the stock of interest with hierarchical models.

Hierarchical Bayesian models (HBM) offer a natural way to model variation at multiple levels (e.g. at the population and meta-population levels) and to "borrow strength" between and within the different levels of a model. This makes them suitable for many applications in statistical ecology; as a result, they have received considerable attention in this field (Clark, 2003; Royle and Dorazio, 2008; Kéry and Schaub, 2012; Parent and Rivot, 2012). Hierarchical models have been used extensively in fisheries science to improve estimates of stock–recruitment parameters (Myers, 2001; Prévost *et al.*, 2003; Michielsens and McAllister, 2004; Su *et al.*, 2005; Hillary *et al.*, 2012; Pulkkinen and Mäntyniemi, 2013; Archambault *et al.*, 2014), growth parameters (Jiao *et al.*, 2009), length–weight and length–fecundity relationships (Pulkkinen *et al.*, 2011), selectivity curves (Harley and Myers, 2001), and catchability (McAllister *et al.*, 2004; Robert *et al.*, 2010).

Figure 3.2 illustrates the basics concepts of hierarchical models. Let us suppose that data are available for $i = 1,…, k$ units, and define y_i as the set of available data for unit i that can be used to estimate some parameters θ_i defined for each unit i. The k units can represent different sites, populations, or individuals. For example, in a typical fisheries problem, the k units might be different stocks of the same species for which we wish to estimate stock-specific stock–recruitment parameters θ_i from the available data y_i.

Hierarchical models seek to build a statistical model for all units $i = 1,…, k$ by making an explicit hypothesis about the between-unit variability of the parameters θ_i. It is often assumed that the θ_i are drawn *a priori* from a common probability distribution that depends on common parameters ϕ (often referred to as hyperparameters that are also assigned prior distributions, which are sometimes referred to as hyperpriors) and possibly on some covariates x_i specific to each unit (e.g. time, environmental covariates).

The hierarchical structure sets the dependency between the units i by expressing both similarity and heterogeneity among the θ_i. The hyperparameters ϕ most often define the mean and variance of the prior distribution of the θ_i parameters $\phi = (\mu_\theta, \sigma_\theta)$. A small variance will express a closer resemblance between groups. The prior distribution on the common hyperparameters ϕ will be updated by the observations on all units $i = 1,…,k$. This updating of the hyperparameters has two main advantages:

- First, it allows transfer of information between the different units (the concept of borrowing strength from data-rich to data-poor units; McAllister *et al.*, 2004; Punt *et al.*, 2011; Parent and Rivot, 2012) that can improve the estimates of θ_i for data-poor units.

- Second, it provides a structure for proposing an informative prior distribution for parameters θ^{new} for a new unit for which no data are available, but that can be considered as belonging to the same family (e.g. stock–recruitment parameters for a new stock). This is accomplished through the posterior predictive distribution $P(\theta^{new}|y_{1:k})$, where $y_{1:k}$ denotes the observations on all units I = 1,…,k. The posterior predictive distribution is a natural product of a hierarchical model. This is the distribution of a new θ that would be drawn in the same population as $(\theta_1, … , \theta_k)$ and then synthesizes the information brought by the data from all units. As detailed below, the

posterior predictive distribution depends on the structure of the hierarchical model. In the following text, these concepts are illustrated using examples of varying complexity.

3.2.1 Exchangeable hierarchical models

When applying a hierarchical model, exchangeability is often assumed for the set of parameters $\theta_{1:k}$. This means that the parameters θ_i for each unit i are considered to be independently drawn from the same prior distribution conditioned by common hyperparameters ϕ:

$$p(\theta_{1:k}|\phi) = \prod_{i=1}^{k} p(\theta_i|\phi) \tag{9}$$

Parameters for each unit i are then associated with observation y_i through a likelihood function $p(y_i|\theta_i)$ (Figure 3.2). Therefore, the observation model for unit i depends only on θ_i, but the hyperparameter ϕ (on which a (hyper)prior $p(\phi)$ is set) links all the units.

Exchangeability implies that there is no prior knowledge of any feature that would account for differences between the parameters of the different units $i = 1,..,k$. Exchangeability can be considered as something intermediate between having identical parameters in all the units and total independence between the unit-specific parameters – parameters for different units have something in common even though they differ from each other. Looking at equation (9), this implies that all permutations of the unit indices i will leave the joint prior distribution $p(\theta_{1:k}|\phi)$ unchanged.

Exchangeable hierarchical models constitute a consistent framework for prediction (Figure 3.2). Inference about the parameter for a new unit, denoted as θ^{new}, can be derived through the posterior predictive distribution conditioned by data from all units $y_{1:k}$. This is an integration (an average) of the distribution $p(\theta^{new}|\phi)$ over the posterior distribution of the hyperparameters ϕ, $p(\phi|y_{1:k})$:

$$p(\theta^{new}|y_{1:k}) = \int_{\phi} p(\theta^{new}|\phi) \cdot p(\phi|y_{1:k}) \cdot d\phi \tag{10}$$

The posterior predictive distribution in equation (10) is a natural product of a hierarchical Bayesian model. The posterior distribution of the hyperparameters $p(\phi|y_{1:k})$ results from updating the prior set on the hyperparameters $p(\phi)$. For example, if the hyperparameters ϕ define the mean and variance of the prior distribution of the θ_i parameters, e.g. $\phi = (\mu_\theta, \sigma_\theta)$ as denoted above, the data available for all units $y_{1:k}$ will be used to learn about the mean μ_θ and variance σ_θ. The posterior predictive distribution in equation (9) then combines the distribution of a new θ with respect to the variability of the θ between units as defined in the model $[p(\theta^{new}|\phi)]$ with what has been learned from all available data about the parameters that control this variability between units $[p(\phi|y_{1:k})]$.

Inference in hierarchical models can be thought of as a flow of information (Figure 3.2). First, information from the observed quantities $y_{1:k}$ flows to the hyperparameters ϕ updating them; the information then continues its journey from the hyperparameters ϕ to the posterior predictive distribution of θ^{new}. The latter distribution can be used as an informative prior distribution for θ^{new} for a new unit for which no data are available, but that can be considered to belong to the same family.

Figure 3.2. Directed acyclic graph for an archetypal exchangeable hierarchical model. Each unit i = 1,…,k has its own parameter θ_i and observations y_i. The parameters θ_i are supposed to be independently drawn from the same prior distribution conditioned by hyperparameters ϕ (exchangeability). This provides a structure for proposing an informative prior distribution for θ^{new} for a new unit for which no data are available, but that can be considered to belong to the same family. The green shaded arrows indicate the flow of information.
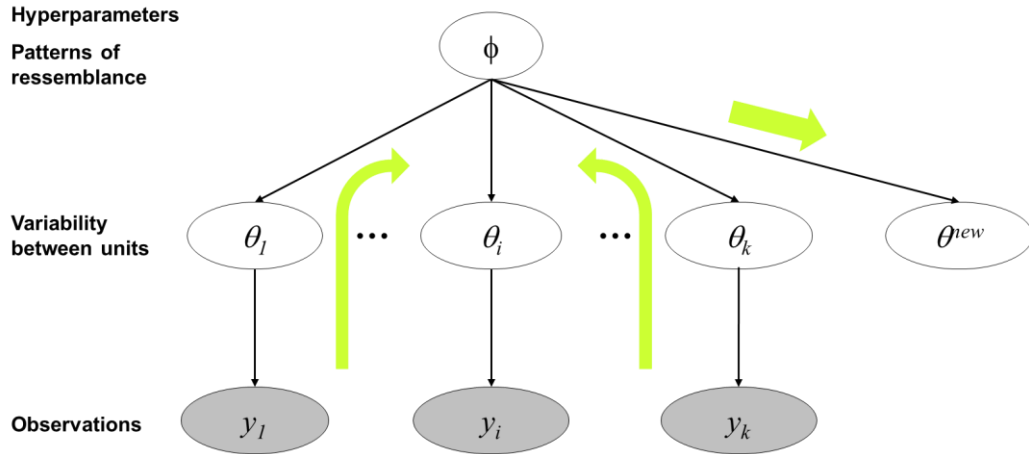
In order to examine the exchangeability of units in a hierarchical model, Michielsens and McAllister (2004) suggest running the model excluding one dataset at a time. If a parameter is exchangeable, the exclusion of any one dataset should not substantially alter its posterior predictive distribution (Michielsens and McAllister, 2004). In other terms, let us define y_{-i} the dataset obtained from $y_{1:k}$ by excluding the data of unit i. If parameter θ is exchangeable with regard to the population of parameters $\theta_{1:k}$, the posterior predictive distribution of a new θ obtained with the altered data set y_{-i}, that is $p(\theta^{new}|y_{-i})$, should not vary much when i varies in $1,…,k$.

Example 3.2.1: Estimating predictive distribution for the slope of the stock–recruitment relationship of Atlantic herring

Let us suppose we are interested in obtaining an informative prior distribution for the slope at the origin parameter of the Beverton–Holt stock–recruit (SR) relationship for a herring (*Clupea harengus*) stock. Further suppose that data are available from $i = 1,…,7$ stocks other than the stock of interest and that the data for those stocks are desired to be used to form an informative prior distribution for the analysis of the stock of interest. In this example, a model is provided to analyze those data and to estimate a posterior predictive distribution for the slope at the origin parameter. This example is similar to the one in Pulkkinen and Mäntyniemi (2013), except that the hierarchical model contains only the Beverton–Holt SR function. BUGS/JAGS code for the example can be found in Annex 1.

Based on the Beverton–Holt SR function, the expected number of recruits from stock i in year t, $\mu_{i,t}$, depends on the maximum production K_i, slope at the origin α_i, and the egg production $E_{i,t}$ of the corresponding spawning cohort:

$$\mu_{i,t} = \frac{K_i E_{i,t}}{\frac{K_i}{\alpha_i} + E_{i,t}} \tag{11}$$

As the spawning–stock size is considered in terms of the number of eggs, the slope at the origin can be interpreted as the maximum survival of eggs to recruits, taking values in the interval [0,1]. By using a logit() transformation for α_i, the resulting logit-transformed parameter can take any positive or negative value, so a normal prior distribu-

tion seems like an appropriate choice for the logit-transformed parameter. An uninformative prior distribution can be used for the hyperparameters μ_α and σ_α that describe the mean and variation of the logit-transformed slope parameter among different stocks:

$$logit(\alpha_i) = A_i \tag{12}$$
$$A_i \sim N(\mu_\alpha, \sigma_\alpha)$$
$$\mu_\alpha \sim N(-1,1000)$$
$$\sigma_\alpha \sim logN(-2,20).$$

The posterior predictive distribution for a new herring stock (α^{new}) that could be used as an informative prior in further analyses can then be simply obtained as:

$$logit(\alpha^{new}) = A^{new} \tag{13}$$
$$A^{new} \sim N(\mu_\alpha, \sigma_\alpha),$$

where the hyperparameters ($\mu_\alpha, \sigma_\alpha$) are integrated out using their posterior distribution after observing the available data for the seven herring stocks; this is similar to how the hyperparameters ϕ were integrated out in equation (10). Figure 3.3 illustrates the estimated slope at the origin parameters for the seven herring stocks from which data are available and the posterior predictive distribution for a new stock of herring.
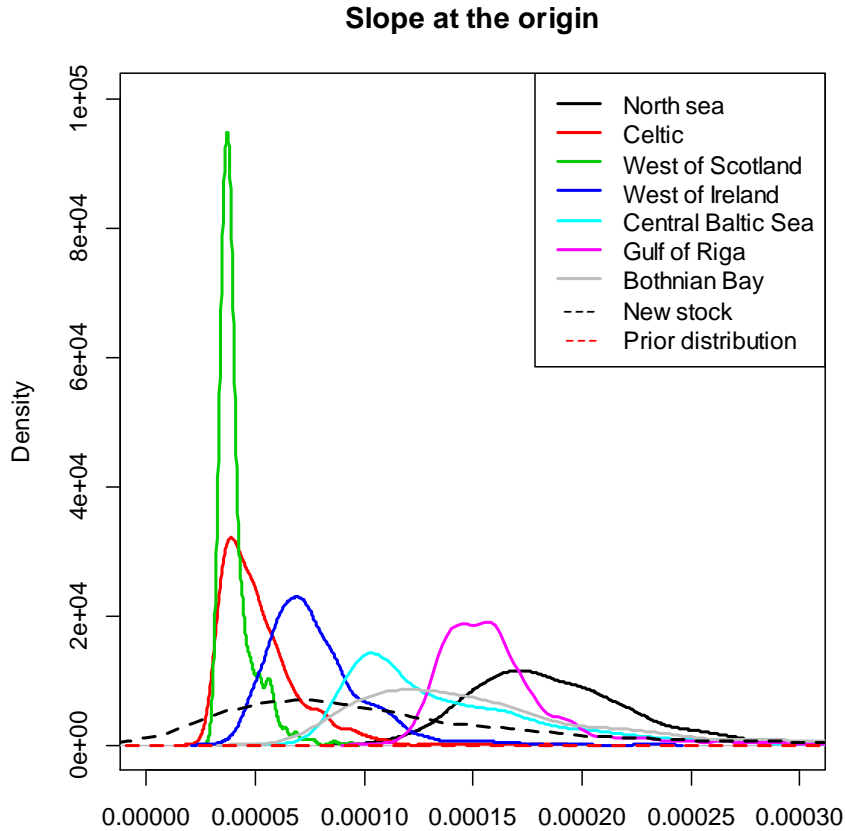
Slope at the origin

Figure 3.3. An uninformative prior (red dashed) and posterior distributions (solid lines) for the stock-specific slope at the origin for seven stocks, with a predictive distribution for a new herring stock (black dashed).

3.2.2   Choice of prior on hyperparameters

Inferences derived from a hierarchical model are generally robust to the choice of prior distributions on the hyperparameters ϕ.

Inferences derived from models that consider independent statistical units can be sensitive to the choice of priors, particularly for units associated with weakly informative data. Thanks to their ability to share information between units, HBMs limit these undesirable effects. In the HBM framework, individual posterior distributions of the parameters for each unit are obtained by combining the likelihood with a modified prior that is constructed by combining information from all other units in a consistent manner (see e.g. Rivot and Prévost, 2002). Inferences about the unit-specific parameters θ_i can be significantly improved in terms of both precision and robustness by comparison with inferences derived from a model where all specific parameters are considered independent.

A common approach consists of using weakly informative priors for the hyperparameters ϕ. However, it is strongly recommended that the predictive prior distribution of the unit-specific set of parameters $(\theta) = \int_\phi p(\theta|\phi) \cdot p(\phi) \cdot d\phi$ constructed from the combination of the hierarchical structure $p(\theta|\phi)$ and the prior on hyperparameters $p(\phi)$, should always be evaluated. In particular, one must avoid an inappropriate choice of priors on ϕ that would strongly limit *a priori* the between-unit variance of the parameters θ because such a prior may result in a very strong and possibly undesirable shrinkage effect among units.

3.2.3   Using covariates in partially exchangeable models

Exchangeable hierarchical models make the strong assumption that the unit-specific parameters θ_i can be considered *a priori* as independent random draws from a common probability distribution $p(\theta|\phi)$. This assumes that there is no *a priori* knowledge of any feature that would explain the differences among units *i = 1,..,k*.

In many cases, however, some covariates x_i (typically environmental covariates) are known *a priori* for each unit *i* and can be utilized to explain some of the differences between the θ_i. Then, a partially exchangeable model can be constructed which assumes the θ_i are *a priori* independently drawn from a prior distribution conditioned by hyperparameters ϕ and unit-specific covariates x_i. A typical case is when the expected mean of the θ_is is not constant (as in an exchangeable hierarchical model), but can be thought of as a parametric function of x_i (e.g. $\mu_{\theta_i} = \nu_\theta + \lambda_\theta \cdot x_i$).

Such partially exchangeable models also provide a consistent framework for deriving an informative prior for θ^{new} based on the posterior predictive distribution (as in the exchangeable case), provided that the covariate x^{new} is available for the new unit. The posterior predictive of θ^{new} is, therefore:

$$p(\theta^{new}|y_{1:k}, x_{1:k}, x^{new}) = \int_\phi p(\theta^{new}|\phi, x^{new}) \cdot p(\phi|y_{1:k}, x_{1:k}) \cdot d \qquad (14)$$

Example 3.2.2: Estimating predictive distributions for Atlantic salmon management reference points

This example develops a hierarchical stock–recruitment model for Atlantic salmon (*Salmo salar*) stocks using a more sophisticated hierarchical structure than the ex-

changeability example in Section 3.2.1. In the following example, the hierarchical probabilistic structure employed to capture the variability among stocks is designed conditionally on some covariates, namely the latitude of the river associated with each stock.

The example is directly drawn from Chapter 9 of Parent and Rivot (2012) and is inspired by the data and model originally published by Prévost *et al.* (2003).

The model uses available stock–recruitment series for 13 monitored rivers throughout the European range of the species. This sample of rivers covers a broad area including Spain, France, UK, Ireland, Norway, the west coast of Sweden, and the southwest coast of Iceland.

SR relationships are used to estimate reference points for the management of salmon populations, such as the spawning target S^*, a biological reference point for the number of spawners necessary to guarantee a sustainable exploitation rate, and the maximum sustainable exploitation rate h^*. Here, the hierarchical model is used as a tool to forecast biological reference points for a new river without any SR data, but for which relevant covariates (latitude) are available.

Let us denote the indices for the 13 rivers as $i = 1,...,13$. Within a river i, recruitment is modeled by a Ricker function with independent log-normal process errors. We use the reformulation of the Ricker model with management parameters (S^*, h^*), as defined in Schnute and Kronlund (1996):

$$\log(R_{i,t}) \sim N\left(\mu_{R_{i,t}}, \sigma^2\right) \qquad \mu_{R_{i,t}} = h_i^* + log\left(\frac{S_{i,t}}{1-h_i^*}\right) - \frac{h_i^*}{S_i^*} \cdot S_{i,t} \qquad (15)$$

The prior distribution on the parameters S_i^* and h_i^* explicitly incorporates the latitudinal gradient, as shown next.

For S_i^*, the latitudinal gradient can be incorporated by writing a log-linear relationship between the prior expectation of S_i^* and the latitude of river i denoted x_i. As no information is available on the slope α and intercept β of the linear regression on the log scale (those are hyperparameters as defined above), they are given flat priors:

$$\log\left(\mu_{S_i^*}\right) = \alpha \cdot x_i + \beta \qquad (16)$$

$$\alpha \sim uniform(-5,5)$$

$$\beta \sim uniform(-50,50)$$

Given the prior mean $\mu_{S_i^*}$ and coefficient of variation CV_{S^*}, the parameter S_i^* is drawn *a priori* from a gamma distribution with parameters a_i and b_i defined as follows:

$$CV_{S^*} \sim uniform(0,20) \qquad (17)$$

$$a_i = \frac{1}{CV_{S^*}^2}$$

$$b_i = \frac{1}{\mu_{S_i^*} \cdot CV_{S^*}^2}$$

$$S_i^* \sim gamma(a_i, b_i)$$

Because h_i^* varies between 0 and 1, the logit() transform of h is used to model the gradient with increasing latitude as a linear regression on μ_{h^*} on the logit scale:

$$\mu_{h_i^*} = \delta \cdot x_i + \kappa \qquad (18)$$

$$\delta \sim uniform(-5,5)$$

$$\kappa \sim uniform(-50,50)$$

Given the prior mean $\mu_{h_i^*}$, a river-specific contribution is modeled as a normal distribution with precision τ_{h^*} (variance $1/\tau_{h^*}$) that expresses the residual degree of similarity between rivers as soon as the latitudinal gradient is accounted for:

$$logit(h_i^*) \sim normal(\mu_{h^*}, 1/\tau_{h^*}) \tag{19}$$

$$\tau_{h^*} \sim gamma(0.001, 0.001)$$

Figure 3.4 shows that the posterior probability $p(\alpha < 0|y_{1:k})$ is null, while $p(\delta < 0|y_{1:k}) = 0.05$, which indicates that the covariate latitude offers a good statistical explanation of variation between rivers in both h^* and S^*.

Figure 3.5 shows the posterior distribution of parameters (S^*, h^*) for the 13 rivers and the posterior predictive distributions, which represent our uncertainty/knowledge without SR observations. The marginal posterior predictive distributions of h^{*new} and S^{*new} at various latitudes covering the salmon range in the Northeast Atlantic area (46, 52, 59, and 63°N) are shown. They indicate that, when moving north, salmon stocks can sustain higher exploitation rates h^*, but at the same time, higher conservation limits S^* should be set. These posterior predictive distributions could be used as informative priors for any river for which the latitude is known.

Even if the case study is different, it is useful to point out the similarities with the herring example in Figure 3.3, where the hierarchical structure is considered exchangeable; the informative prior distribution of a SR parameter for a new herring stock will not depend on any covariates. In the salmon example in Figure 3.5, knowing the latitude of the river will greatly improve the prediction of the associated salmon stock.
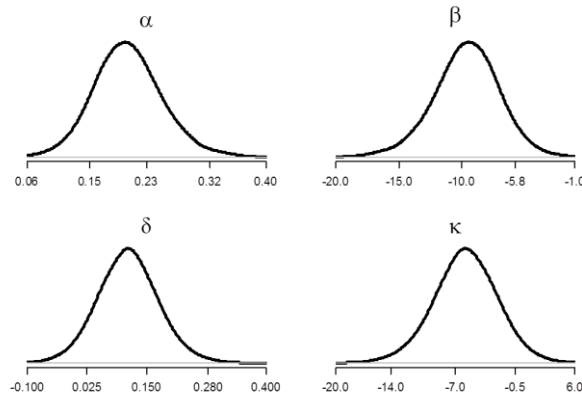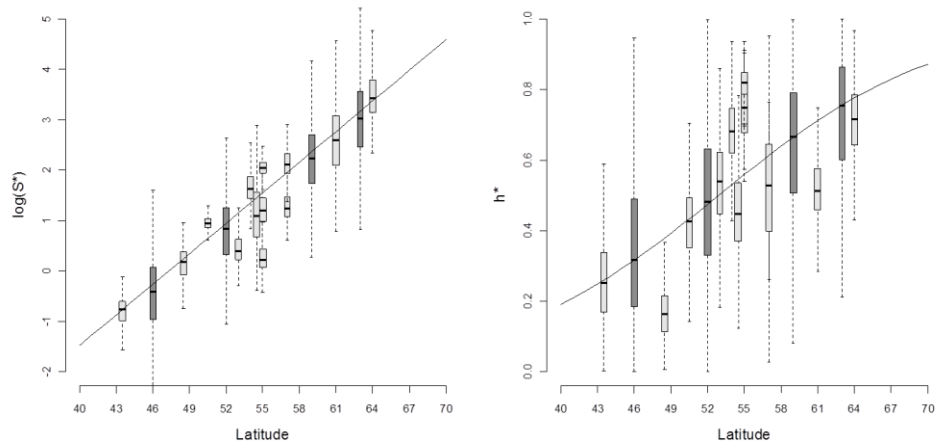

Figure 3.4. Marginal posterior probability shapes of the parameters (α, β), (δ, κ), **from the hierarchical model (source: Parent and Rivot, 2012, Chapter 9).**

Figure 3.5. Marginal posterior distributions of $\log(S^*)$ and h^* for the 13 rivers obtained under the hierarchical model with latitude as a covariate (light grey). The posterior pdfs are graphed as a function of latitude. The thin lines are the regression in equations (16) and (18) with parameters (α, β) and (δ, κ) set to their posterior medians. Boxplots in dark grey are the posterior predictive for $\log(S^{*new})$ and h^{*new} obtained with latitudes 46, 52, 59, and 63°N (source: Parent and Rivot, 2012, Chapter 9).

3.2.4 Modelling correlation between biological parameters

A sufficient amount of data to derive informative prior distributions is not always available from the stock of interest or conspecific stocks. In such situations, it may be useful to consider data from other stocks of the same species and/or related species to reduce our uncertainty. In the case of life history parameters (growth, length–weight, etc.), correlations between parameters might be expected *a priori* e.g. because of life history characteristics and trade-offs. Explicitly modelling these correlations in a hierarchical model can allow more effective use of the available information or more knowledge with the same data (Pulkkinen *et al.,* 2011). In Pulkkinen *et al.* (2011), correlations between length–weight and length–fecundity relationships were utilized to estimate posterior predictive distributions for the length–weight and length–fecundity parameters of a data-poor stock of round sardinella (*Sardinella aurita*). Information about the correlations between parameters was drawn from relatively data-rich Atlantic herring and European pilchard (*Sardina pilchardus*) stocks. Figure 3.6 illustrates how uncertainty is reduced for the mean parameters of length–weight (μ_{a_w}, μ_{b_w}) and length–fecundity relationships (μ_{a_f}, μ_{b_f}) when information about correlations is utilized.
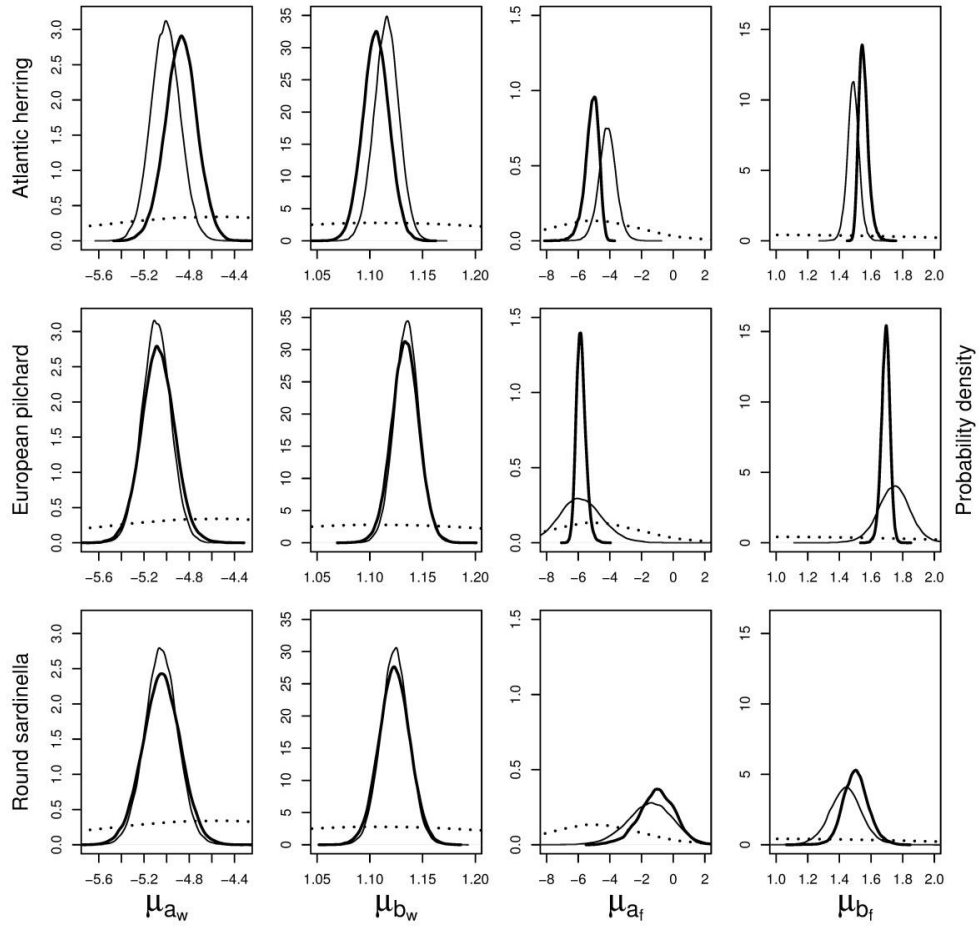
Figure 3.6. Posterior distributions for mean parameters of length–weight $\left(\mu_{a_w}, \mu_{b_w}\right)$ and length–fecundity relationships $\left(\mu_{a_f}, \mu_{b_f}\right)$ for Atlantic herring, European pilchard, and round sardinella species. Bold lines illustrate the posterior distribution from a model with correlation; thin lines show the corresponding distribution from a model without correlations. Dotted lines illustrate prior distributions, those being the same in both models.

Utilization of correlations is based on a computationally convenient assumption that the biological parameters follow a common multivariate normal distribution. If the parameters of interest cannot take negative values, a multivariate normal distribution should be assumed instead for the log-transformed parameters.

Thus, observed values for stock-specific biological parameters are considered to follow a common multivariate normal distribution:

$$\theta_k \sim MVN(\mu_k, \Sigma) \tag{20}$$

where θ_k is the vector of biological parameters for stock k and vector μ_k contains species-specific means for all parameters. If μ_k are considered exchangeable among species, a hierarchical prior distribution can be given for the expected values of the biological parameters. Information from other species (except the ones included in the dataset of the current analysis) can be utilized when setting up the hyperpriors for μ_ks.

As in Pulkkinen *et al.* (2011), an inverse-Wishart distribution can be used as a prior distribution for the variance-covariance matrix

$$\Sigma^{-1} \sim Wishart\,(\Omega, d) \tag{21}$$

where Ω is a positive definite-scale matrix and d is the degrees of freedom parameter.

Even though an inverse-Wishart distribution is a computationally convenient choice to ensure positive definiteness of the variance–covariance matrix, its usage can be problematic. For example, the inverse-Wishart distribution does not allow prior distributions to be set up separately for the marginal variances and covariances of different biological parameters (this is desirable as information about the plausible values of these parameters often exists). There are, however, alternative options for parameterizing the variance–covariance matrix, many of which utilize Cholesky decomposition. More information can be found, for example, from Pinheiro and Bates (1996) or Lu and Ades (2009).

3.3 Getting priors from FishBase

Rainer Froese and Konstantinos Stergiou

Stock assessment, especially for data-limited stocks, depends on information derived from the literature (Section 2.2) as well as from literature compilers (i.e. databases) such as FishBase (Section 2.3). This section explains how to obtain priors for length–weight, growth, and mortality from FishBase. Given that only part of the existing literature is incorporated into databases (Section 2.2), it will always be useful to undertake a review of all available literature on the target stocks and, after incorporating it into FishBase (or SeaLifeBase), to proceed with obtaining priors from FishBase. One important issue to consider here is the credibility of existing data. The quality of the data can be examined, for example, by checking and evaluating the sample size and frequency, the sampling gear used for collecting the data, the representativeness of lengths sampled, and the method used for estimation, depending on the parameter of concern (e.g. Froese *et al.*, 2011; Tsikliras *et al.*, 2013).

3.3.1.1 Length–weight

Froese *et al.* (2014) present a six-step procedure describing how related information can be summarized in a hierarchical Bayesian process for establishing priors for the Bayesian analysis of new length–weight data. The length–weight equation has the form:

$$W = a\, L^b \tag{22}$$

where W and L are variables representing body weight and length, respectively, and a and b are parameters. A Bayesian estimation of these two parameters may proceed in the following six steps:

1) Textbook knowledge is translated into broad overall priors. Exponent b is known to be normally distributed around 3, with a typical range of 2.5–3.5. This is translated into a prior distribution with mean = 3.0 and s.d. = 0.5. Parameter a is known to be log-normally distributed, ranging from $a = 0.001$ in eel-like fish to $a = 0.1$ in spherical fish, if length is in cm and weight is in g. This general information is translated into an overall prior log-normal distribution with mean $\log_{10}(a) = -2$ and s.d. = 1.

2) These overall prior distributions are updated with the parameters from over 5000 length–weight relations (LWR) studies compiled in FishBase.

3) The posterior across-all-species distributions from step 2 are updated for different typical body shape groups of fish, such as eel-like, elongated, fusiform, and short and deep. The posterior distributions obtained from this exercise can

be used directly as preliminary LWR parameters for species for which the body shape is known, but no LWR exists.

4) Existing LWR estimates for a given species are combined with the respective body-shape priors to get posterior distributions for this species.

5) This step is like the fourth step, but applies to species with fewer than five specific LWR estimates. In this case, the body-shape priors are updated with the few LWR studies for the species plus LWR studies from closely related species, such as congeners or family members.

6) The appropriate posterior distributions from steps 3–5 are used to inform the analysis of new length–weight data.

Steps 1–5 are implemented in FishBase and are re-run whenever new LWR data are entered. Users can see the respective posterior distributions in the species summary page in the section "Estimates of some properties based on models". An example for Atlantic herring is shown in Figure 3.7.
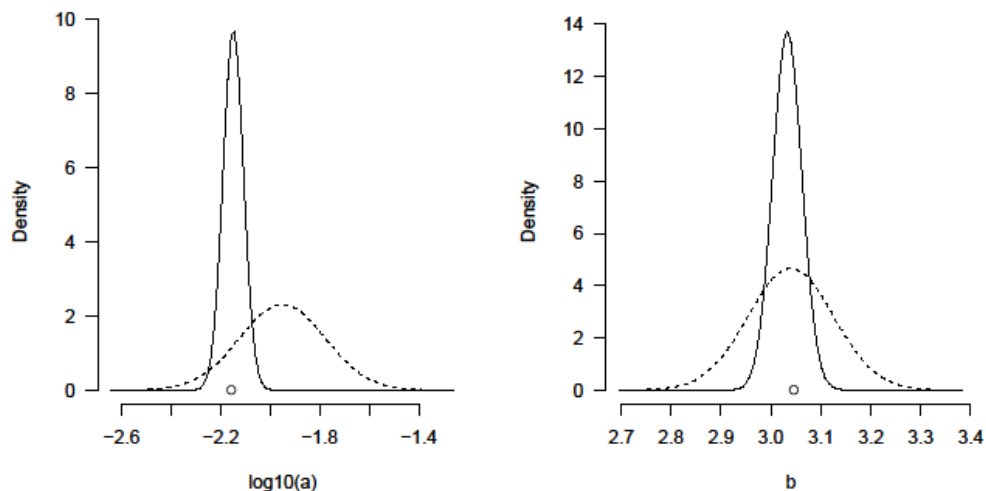
Figure 2.7. Example of posterior distribution of length–weight parameters *a* and *b* for herring (*Clupea harengus*).

Alternatively, FishBase allows the user to select and prioritize existing length–weight studies and perform a new online Bayesian analysis according to the steps outlined above. These Bayesian tools are available from the footer of the length–weight table in FishBase.

3.3.1.2 Growth parameters

A publication and an online web tool in FishBase similar to the one for length–weight relationships are in preparation for growth (R. Froese, pers. comm.). The von Bertalanffy growth equation for body length is:

$$L_t = L_\infty\left(1 - e^{-K\,(t - t_0)}\right) \tag{23}$$

where L_t is the body length at age t, L_∞ is a parameter representing asymptotic length, K is a parameter describing the exponential decay of the difference between L_t and L_∞, and t_0 is a parameter accounting for the fact that fish at birth already have a body length.

A four-step process is used to derive and update priors for the parameters L_∞ and K:

1) General knowledge about the growth of fish is used to obtain a first crude prior for L_∞ as being equal to the known maximum length for the species in question, with mean $L_\infty/L_{max} = 1$ and s.d. = 0.5. L_∞ and K are known to be negatively correlated, i.e. populations with high L_∞ have low K, and vice versa.

   Based on the growth parameters in Pauly (1978), the largest compilation of its kind covering 1501 "stocks" for 515 fish species, Pauly (1979) concluded that the values of $\log_{10}(K)$ vs. $\log_{10}(L_\infty)$ cluster in ellipsoid patterns, whose major axes had a mean slope of –2 . Thus, he suggested that a preliminary estimate of K could be obtained for an unstudied stock for which an estimate of L_∞ was available (see above) together with the growth parameters of another stock of the same species, by assuming that Φ' is a species-specific constant defined as $\Phi' = \log_{10}(K) + 2 \log_{10}(L_\infty)$ (Pauly, 1979), where Φ' is the intercept in a linear regression of $\log_{10}(K)$ over $\log_{10}(L_\infty)$ with a fixed slope of –2. When several sets of growth parameters for a given species are available, a mean value of Φ' can be calculated and used to infer K from L_∞ for an unstudied stock, along with its range of uncertainty.

2) A Bayesian multiple regression is performed with $\log(K)$ as a function of $\log(L_\infty)$, environmental temperature, and habitat use, with the priors from step 1. This results in narrower distributions for K for species within a certain climate zone and habitat.

3) The priors from step 2 are updated with growth studies for the species in question.

4) As with the length–weight relationship, the posterior parameter distributions from this hierarchical analysis will be made available in FishBase in the species summary page in the section "Estimates of some properties based on models" (R. Froese, pers. comm.). Meanwhile for species with many available growth studies, users can perform a Bayesian analysis estimating mean values of $\log(K)$, $\log(L_\infty)$, and the slope of the covariance between $\log(K)$ and $\log(L_\infty)$ with the priors derived in step 1. The posterior distributions can then be used as priors in the analysis of new length-at-age data.

3.3.1.3 Natural mortality

Reasonable estimates of M of adults are rare, and stock assessments have often relied on empirical equations. In FishBase, estimates of M are presented as part of the Growth Table. When opening a respective growth record, the method used for estimating M is clearly indicated, as well as an assessment by FishBase staff whether the estimate was considered doubtful, together with a short comment. The reference from which the mortality estimate was taken is indicated. Clearly, given the importance of M in stock assessments, the user should carefully consider all available information about every published M estimate before accepting it for the analysis. Gislason *et al.* (2010) has published a set of useful criteria for that purpose. For the purpose of deriving a preliminary prior distribution for M, taking the median of published estimates with a CV of ±0.5 seems a reasonable option. If no or few published estimates are available for the species under consideration, M estimates from close relatives with similar maximum length and from the same climate zone may be included in the median. In addition, the empirical equations by Pauly (1980), Hoenig (1983), and Gislason *et al.* (2010) can be used to introduce information from life history traits such as growth and maximum age that

are highly correlated with M. The geometric mean and s.d. across all these estimates may then serve as a useful prior summarizing the existing information on *M*.

3.4   Eliciting expert opinion

Rebecca Whitlock, Polina Levontin, Adrian Leach, Juho Kopra, Charis Apostolidis, and John Mumford

3.4.1   Introduction

"Elicitation" refers to a set of methods designed to capture an expert's knowledge and uncertainty about some unknown parameter(s) or process. Representation of expert judgments of uncertainty as probability distributions (following a Bayesian interpretation of probability as a degree of belief) has been advocated as the most appropriate way to capture expert knowledge (O'Hagan, 1998). This is also the form needed for prior distributions in Bayesian analyses. The Bayesian approach is thus naturally suited to the incorporation of expert judgments about parameters of interest or hypotheses about model structure (O'Hagan *et al.,* 2006; Kuhnert *et al.,* 2010).

Although expert elicitation methods are most commonly used to learn about model parameters, expert opinion enters the modelling process (albeit informally) at a number of stages, including problem definition and model development; this applies to Bayesian and non-Bayesian approaches. Expert opinion plays a fundamental role in the construction of models, in the assumptions that underlie model structure, and the choice between multiple ways to represent a particular process, given those assumptions. As a result, the model itself can be thought of as a prior (i.e. one hypothesis of many alternatives about model structure). Structural uncertainty may be of key importance in ecological problems; explicit recognition and documentation of the role of expert judgement in informing model structure, selection of the set of candidate models, and quantification of prior model probabilities has been recommended (Krueger *et al.,* 2012).

In this chapter, we discuss the practical aspects of eliciting expert opinion, providing examples from the ECOKNOWS project in which expert knowledge was elicited about model parameters and model structure. Problem definition and structuring are dealt with first, followed by elicitation techniques and examples, and finally how to combine information from multiple experts.

3.4.2   The elicitation process

3.4.2.1   Defining and structuring the problem

In order to elicit the desired information from experts, it is important that all experts involved share a common understanding of the quantity to be elicited. O'Hagan (1998) recommends that the elicitor should "ask about quantities that the experts understand best, and in language that is as simple and familiar as possible". Getting the question right is one of the most difficult parts of elicitation (Kuhnert *et al.,* 2010). Linguistic uncertainty (uncertainty that arises because words have imprecise or unclear meanings) can contribute significantly to the uncertainty associated with expert opinion and/or lead to unwanted biases in results (Kuhnert *et al.,* 2010).

For complex problems, the requirement for conceptual clarity may suggest some structuring so that the problem can be presented in terms of quantities with which experts feel most comfortable (and that are thus easier to elicit). Structuring (or elaboration) is also a valuable method for multivariate elicitation (O'Hagan, 1998). In principle, a full

joint probability distribution is needed for two or more uncertain quantities. However, this can be difficult to elicit in practice, so restructuring to reformulate the problem in terms of independent quantities is common (note that in this context, independence means that learning about one of the quantities would not alter the expert's judgment about the other) (O'Hagan, 1998).

In the context of fish stock assessment, we may wish to learn about complex parameters that subsume several different processes and/or attributes, e.g. catchability of a fishing gear, or potential productivity of a particular habitat. An example of structuring can be found in Uusitalo *et al.* (2005), who structured elicitation of smolt production capacity for salmon rivers in the Baltic Sea into eight subquestions relating to the physical and biological attributes of rivers and the likely response of juvenile salmon to those attributes. They suggested that structuring can isolate parts of the problem that are most uncertain and/or controversial. Identification and separate elicitation of the main sources of uncertainty that contribute to overall uncertainty has also been recommended to reduce overconfidence (see Section 2.4.3) (O'Hagan, 1998). This is based on the idea that separate consideration of possible sources of uncertainty will lead the expert to admit more reasons why extreme values might arise, and a correspondingly greater uncertainty about the elicited quantity. Finally, highlighting subcomponents of a problem for experts can help them see the overall picture, as well as the individual parts.

3.4.2.2 Elicitation techniques

Methods for eliciting expert opinion include interviews, workshops, and questionnaires, all designed to ensure that experts give consistent responses (McBride and Burgman, 2012). Elicitation by face-to-face interview with the expert involves a substantial investment of human resources, but may be more cost-effective than a questionnaire in the case of a single expert (O'Hagan *et al.*, 2006) and can increase the expert's motivation and sense of responsibility (Knol *et al.*, 2010). Remote elicitation methods (e.g. questionnaires) are often lower in cost than face-to-face interviews, but this may be countered by low response rates and potential bias as a result (Page *et al.*, 2012). Overall, in-person interviews are preferred for the reasons above and because they provide more opportunities for feedback and interaction than questionnaires or telephone interviews. An emphasis for making the reasoning behind expert judgments transparent and ensuring that judgments are given on the basis of all relevant information is common to all methods (Spiegelhalter *et al.*, 2011).

Elicitation of expert opinion with individuals allows for more interaction between facilitator and expert. Benefits of this include a greater opportunity for explanation to ensure the expert properly understands the subject of elicitation, the possibility for more targeted questioning, and the chance for more detailed feedback (Knol *et al.*, 2010). The possibility of the elicitation being influenced by the ideas of the facilitator could be greater in interactions with a single expert than with a group (Ayyub, 2001).

Group elicitation (or behavioural aggregation) methods allow some interactions between a group of experts, with the aim of forming a consensus, so that a single distribution is elicited from the group. Diverse views can be accommodated in group elicitation where the group's uncertainty may reflect differing opinions by participants with different knowledge, experience, or interpretations. Group elicitation techniques can involve bringing experts together in one place, e.g. focus groups and consensus-building workshops with an unbiased facilitator (Nuseibeh and Easterbrook, 2000), or involve more controlled interaction (e.g. the Delphi method). In the Delphi method, each expert first provides their own judgment, together with a rationale. Each expert's

judgment is then made available to all other experts in the group, who are then given the chance to revise their own assessment. This process continues until a consensus is reached or the facilitator decides to combine the individual inputs. Some advantages and disadvantages of group elicitation are as follows (Janis, 1971; Esser, 1998; Ayyub, 2001):

Advantages:

    a) Group discussion allows redundant information to be discounted through sharing of information.

    b) Group elicitations can be more efficient when dealing with participants with different backgrounds by making disciplinary biases more explicit.

    c) Group elicitation can be seen by participants as more equitable, giving equal voice for all participants, especially when the results can be assessed directly by the entire group.

    d) Group elicitation allows a more natural interaction between participants than formal interview.

Disadvantages:

    a) Groups may be dominated by one or few individuals, while some individuals may be underrepresented.

    b) Disagreement may arise over the structure of the problem, making it difficult to proceed with an informative discussion on components of the problem.

    c) Conflicting views might not be resolved during the allotted time, demanding further planning and follow-up meetings.

    d) There is potential for group-reinforced bias when members share a common background. The tendency towards overconfidence may be greater in groups than in individuals.

As a result of the potential biases arising from group interaction, the role of an impartial facilitator in managing the elicitation and ensuring that all group members have an opportunity to contribute is the key to successful elicitation.

3.4.2.3 What to elicit?

While the probability distribution that best describes an expert's views is, in reality, comprised of a large (usually infinite) number of probabilities, in practice, a satisfactory representation of the expert's beliefs can be obtained using a small number of individual probabilities. Measures of location (such as mean and median) and scale (standard deviation, variance) or probabilities (quantiles) represent alternative summaries of expert distributions. Generally, summaries based on probabilities are most widely used, owing to potential biases in elicited location and scale parameters (e.g. assessments of the mean tend to be biased towards the median for skewed distributions (Peterson and Miller, 1964). Other studies suggest that people are poor at interpreting the meaning of variance and assigning numerical values to it (O'Hagan *et al.*, 2006).

Probabilities can be elicited by specifying values x (for the quantity to be elicited) at which the expert's probabilities $p(X \leq x)$ are requested. Alternatively, the elicitor may opt to elicit quantiles of the expert's distribution (i.e. the expert is asked for the value x for a specified probability p, such that $p(X \leq x) = p$. The most commonly used quantile is the median, $p = 0.50$ (O'Hagan *et al.*, 2006). The median can be used as the starting point for sequentially bisecting the expert's probability to elicit a number of quantiles (O'Hagan *et al.*, 2006), whereby each question takes the form of asking the expert

to determine a value for x such that X is equally likely to be less than or greater than this point (Figure 3.8).

An additional method for eliciting quantiles is the credible interval method. For probability q, this involves eliciting two values for x such that $p(x_1 \leq X \leq x_2) = q$, where the range is a central interval, i.e. it is equally likely that X is lower than x_1 or higher than x_2. This approach can also be used to obtain a credible interval, whereby the probability q corresponding to $p(x_1 \leq X \leq x_2) = q$ is elicited for given x_1 and x_2.

In order to use quantile-based elicitation methods, the elicitor needs to decide on the probabilities p or credible interval coverage probability q to use. While the median and central 50% intervals are the most commonly used, some studies have indicated that assessment of the 33rd and 67th percentiles can reduce overconfidence (e.g. Barclay and Peterson, 1973).

Once the elicitor has obtained summaries of the expert's distribution, a parametric distribution is usually fitted to represent the expert's opinion (Figure 3.8). The raw output of the elicitation may also be used, e.g. in the form of a histogram; however, it has been argued that a smooth density function provides a more realistic representation of an expert's opinion (e.g. O'Hagan, 1998). The choice of distribution may be affected by the range of values that a quantity can take (e.g. a beta distribution is a natural choice for a proportion taking values between 0 and 1, while a log-normal or gamma distribution is suitable for a quantity that can only take positive values). Where such a preference does not exist, the choice of distribution can be guided by the implied shape of the expert's probability distribution. For example, the choice between normal and log-normal distributions may be made on the basis of the distances between the expert's prior estimate (e.g. median or mode) and his/her lower and upper quantiles, so that if these distances are approximately equal, a normal distribution would be implied, while a greater distance between the prior estimate and upper quantile would imply a log-normal distribution (e.g. O'Hagan, 1998).
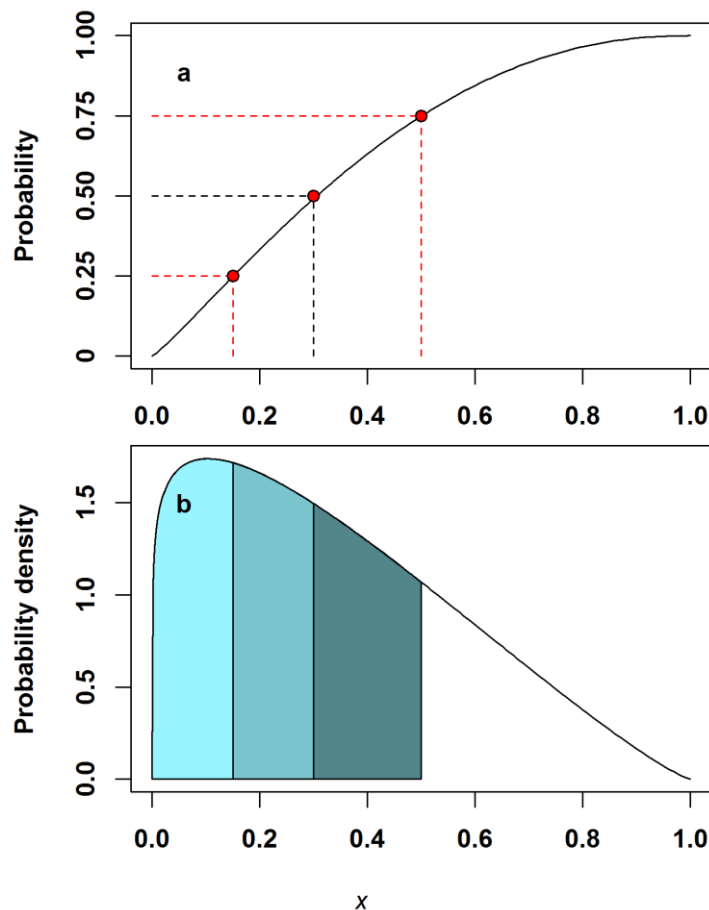
Figure 3.8. Fitted Beta(1.1, 2.2) distribution for the 0.25, 0.50, and 0.75 quantiles (red points) of a distribution elicited from an expert. (a) Cumulative distribution function; b) probability density function. The median ($p(X \leq x)$ = 0.50) bisects the probability such that X is equally likely to be lower or higher than the median value. The regions below and above the median are then each bisected to obtain the lower and upper quartiles (0.25 and 0.75). In the example, the expert believes that it is equally likely that X is < or >0.30. Supposing that X is > the expert's median (0.30), he/she believes that it is equally likely that X is < or >0.50.

3.4.2.4 Dealing with inaccuracy and imprecision

As noted in Section 2.4.3, distributions fitted to represent expert knowledge will be subject to inaccuracy and imprecision (because it is difficult for experts to give precise numerical values for their probabilities and because only a few probability judgments are usually elicited). Two commonly-used methods to address the second problem are feedback and overfitting. Feedback involves showing the expert some summary or implication of their fitted distribution (e.g. an additional quantile of the distribution or a visual summary) to allow them to confirm whether the fitted distribution is an accurate representation of their beliefs (examples using visual feedback can be found in Section 3.4.3). Feedback should be accompanied by an opportunity for experts to revise their summaries (e.g. probabilities or quantiles) or request a more flexible probability model if they feel that their beliefs have not been accurately captured. Overfitting involves eliciting more summaries than are strictly needed to fit an expert's distribution. For example, to fit a beta distribution, the elicitor might elicit three or more quantiles rather than the sufficient two. It is unlikely that a single beta distribution will fit the expert's quantiles perfectly; in practice, there may be several beta distributions that fit equally

well. The elicitor can then use some method of minimizing the difference between the expert's quantiles and a fitted beta distribution, although the final distribution selected will depend on the metric used for minimization (sum of squares, sum of percentage differences, etc.). An example of overfitting is provided in Figure 3.9. Overfitting has been advocated as preferable to feeding back a quantile to the expert because it avoids the need to ask a question that may suffer from problems of anchoring (i.e. the expert's response is likely to be anchored to the value that the elicitor feeds back to him/her) (O'Hagan *et al.*, 2006). However, eliciting a large number of quantiles is a complex task that is more demanding of the elicitor's time.

An additional approach is to formally quantify the uncertainty about the expert's distribution by updating the analyst's prior beliefs about the expert's density function with "observations" (elicited summaries of the expert's distribution) obtained during elicitation, using Bayes theorem (O'Hagan and Oakley, 2004). The remaining uncertainty about the expert's distribution is then represented through the analyst's posterior distribution.

Density of fitted distribution
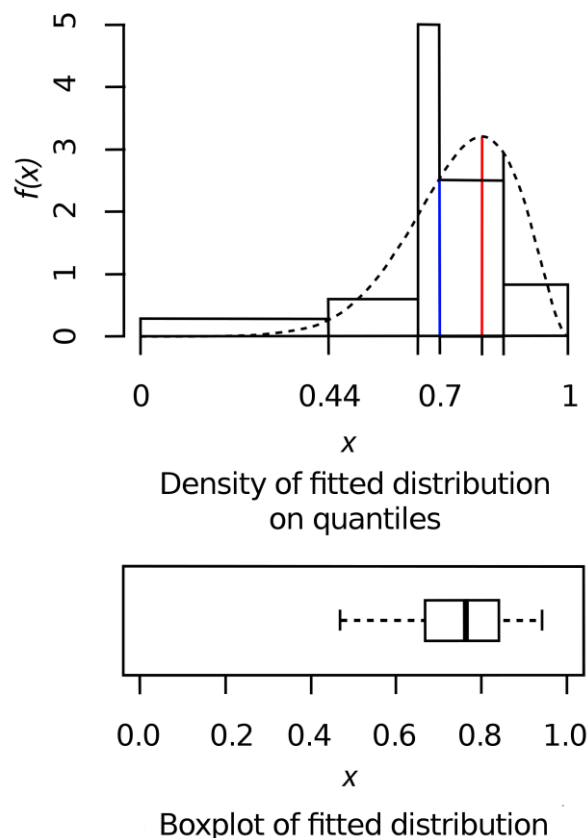on quantiles

Boxplot of fitted distribution

Figure 3.9. Example of overfitting. Five quantiles (12.5, 25, 50, 75, and 87.5%) were elicited (shown by the histogram) and a beta distribution was fitted using weighted sums of squares (dashed line). The median of the fitted distribution is indicated by the vertical blue line, and the mode is indicated by the vertical red line (top panel). Boxplot describing the 95% credible interval and quartiles (ends of the box) of the fitted distribution (lower panel). This figure was immediately shown to the expert, who was offered the chance to revise one or more of the quantiles if the fitted distribution did not correspond to his/her beliefs.

Sensitivity analysis can be carried out to explore the effect of alternative fitted distributions on the problem at hand (recognizing that there is uncertainty in the elicited distribution). This may be particularly relevant in a Bayesian context when the analyst

wishes to evaluate the influence of the elicited prior distribution on the posterior. Global sensitivity analysis (Berger, 1994) builds on this idea to evaluate the effect of a class of prior distributions in terms of the range of posterior distributions that can be obtained.

3.4.3 Expert elicitation tools with examples

3.4.3.1 ECOKNOWS elicitor

The ECOKNOWS elicitor is an Excel-based method for guiding experts through the process of expressing their beliefs about variables of interest, such as stock–recruitment dynamics, natural mortality, maturation (Figure 3.10). Its main advantage is providing an immediate visual feedback of elicited distributions enabling experts to iteratively arrive at a distribution that most closely reflects their knowledge and beliefs.

Life history and ecological elicitation tool

Three distributions are available to represent experts' beliefs: normal, log-normal, and beta. For normal and log-normal variables, experts are invited to choose to work with standard parameterizations such as the median and *CV* (for log-normal) or mean and standard deviation (for normal), or alternatively to specify two quantiles of their choice: e.g. the 0.25 and 0.75 quantiles. The beta distribution is parameterized using a method that requires the expert to propose mode and "mode weight" parameters which determine how much probability mass is centred on the mode.
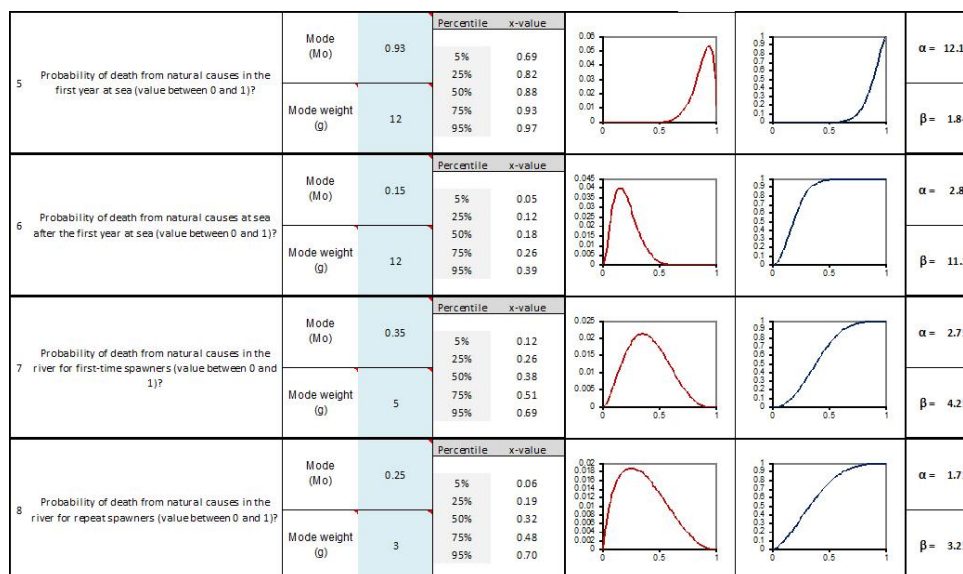
Figure 3.10. Screenshot from the ECOKNOWS elicitation tool used to quantify expert knowledge about sea trout (*Salmo trutta*) stocks in Rivers Isojoki and Lestijoki in Finland.

Stock–recruitment elicitation tool

The ECOKNOWS stock–recruitment tool provides a specific elicitation platform for learning about the relationship between the reproductive potential of a fish (or wildlife) population and recruitment of new individuals. The Beverton–Holt and Ricker models are currently implemented, although the elicited quantities could readily be used to parameterize other models (Pulkkinen and Mäntyniemi, 2013). Stock–recruitment models are parameterized in terms of a density-independent parameter α or a that is proportional to fecundity and a density-dependent parameter β or b that is proportional to both fecundity and density-dependent mortality (Quinn and Deriso, 1999). The Beverton–Holt model is then:

$$R = \frac{\alpha S}{1 + \beta S} \quad (24)$$

and the Ricker model:

$$R = aSe^{-bS} \quad (25)$$

where R is the number of recruits, S is a measure of spawning potential (e.g. spawning–stock biomass or fecundity), α and a are slopes of recruits per unit of e.g. spawning–stock biomass at the origin (density-independent parameters), and β and b are parameters that control the degree of density-dependence. The stock–recruitment elicitation tool requires input of probability distributions for the maximum lifetime reproductive rate, spawner biomass per recruit, and either maximum recruitment or recruitment at the unfished demographic equilibrium. Equations can be found in Annex 2. Input distributions for any of the quantities may be empirically based (e.g. information about spawner biomass per recruit). The ECOKNOWS stock–recruitment tool elicits uncertainty about the functional form of the stock–recruitment relationship by asking experts about the most likely mechanism for density-dependence in the population of interest (Figure 3.11). Visual feedback is provided in the form of probability distributions for derived parameters (e.g. stock–recruit steepness) and spawner–recruit plots for elicited stock–recruitment functions.
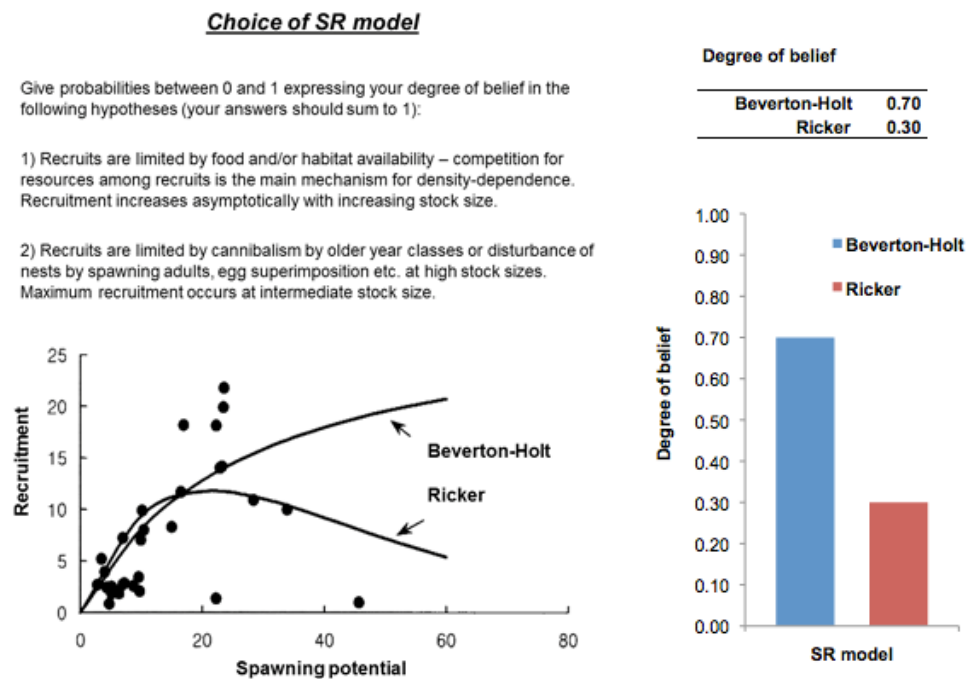

Figure 3.11. Screenshot from the ECOKNOWS stock–recruitment elicitation tool model choice interface.

The life history and ecological elicitation tool is available at http://www.ecoknows.eu/Results/LifeHistoryElicitationTool.php.

3.4.3.2 ECOKNOWS selectivity elicitor tool

The ECOKNOWS selectivity elicitor is an R tool for extracting expert opinion about the selectivity of fishing gears. Three functional forms for selectivity curves are currently implemented: normal, logistic, and double normal. Selectivity can be expressed as a function of size or age.

Selectivity elicitation involves simulating the size distribution of the true population using estimates or priors for growth parameters and age-specific rates of total mortality. The simulated size distribution is plotted, can be shown to the experts, and iteratively adjusted (e.g. via the total mortality rate) to arrive at a distribution the expert believes to be representative of the actual size structure. Experts are then asked for their beliefs about the functional form of selectivity for the gear in question (i.e. logistic, normal or double normal) and for summaries of their distributions for lengths corresponding to two selectivity probabilities, e.g. 50 and 90% selectivity. The expert's answers are used to obtain parametric distributions for the parameters of the selectivity curve [length at 50% selectivity (log-normal distribution) and steepness (log-normal distribution) for the logistic curve, mean (normal distribution) and variance (inverse gamma distribution) for the normal curve, and minimum and maximum lengths at full selectivity (both normal distributions) plus left- and right-hand-side variances (both inverse gamma distributions) for the double normal curve]. The tool produces a plot of the range of selectivity curves implied by the expert's values, along with the expected size distribution in the catch and of missed (unselected) fish. Experts can then be given an opportunity to revise their values on the basis of the visual feedback.

Methods used to obtain priors for selectivity curve parameters using summaries of an expert's distribution for length at a particular selectivity value are described below for the logistic functional form and in Annex 3 for the normal and double normal functional forms. The equations employed to derive priors using experts' inputs to the R selectivity elicitor tool are described in Annex 4 for normal and logistic selectivity.

<u>Logistic selectivity</u>

The logistic distribution is parameterized by the length at 50% selectivity $L50$, and the steepness of the curve υ. Log-normal priors are specified for $L50$ and υ using the elicited 0.10, 0.50 (median), and 0.90 quantiles for the length at which selectivity equals 0.50 plus the 0.10, 0.50 (median), and 0.90 quantiles for the length at which selectivity equals 0.90 (Figure 3.12).
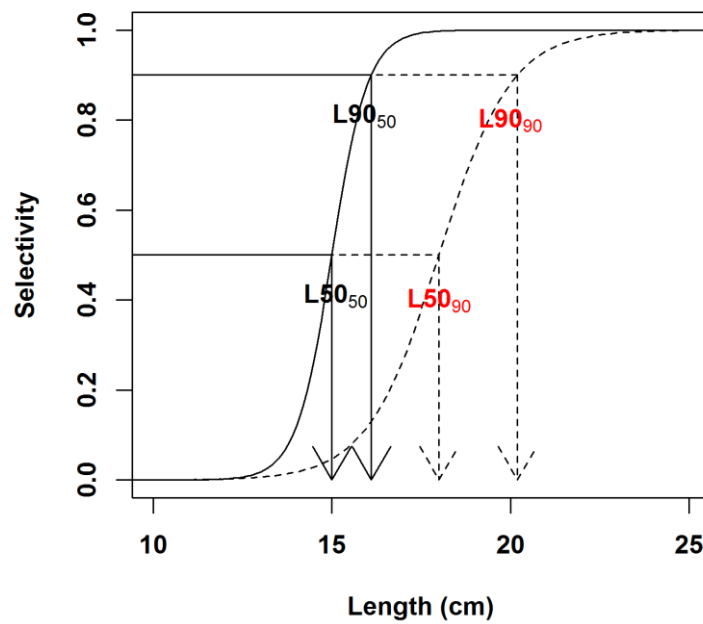
Figure 3.12. An example of a logistic selectivity curve where the most likely (modal) length at 50% selectivity ($\hat{L}50$) is 15 cm, and the modal length at 90% selectivity ($\hat{L}90$) is 16 cm. The 0.90 quantile for $L50$ ($L50_{90}$) is 18 cm and the 0.90 quantile for $L90$ ($L90_{90}$) is 20 cm.

Selectivity elicitor tool example

To illustrate the selectivity elicitor tool, suppose an expert provides 0.50 and 0.90 quantiles for length at 50 and 90% selectivity that are the same as those in Figure 3.12, with 0.10 quantiles for length at 50 and 90% selectivity of 12 cm and 15 cm, respectively. The resulting selectivity curve is shown in Figure 3.13, based on the fitted priors for $L50$ and v of $L50$~log-normal(2.71,0.14) and v~log-normal(0.79,1.26). Numbers in parentheses are mean and standard deviation of $\log(x)$.

Figure 3.13. Example elicited logistic selectivity curve. Panel (a) shows the median selectivity curve (red) plus randomly sampled selectivity curves from the fitted prior distributions for $L50$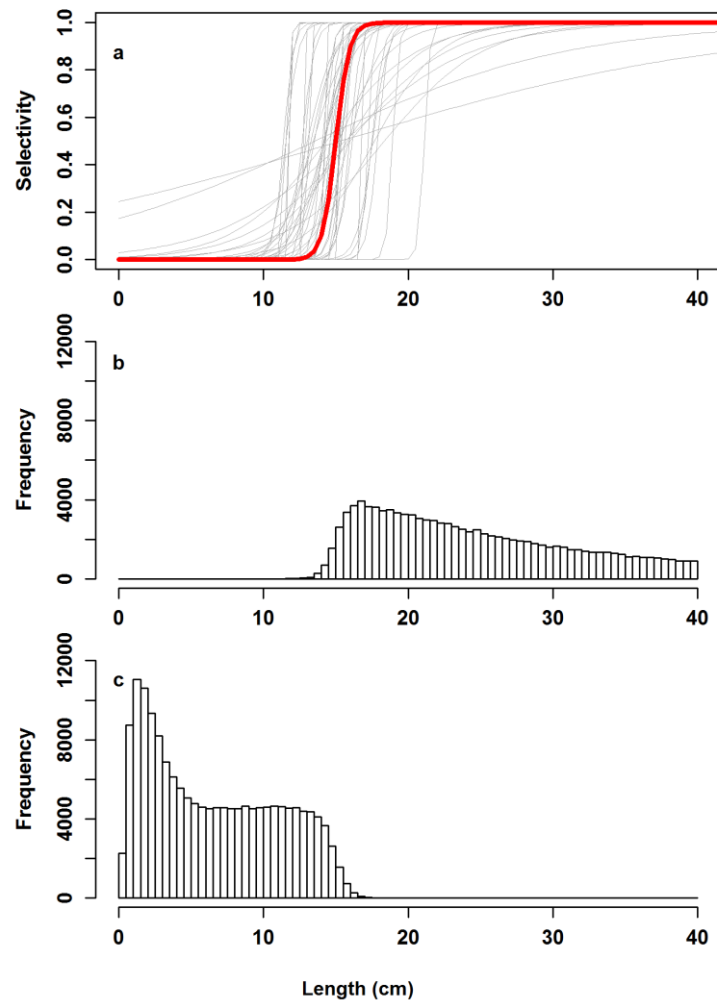 and v. Panel (b) shows the length frequency distribution of fish that are retained in the catch, given the median selectivity curve and simulated length distribution. Panel (c) shows the length frequency distribution of fish that are not retained in the catch, given the median selectivity curve and simulated length distribution.

3.4.4 Combining the opinions of multiple experts (mathematical aggregation)

When judgments are elicited from more than one expert individually, it is usually convenient to combine them to yield one prior distribution for use in analyses. In this section, we discuss a selection of some of the most common methods for combining multiple expert distributions (mathematical aggregation). Reviews of different methodologies available to aggregate expert opinions can be found in Clemen and Winkler (1999), Burgman (2005), and O'Hagan *et al.* (2006).

Weighting expert opinions

The mathematical aggregation methods considered here share a first step of determining and expressing numerically how much weight to place on the opinion of each expert. A common choice is to consider the opinions of all experts as equally valuable and assign them equal weights. However, methods that use performance-based weights have been shown to outperform methods that assume equal weights (Goossens *et al.,* 1998). Typically, performance-based weighting methods take into account an expert's calibration as well as his/her informativeness. Calibration measures how

closely an expert's opinion corresponds to reality, while informativeness is a measure of the information contained in an expert's opinion and is related to the precision of the distribution. Cooke (1991) proposed the use of seeding variables (quantities chosen from the same subject area as the quantity to be elicited), whose true values are known to the facilitator, but not to the experts. Quantiles for the seeding variables are elicited from experts; an expert would be down-weighted for being poorly calibrated or giving a wide distribution that provides little information about the quantity of interest. The weight awarded to each expert is then a combined measure of his/her calibration and information scores.

Aspinall (2010) describes an application of Cooke's method (1991) in which expert risk assessors were given a series of seed questions to evaluate their proficiency in predicting time to failure in old earthen dams in the UK. The experts' responses (best estimate and 90% credible interval) were weighted according to their success in answering the seed questions. Experts who provided longer-time-to-failure estimates with wider credible intervals did better on seed questions, earning greater performance weights (Figure 3.14).
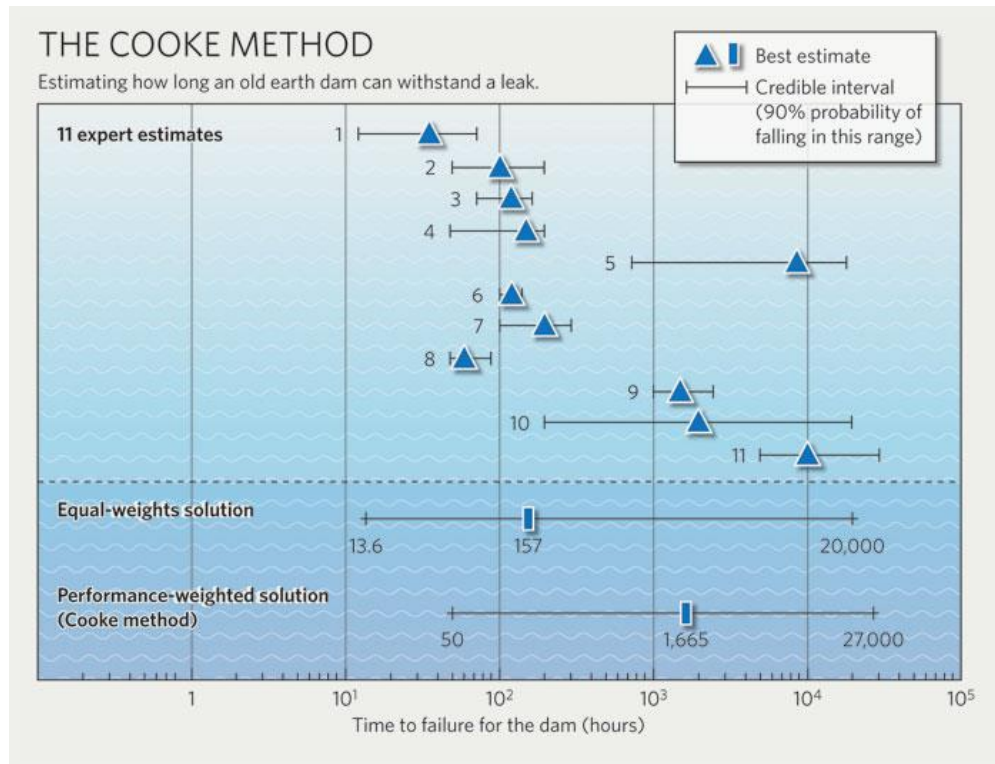
Figure 3.14. Estimates from 11 experts of the time-to-failure of an earthen dam after the core begins to leak. The performance-weighted best judgment is about 70 days — much longer than the equal-weights solution of ca. one week. Reprinted by permission of Macmillan Publishers Ltd: Nature (Aspinall, W. 2010. A route to more tractable expert advice. Nature, 463: 294–295), copyright (2010).

Linear and logarithmic opinion pools

Opinion pooling involves combining the distributions of individual experts to form a consensus distribution. The simplest method for achieving this is the linear opinion pool, is a linear weighted average of the individual distributions with weights w_i summing to 1:

$$f(\theta) = \sum_{i=1}^{n} w_i f_i(\theta) \tag{26}$$

where $f_i(\theta)$ is the distribution of the ith expert. In the case of equal weights w_i, this is a simple average of the individual distributions. The logarithmic opinion pool is based on a weighted geometric mean of the individual distributions (i.e. the weights are exponents):

$$f(\theta) = k \prod_{i=1}^{n} f_i(\theta)^{w_i} \tag{27}$$

where k is a normalizing constant which ensures that the combined distribution $f(\theta)$ integrates to 1 and that the weights w_i have the same interpretation as above. Linear and logarithmic opinion pools are appealing because of their simplicity, but they have some drawbacks. The linear opinion pool is not externally Bayesian (i.e. updating individual distributions with a new observation before combining, combining individual distributions, and then updating the combined distribution do not always yield the same result) (Madansky, 1964). The logarithmic opinion pool is externally Bayesian, but does not satisfy the marginalization requirement (i.e. summing the probabilities of two mutually exclusive events $p(A)$ and $p(B)$ elicited from several experts to obtain the probability of the event "A or B" does not always yield the same result before and after pooling the experts' opinions) (O'Hagan *et al.*, 2006). Logarithmic pooling tends to result in a narrower consensus distribution than linear pooling, implying stronger information than that given by any individual expert, whereas the linear pooling consensus distribution implies less knowledge than any of the individual distributions (O'Hagan *et al.*, 2006). It is also worth noting that the consensus distribution formed from the logarithmic opinion pool treats as implausible values of θ that are considered implausible by any one expert.

Bayesian methods

An alternative way of combining the distributions of different experts is to treat them as alternative hypotheses about the true state of nature and merge them using Bayesian model averaging (BMA) (Carlin and Chib, 1995; Punt and Hilborn, 1997; Hoeting *et al.*, 1999; Mäntyniemi *et al.*, 2012). BMA is a consistent method for taking structural uncertainty into account in a Bayesian framework (Hoeting *et al.*, 1999). The analyst assigns a prior probability $p(M_i)$ or weight to each of E experts ($M = 1, 2, \ldots, E$) expressing their beliefs about the extent to which expert i's assessment $p(\theta|M_i)$ corresponds to the true state of nature (i.e. the probability that expert i is right), before observing any data. A common choice is to set $p(M_i) = 1/E$. Since M is a discrete variable, the prior knowledge of the analyst about θ can be found using the weighted average of the beliefs of different experts (see Mäntyniemi *et al.*, 2012):

$$p(\theta) = \sum_{i=1}^{E} p(\theta|M_i)p(M_i) \tag{28}$$

This approach thus belongs to the class of linear pooling methods, but differs in the conceptual interpretation of the weights as the (analyst's) probability that an individual expert is correct. The weights of different experts in BMA can be updated using Bayes theorem as soon as new observations become available, so that the posterior weights will reflect the experts' performance in predicting the observations y (the posterior weights are obtained using the marginal likelihood of the data given M's beliefs, $p(y|M_i)$). Since BMA allows updating of experts' weights (rather than keeping them fixed) when new evidence is obtained, BMA adheres to the externally Bayesian criterion. A combined distribution obtained using BMA is shown in Figure 3.15. JAGS code for this example can be found in Annex 5.
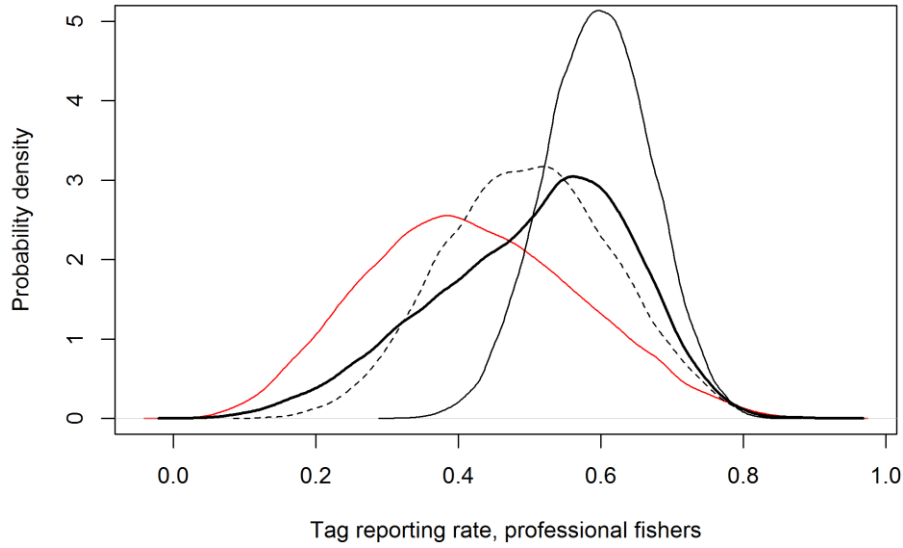
Figure 3.15. Prior probability distributions from three experts (solid red line, dashed black line, and solid non-bolded black line). The prior distribution obtained using Bayesian model averaging (BMA) is shown by the thick black line.

In Bayesian decision theory, it is the decision maker's beliefs rather than the experts' that are ultimately required, so the correct procedure for combining expert opinions is for the decision maker to update his/her own prior distribution with data in the form of expert opinions (O'Hagan *et al.*, 2006). To illustrate this with a simple example, point estimates of an unknown quantity X from different experts ($x_1, x_2, \dots x_e$) could be treated as observations and used to update the decision maker's prior for $p(x)$, yielding a posterior distribution for X (Winkler, 1981; Mosleh and Apostolakis, 1986):

$$p(x \mid x_1, x_2, \dots x_e) \propto p(x_1, x_2, \dots x_e \mid x) \, p(x) \,. \tag{29}$$

The likelihood terms from different experts $p(x_i|x)$ represent the likelihood of expert i giving an estimate of x_i, given that the true value is x. They are formulated to reflect the decision maker's appraisal of expert i's competence; two models, additive or multiplicative errors, have been suggested by Mosleh and Apostolakis (1986). Under the additive error model, expert i's assessment of x is modelled as the sum of two terms:

$$x_i = x + \varepsilon_i \tag{30}$$

where ε_i is an additive error term that follows a normal distribution with mean μ_i and standard deviation σ_i chosen to reflect the decision maker's beliefs about i's bias and precision. If the opinions of different experts are assumed to be independent, the likelihood terms for different experts can simply be multiplied together; otherwise, the decision maker must provide his/her beliefs about the dependence of the assessments (the degree to which the experts' errors are interrelated).

More likely, the problem will involve combining distributions (rather than point estimates) from many experts. In this case, a similar approach can be used where the decision maker updates his/her prior with the distributions of many experts ($\theta_1, \theta_2, \dots \theta_e$), and the notions of bias, precision, and dependence relate to how well an expert's judgments are calibrated and precise, and the tendency of experts to report similar proba-

bilities (Clemen and Winkler, 1999). Examples can be found in Lindley and Singpurwalla (1986) and Clemen and Winkler (1999). Hierarchical Bayesian approaches (Section 3.2) have also been applied (e.g. Lipscomb *et al.*, 1998); these allow the biases of different experts to be treated analogously to random effects and to introduce some dependence between the probabilities of different experts. However, a side-effect of the increased model complexity is the challenge of specifying an increasing number of priors for uncertain parameters.

3.5 Bayesian belief networks

Sakari Kuikka

The idea behind Bayesian belief networks (BBN) is fairly simple. They mimic the human inference; the human mind connects variables by means of logic and in the inference more weight is given to better knowledge. If something is unknown (no information), it should not be reflected in other parts of the model. The knowledge is "collected" from the different parts of the model.

This mimicry is easy to explain; belief networks have been developed in the field of artificial intelligence (AI) research where one of the aims is to describe human logic (see Pearl, 1988 for a detailed text on belief networks). This close relationship between human thinking and belief networks can also be seen in the elicitation process of expert knowledge. It appears to be relatively easy for experts to include their logic and other knowledge in the belief network modification of Varis (1994), where a single link value is used to connect two variables instead of full conditional probability tables. The articles of Rowe *et al.* (1988), Rowe and Watkins (1992), and Shafer and Pearl (1990) all stress the need to model the logic required in expert judgments.

Punt and Hilborn (1997) point out that one of the advantages of Bayesian assessment is that it offers the possibility of including some of the knowledge of non-modelling experts in the scientific analysis. When the non-experts see that their knowledge has also been used in the analysis, their commitment to scientific advice may improve. The role of knowledge obtained from different sources can be tested separately and their impact on the conclusions evaluated. .

BBN models can be used in expert elicitations. For example, Uusitalo *et al.* (2005) used BBN models to describe the probabilistic dependencies of the maximum salmon smolt production of different Baltic Sea rivers. This knowledge has been used as prior information in the modelling of S/R relationships of Baltic Sea salmon stocks. In Bayesian decision analysis and in belief network modelling, both the parameters and the structure of the entire model are uncertain. The amount (quality) of knowledge is described by probability distributions, and the model is used to capture uncertainty in the reasoning process. Clemen (1989) and Dagum *et al.* (1995) discuss the different aspects of combining information in forecasting and underline the importance of the Bayesian paradigm in forecasting.

A short popularized description of BBN models is given in Kuikka and Varis (1996) and Varis and Kuikka (1997). Almond (1995), Shafer (1996), van der Graag (1996), and Darwiche (2009) provide clear presentations on the use of belief networks. In Bayesian belief networks, the relationships between the variables can be presented either by conditional probabilities (used in influence diagrams, i.e. models that also include a decision node) or by links between variables (Pearl, 1988). In the modification of Varis (1994), the prior probabilities for each variable are given first. After this, links are estimated or assessed by the expert in order to create a structure for the model (= input of

causal knowledge). Links enable the calculus of the posterior probabilities, i.e. the knowledge in the other parts of the model, to be taken into account by the information flow through the links and other variables.

The present application area of BBN is wide (see e.g. Jensen, 1996). They are mostly applied to assess the state of a system by using direct observations, prior knowledge, and causal dependencies. They are very effective compared with rule-based systems (if x, then y), especially if there are high uncertainties in the relationships between the variables (see e.g. Chong and Walley, 1996, for a test on wastewater treatment). Another important use of belief networks is in the field of decision analysis, where modifications of belief networks, i.e. influence diagrams, are used to analyze decision problems. The use of influence diagrams is close to the use of decision trees, but in complicated problems, the corresponding graphical representation is more easily understood. Therefore, it is easier for an expert or other stakeholders to assess the suggested model structure.

The aim of Bayesian decision analysis is to model the structure of human reasoning needed for problem solving. The model includes two or more possible alternatives. There is a need to model both the use of knowledge and the controllability (actions), and the causal relationships of the models should be relevant in describing the causal dependencies, from the perspective of the person constructing the model. This is an essential difference with the prevailing modelling practice in ecology or fisheries, where the aim is usually to model only the physical system itself.

BBN models can include the following types of variables:

   a) probabilistic nodes: probabilistic variables, including conditional probabilities or prior probabilities;
   b) deterministic nodes: including arithmetic functions;
   c) decision nodes: control variables, values selected by decision makers;
   d) expected value/utility nodes: objective variables, nodes to be maximized or minimized.

If the model is used only to assess the state of the system, it can include only probabilistic nodes. Moreover, the types of variables can also be algorithm- and software-specific. For example, Hugin software (http://www.hugin.dk/) is based on the algorithm presented in Lauritzen and Spiegelhalter (1988). This software allows the observation of new information for any of the probabilistic variables to be modeled. After observing one value, the beliefs on the state of the other variables are updated on the basis of the new information. In this algorithm, the information obtained can also go "against" the direction of conditional probabilities shown by arrows. Computation is based on so-called "junction trees" (e.g. Jensen, 1996). Lam and Bacchus (1994) and Cooper and Herskovits (1992) discuss the different aspects of estimation in belief-network modelling. Cooper and Herskovist (1992) discuss direct learning of Bayesian belief-network models from data, which is one possibility if the datasets are sufficiently large.

In influence diagrams, where the focus is on decision analysis, uncertainty is described by conditional probabilities relating the flow of information in the inference. Arcs show the direction of the information flow from predecessor node (parents) to successor node (child) (Jensen, 1996). Models include at least one decision node and a node describing the objectives. Objective functions (in some cases called utility variables) are either minimized or maximized by the decision nodes. The ability to control the system is described by the model, i.e. how effectively the changes in decision variables are reflected in the probability distributions of interest variables. If the degree of uncertainty in the chain between the decision and interest nodes is high (wide distributions

in conditional probabilities), the controllability is poor, and it is difficult to adjust the system to a desired state by means of the decision variables.

BBN are at their best in modelling complex phenomena, where the models and data for different parts of the problem can be of very different quality and where expert judgments are needed. Rowe *et al.* (1988) discuss the management use of artificial intelligence in more detail. The call of Stephenson and Lane (1995) for interdisciplinary methodology is fulfilled by belief networks; conditional probabilities facilitate the use of, for example, socio-economic information.

Burns and Clemen (1993) discussed the interesting relationship between covariance structure models and Gaussian influence diagrams (conditional distributions are normal). They stressed that these tools are closely related. Both tools are based on multivariate normality. Burns *et al.* (1993) applied a similar approach to the modelling of public responses to risk signals in the media. This approach seems to be an effective tool for modelling large datasets and might offer good possibilities for modelling complicated ecological or fisheries systems.

The Markov Chain Monte Carlo (MCMC) estimation procedures (Besag *et al.*, 1995) are also used in the estimation of probabilities for belief networks. For example, Gibbs sampling (see Besag *et al.*, 1995) is used in the estimation of probabilities in mixed graphical models, where some of the variables are continuous and some are discrete.

In decision analysis, the robustness of the methodology is essential; models should not be too sensitive to, for example, assumptions. Pradhan *et al.* (1996) investigated the sensitivity of belief networks and concluded that they are robust tools and are not very sensitive to imprecise input probabilities. Experiences reported in Kuikka and Varis (1996) support this conclusion. Even though the structure varied, the models gave relatively similar results. This finding suggests that belief networks can also be used as an effective tool for modelling the knowledge of non-mathematical experts.

The elicitation process linked to expert knowledge (Kuikka and Varis, 1996; O'Hagan, 2012) is relatively straightforward if the total number of arcs is not very large. Large numbers of parent nodes and their associated arcs lead, in turn, to large and unmanageable conditional probability tables. As a rule of thumb, if the link value (Kuikka and Varis, 1996) is <0.1 (the parent node explains <10% of the child node), there is no reason to make the model more complex by adding links.

4 Summary: impacts and interpretation

Sakari Kuikka, Samu Mäntyniemi, and Etienne Rivot

This report has demonstrated and discussed the logic of the Bayesian analysis, with particular focus on the role of prior information in stock assessment and has demonstrated the collection of prior knowledge from primary data, literature, databases, and experts.

The Bayesian analysis is a mathematical formalization of the sequential learning process. Learning is one of the key aims of scientific activities. It is natural that new publications take into account earlier publications and discuss the results in light of earlier scientific results. In spite of this, it is still relatively uncommon to see publications that take former results fully into account in a formal, mathematical way. In papers, literature is usually only discussed without the results incorporated quantitatively by applying classical statistics, while databases have been used to estimate parameters for which there are no case-specific data. The role of expert knowledge is not formally taken into account in classical statistics, even though it is generally accepted that expert knowledge can be used effectively in the decision-making process to collect certain types of data and in the selection of methods to analyse the data. One can say that the effectiveness of scientific learning may be enhanced by taking account of expert experience in a formal analytical process.

Publication bias can be a severe problem when collecting information from published results to build informative priors. For example, null hypothesis statistical testing (NHST) remains one of the most commonly used methods for testing a hypothesis when analyzing ecological data. P-values are one of the main statistical criteria used for NHST because of their apparent simplicity of interpretation; low p-values demonstrate that the observed data are unlikely, given the null hypothesis. In practice, a low p-value is an essential criterion that will facilitate publication of any particular work. Although they are as much informative as results with low p-values, results with high p-values ("negative results") are less likely to be published because they are less likely to be submitted by authors or accepted by editors. Then, literature does not keep track of "negative results" as it should, and there is a high risk that our world view becomes biased. This may not be a problem in cases where raw data are published for stock assessment purposes, but evaluation of the aim and purpose of the papers needs to be carried out before relying on the sampled literature set.

In addition to the quality of the knowledge available, one must also consider the costs of obtaining new information. Collection of raw data is definitely more expensive than the use of databases or published papers. Moreover, one may regard it the moral duty of a scientist to exploit all information to decrease uncertainties and provide the most realistic estimates possible to decision-makers and fishers.

In the context of a precautionary approach to fishery management, using all available sources of information to reduce uncertainty in parameter estimates and predictions may have impacts in decision-making. The precautionary approach to fishery management means taking a risk-averse attitude in decision-making; i.e. the variance of the predictions plays a role, in addition to the expected value, in selecting the optimal decision alternative. In some cases, it is preferable to accept a decision that implies a smaller expected value than another decision, if the risk for total failure is smaller. In fisheries, impaired recruitment is a good example of such a risk, the probability of

which should be low. The degree of risk aversion may depend on the consequences in each case.

Online databases offer interesting possibilities. For example, current techniques enable an analysis where the impact of new data on the overall estimates of the databases can easily be assessed. If, at the same time, precautionary management is applied, the person providing new data may see the quantitative impact of the new data, e.g. increased catches while maintaining the same level of risk as earlier. The reward from investing in more data is the improvement in either increased yield or reduced risk, or ideally both, that science can provide to society.

It is often difficult to find an expert who has not seen the case-specific dataset before constructing a prior distribution. In this sense, there is a theoretical risk that the same knowledge will be used twice (i.e. to inform the prior and then as part of the data used in the analysis). On the other hand, as the stock assessment results are used to evaluate future outcomes, an expert may be able to identify new uncertainties to include in the evaluations, which have not yet been seen in the data. The impacts of climate change or other environmental disturbances offer good examples of possible sources of uncertainty. Good knowledge of the literature and theoretical education may provide sources of information, e.g. on causal processes, which do not exist in historical datasets. Therefore, it may be useful to use experts at least in model selection.

The use of hierarchical models, and especially the role of correlations, offers effective ways to exploit such features of the data, which usually cannot be obtained from expert elicitation. Even relatively small datasets offer possibilities for learning about correlations. Some of these correlations may be similar for many species, offering the possibility of learning from other species or other populations of the same species. Similarity between species may be due to biology or to the environment where populations live.

5 Future research needs

Samu Mäntyniemi and Etienne Rivot

The use of prior information in Bayesian stock assessment faces both philosophical and practical challenges. This manual has studied and proposed techniques that can be used to deal with many practical challenges. All of these techniques can be further studied and fine-tuned, and new developments in interactive, database, and mobile communication technology can potentially be explored to find novel ways of forming prior distributions from different sources of information.

However, the philosophical issues and their sociological consequences have not been dealt with here, although they represent key issues for future research. For example, the fact that a model structure is also a statement of prior knowledge and provides an important component for the joint prior distribution of all variables is not widely acknowledged. Instead, model structures and likelihood functions are often thought of as somehow representing the objective part of the analysis, with the research focused on the prior distributions of the model parameters: how to make the priors contain as little information as possible or, as examined in this manual, how to condense the existing information into a prior distribution of a parameter.

It is well known that when the amount of observed data increases, the weight of the prior distributions of model parameters in the resulting inference decreases. This is often termed "data overriding the prior", which overlooks the fact that the assumptions made about the model structure and the observation processes that define the likelihood function determine the interpretation of the observations. This logically means that the prior information used to construct the model becomes more influential when the amount of data increases.

Future research efforts should also clearly be targeted at the philosophy and practices of model building based on existing prior information. Stock assessments that explicitly account for model uncertainty are still rare, and improving the methodology in this direction is an important avenue for future research. Bayesian model averaging methods have been developed, but are still difficult to implement for complex models with many parameters. Assigning *a priori* the same probabilities for the competing model structures may be justified in some cases. However, one easily forgets the fact that in the process of selecting alternative models, one already uses prior information in a crude way (selection/rejection); this could be elaborated further by using formal, transparent, and consistent methods to assign informative priors for the alternative models.

Non-parametric models do not require strong assumptions about the model structure (e.g. non-parametric stock–recruitment models avoid constraining the analysis to classical Ricker and Beverton–Holt forms) and can naturally cope with model uncertainty. More research is needed to make Bayesian analysis of non-parametric models more accessible in practice.

Perhaps one of the key obstacles to the systematic and widespread use of prior information lies in the attitudes of both scientists and the users of the scientific information. Scientific knowledge, such as stock assessment results, is often expected to be independent of the people involved in the production of the results. As a result, scientists are often hesitant to use their expertise to the fullest extent and are more willing to "let the data speak".

However, new approaches to science are emerging. The importance of taking stakeholder perspectives into account when dealing with highly uncertain and ambiguous

environmental problems is increasingly being stressed. Submitting a model to the critical review of an extended peer group enables assessing and potentially improving the credibility of the model. Moreover, building a model in a participatory process with stakeholders provides the possibility of integrating as many perspectives as is considered relevant in model building, thereby avoiding a one-sided interpretation. It can be assumed that a model built in a participatory process can increase the legitimacy of the results.

The conventional juxtaposition related to subjective and objective science needs reanalysis. Since Bayesian stock assessment models (like all other assessment models) are made from existing human knowledge, prior distributions for parameters and model structures may play a key role in the process of reviewing and collectively building those models with stakeholders. Research on theories and practices of such processes will be needed in the future.

6   Acknowledgements

7 References

Almond, R. G. 1995. Graphical Belief Modeling. Chapman & Hall, London. 427 pp.

Archambault, B., Le Pape, O., Bousquet, N., and Rivot, E. 2014. Density-dependence can be revealed by modelling the variance in the stock–recruitment process: an application to flatfish. ICES Journal of Marine Science, 71: 2127–2140.

Alpert, M., and Raiffa, H. 1982. A progress report on the training of probability assessors. *In* Judgment Under Uncertainty: Heuristics and Biases, pp. 294–305. Ed. by D. Kahneman, P. Slovic, and A. Tversky. Cambridge University Press, Cambridge. 544 pp.

Apostolidis, C. A., and Stergiou, K. I. 2008. A preliminary compilation of life-history data for Mediterranean marine invertebrates. *In* Von Bertalanffy Growth Parameters of Non–fish Marine Organisms, pp. 92–111. Ed. by M. L. D. Palomares, and D. Pauly. Fisheries Centre Research Reports 16(10). Fisheries Centre, University of British Columbia.

Aspinall, W. 2010. A route to more tractable expert advice. Nature, 463: 294–295.

Ayyub, B. M. 2001. Elicitation of Expert Opinions for Uncertainty and Risks. CRC Press, Boca Raton, FL. 328 pp.

Barclay, S., and Peterson, C. R. 1973. Two methods for assessing probability distributions. Technical Report 73-1, Decisions and Designs, McLean, VA.

Bedford, T., and Cooke, R. 2001. Probabilistic Risk Analysis: Foundations and Methods. Cambridge University Press, Cambridge. 414 pp.

Berger, J. O. 1985. Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York. 618 pp.

Berger, J. O. 1994. An overview of robust Bayesian analysis (with discussion). Test, 3: 5–124.

Berger, J. O., and Wolpert, R. L. 1988. The Likelihood Principle: A Review, Generalizations, and Statistical Implications. The Institute of Mathematical Statistics, Haywood, CA. 208 pp.

Bernardo, J. M. 1979. Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society, Series B, 41: 113–147.

Besag, J., Green, P., Hidgon, D., and Mengersen, K. 1995. Bayesian computation and stochastic systems. Statistical Science, 10: 3–66.

Beverton, R. J., and Holt, S. J. 1957. On the dynamics of exploited fish populations. Ministry of Agriculture, Fisheries and Food. Fishery Investigations, London, Series II, XIX. 533 pp.

Booker, J. M., and McNamara, L. A. 2004. Solving black box computation problems using expert knowledge theory and methods. Reliability Engineering and System Safety, 85(1–3): 331–340.

Box, G. E. P., and Tiao, G. C. 1973. Bayesian Inference in Statistical Analysis. John Wiley & Sons, New York. 588 pp.

Browman, H. I. (Ed). 1999. The uncertain position, status and impact of negative results in marine ecology: philosophical and practical considerations. Marine Ecology Progress Series, 191: 301–309.

Burgman, M. A. 2005. Risks and Decisions for Conservation and Environmental Management. Cambridge University Press, Cambridge. 504 pp.

Burns, W. J., and Clemen, R. 1993. Covariance structure models and influence diagrams. Management Science, 39: 816–834.

Burns, W. J., Slovic, P., Kasperson, R. E., Kasperson, J. X., Renn, O., and Emani, S. 1993. Incorporating structural models into research on the social amplifications of risk: implications for theory construction and decision making. Risk Analysis, 13: 611–623.

Carlin, B. P., and Chib, S. 1995. Bayesian model choice via markov-chain monte-carlo methods. Journal of the Royal Statistical Society, Series B (Methodological), 57(3): 473–484.

Chong, H. G., and Walley, W. J. 1996. Rule-based versus probabilistic approaches to the diagnosis of faults in wastewater treatment processes. Artificial Intelligence in Engineering, 1: 265–273.

Clark, J. S. 2003. Uncertainty and variability in demography and population growth: a hierarchical approach. Ecology, 84(6): 1370–1381.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5: 559–583.

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

Cooke, R. M. 1991. Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press, Oxford. 336 pp.

Cooper, G. F., and Herskowitz, E. 1992. A Bayesian method for the induction of probabilistic networks from data. Machine Learning, 9: 309–347.

Dagum, P., Galper, A., Horwitz, E., and Seiver, A. 1995. Uncertain reasoning and forecasting. International Journal of Forecasting, 11: 73–87.

Darwiche, A. 2009. Modeling and Reasoning with Bayesian Network. Cambridge University Press, Cambridge. 548 pp.

Edwards, W. 1982. Conservatism in human information processing. *In* Judgment Under Uncertainty: Heuristics and Biases, pp. 359–369. Ed. by D. Kahneman, P. Slovic, and A. Tversky. Cambridge University Press, Cambridge. 544 pp.

Esser, J. K. 1998. Alive and well after 25 years: a review of groupthink research. Organizational Behavior and Human Decision Processes, 73(2–3): 116–141.

Fazey, I., Fazey, J. A., Salisbury, J. G., Lindenmayer, D. B., and Dovers, S. 2006. The nature and role of experiential knowledge for environmental conservation. Environmental Conservation, 33(1): 1–10.

Fischhoff, B., Slovic, P., and Lichtenstein, S. 1978. Fault trees: Sensitivity of estimated failure probabilities to problem representation. Journal of Experimental Psychology: Human Perception and Performance, 4: 330–344.

Fletcher, W. J. 2005. The application of qualitative risk assessment methodology to prioritize issues for fisheries management. ICES Journal of Marine Science, 62: 1576–1587.

Froese, R., and Pauly, D. (Eds). 2014. FishBase. World Wide Web electronic publication. www.fishbase.org.

Froese, R., and Proelss, A. 2010. Rebuilding fish stocks no later than 2015: will Europe meet the deadline? Fish and Fisheries, 11: 194–202.

Froese, R., and Sampang, A. 2013. Potential indicators and reference points for good environmental status of commercially exploited marine fishes and invertebrates in the German EEZ. World Wide Web electronic publication, available from http://oceanrep.geomar.de/22079/.

Froese, R., Thorson, J. T., and Reyes, R. B. Jr. 2014. A Bayesian approach for estimating length-weight relationships in fishes. Journal of Applied Ichthyology, 30(1): 78–85.

Froese, R., Tsikliras, A. C., and Stergiou, K. I. 2011. Editorial note on weight–length relations of fishes. Acta Ichthyologica et Piscatoria, 41: 261–263.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. Bayesian Data Analysis, 2nd edn. Chapman & Hall/CRC Press, New York. 696 pp.

Gilovich, T., Griffin, D., and Kahneman, D. (Eds). 2002. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, New York. 882 pp.

Gislason, H., Daan, N., Rice, J., and Pope, J. 2010. Size, growth, temperature and the natural mortality of marine fish. Fish and Fisheries, 11: 149–158.

Goossens, L. H. J., Cooke, R. M., and Kraan, B. C. P. 1998. Evaluation of weighting schemes for expert judgment studies. *In* Probabilistic Safety Assessment and Management, Probabilistic Safety Assessment and Management: Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management (PSAM), 13-18 September 1998, New York City, USA, pp. 1937–1942. Ed. by A. Mosleh, and R. A. Bari. Springer Verlag, New York. 2933 pp.

Gould, S. J. 1993. Cordelia's dilemma. Natural History, 2/93: 10–18.

Harley, S. J., and Myers, R. A. 2001. Hierarchical Bayesian models of length-specific catchability of research trawl surveys. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1569–1584.

Hillary, R. M., Levontin, P., Kuikka, S., Mäntyniemi, S., Mosqueira, I., and Kell, L. 2012. Multi-level stock–recruit analysis: beyond steepness and into model uncertainty. Ecological Modelling, 242: 69–80.

Hoenig, J. M. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin US, 82: 898–903.

Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. 1999. Bayesian model averaging: a tutorial. Statistical Science, 14: 382–417.

Hora, S. C., and von Winterfeldt, D. 1997. Nuclear waste and future societies: A look into the deep future. Technological Forecasting and Social Change, 56: 155–170.

Howson, C., and Urbach, P. 1993. Scientific Reasoning: The Bayesian Approach, 2nd edn. Open Court Publishing Company, Peru, IL. 470 pp.

ICES. 2012a. Report of the ICES Advisory Committee 2012. ICES Advice 2012, Book 2. 114 pp.

ICES. 2012b. Report of the ICES Advisory Committee 2012. ICES Advice 2012, Book 3. 82 pp.

ICES. 2013. Report of the ICES Advisory Committee 2013. ICES Advice 2013. Book 1. 348 pp.

ICES. 2014. Report of the Baltic Salmon and Trout Assessment Working Group (WGBAST), 26 March–2 April 2014, Aarhus, Denmark. ICES Document CM 2014/ACOM: 08. 347 pp.

Janis, I. L. 1971. Groupthink. Psychology Today, 5: 43–46.

Jeffreys, H. 1961. Theory of Probability, 3rd edn. Clarendon Press, Oxford. 447 pp.

Jensen, F. V. 1996. An Introduction to Bayesian Networks. UCL Press, London. 178 pp.

Jiao, Y., Hayes, C., and Cortés, E. 2009. Hierarchical Bayesian approach for population dynamics modelling of fish complexes without species-specific data. ICES Journal of Marine Science, 66(2): 367–377.

Johnson, D. H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management, 63: 763–772.

Jones, S. K., Jones, T. K., and Frisch, D. 1995. Biases of probability assessment: A comparison of frequency and single-case judgments. Organizational Behavior and Human Decision Processes, 61: 109–122.

Kéry, M., and Schaub, M. 2012. Bayesian Population Analysis using WinBUGS: A Hierarchical Perspective. Academic Press, Waltham, MA. 535 pp.

Knol, A. B., Slottje, P., van der Sluijs, J. P., and Lebret, E., 2010. The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environmental Health, 9(19): 9–19.

Krueger, T., Page, T., Hubacek, K., Smith, L., and Hiscock, K. 2012. The role of expert opinion in environmental modelling. Environmental Modelling and Software, 36: 4–18.

Kuhnert, P. M., Martin, T. G., and Griffiths, S. P. 2010. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecology Letters, 13: 900–914.

Kuikka, S., and Varis, O. 1996. Uncertainties of climatic change impacts in Finnish watersheds: a Bayesian network analysis of expert knowledge. Boreal Environmental Research, 2: 109–128.

Lam, W., and Bacchus, F. 1994. Learning Bayesian belief networks, an approach based on the MDL principle. Computational Intelligence, 10: 269–293.

Lauritzen, S. L., and Spiegelhalter, D. J. 1988. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society, Series B (Methodological), 50(2): 157–224.

Lichtenstein, S., and Fischhoff, B. 1980. Training for calibration. Organizational Behavior and Human Performance, 26: 149–171.

Lichtenstein, S., Fischhoff, B., and Phillips, L. D. 1982. Calibration of probabilities: The state of the art to 1980. *In* Judgment Under Uncertainty: Heuristics and Biases, pp. 306–334. Ed. by D. Kahneman, P. Slovic, and A. Tversky. Cambridge University Press, Cambridge. 544 pp.

Lindley, D. V., and Singpurwalla, N. D. 1986. Reliability (and fault tree) analysis using expert opinions. Journal of the American Statistical Association, 81: 87–90.

Lipscomb, J., Parmigiani, G., and Hasselblad, V. 1998. Combining expert judgment by hierarchical modeling: An application to physician staffing. Management Science, 44: 149–161.

Lu, G., and Ades, A. 2009. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics, 10(4): 792–805.

Mackinson, S., Wilson, D. C., Galiay, P., and Deas, B. 2011. Engaging stakeholders in fisheries and marine research. Marine Policy, 35: 18–24.

Madansky, A. 1964. Externally Bayesian groups. Technical Report RM-4141-PR, RAND Corporation.

Mäntyniemi, S., Haapasaari, P., Kuikka, S., Parmanne, R., Lehtiniemi, M., and Kaitaranta, J. 2012. Incorporating stakeholders' knowledge to stock assessment: Central Baltic herring. Canadian Journal of Fisheries and Aquatic Sciences, 70(4): 591–599.

Mäntyniemi, S., Kuikka, S., Rahikainen, M., Kell, L.T., and Kaitala, V. 2009. The value of Information in fisheries management: North Sea herring as an example. ICES Journal of Marine Science, 66: 2278–2283.

Mäntyniemi, S., Romakkaniemi, A., and Arjas, E. 2005. Bayesian removal estimation of population size under unequal catchability. Canadian Journal of Fisheries and Aquatic Sciences, 62: 291–300.

Martell, S., and Froese, R. 2013. A simple method for estimating MSY from catch and resilience. Fish and Fisheries, 14: 504–514.

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

Martin, T. G., Burgman, M. A., Fidler, F., Kuhnert, P. M., Low-Choy, S., McBride, M., and Mengersen, K. 2012. Eliciting expert knowledge in conservation science. Conservation Biology, 26: 29–38.

McAllister, M. K., Hill, S. L., Agnew, D. J., Kirkwood, G. P., and Beddington, J. R. 2004. A Bayesian hierarchical formulation of the DeLury stock assessment model for abundance estimation of Falkland Islands' squid (*Loligo gahi*). Canadian Journal of Fisheries and Aquatic Sciences, 61(6): 1048–1059.

McAllister, M. K., Pikitch, E. K., and Babcock, E. A. 2001, Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1871–1890.

McBride, M. F., and Burgman, M. A. 2012. What is expert knowledge, how is such knowledge gathered, and how do we use it to address questions in landscape ecology? *In* Expert Knowledge and Its Application in Landscape Ecology, pp. 11–38. Ed. by A. H. Perera, D. C. Ashton, and C. J. Johnson. Springer-Verlag, New York. 307 pp.

McDonald, A. D., and Smith, A. D. M. 1997. A tutorial on evaluating expected returns from research for fishery management. Natural Resource Modeling, 10: 185–216.

Michielsens, C. G. J., and McAllister, M. K. 2004. A Bayesian hierarchical analysis of stock-recruit data: Quantifying structural parameter uncertainty. Canadian Journal of Fisheries and Aquatic Sciences, 61: 1032–1047.

Michielsens, C. G. J., McAllister, M. K., Kuikka, S., Mäntyniemi, S., Romakkaniemi, A., Pakarinen, T., Karlsson, L., *et al*. 2008. Combining multiple Bayesian data analyses in a sequential framework for quantitative fisheries stock assessment. Canadian Journal of Fisheries and Aquatic Sciences, 65: 962–974.

Mosleh, A., and Aspostolakis, G. 1986. The assessment of probability distributions from expert opinions with an application to seismic fragility curves. Risk Analysis, 6(4): 447–461.

Myers, R. A. 2001. Stock and recruitment: generalization about maximum reproductive rates, density dependence, and variability using a meta-analytic approaches. ICES Journal of Marine Science, 58: 937–951.

Myers, R. A., Bowen, K. G., and Barrowman, N. J. 1999. Maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences, 56: 2404–2419.

Nuseibeh, B., and Easterbrook, S. 2000. Requirements engineering: a roadmap. ICSE 2000 Proceedings of the Conference on the Future of Software Engineering June 4th-11th, 2000 in Limerick, Ireland, pp. 35–46. ACM New York, NY.

O'Hagan, A. 1998. Eliciting expert beliefs in substantial practical applications. The Statistician, 47(1): 21–35.

O'Hagan, A., 2012. Probabilistic uncertainty specification: expert judgement of input uncertainties in a mechanistic model of carbon flux. Environmental Modelling and Software, 36: 35–48.

O'Hagan, A., and Oakley, J. E. 2004. Probability is perfect, but we can't elicit it perfectly. Reliability Engineering and System Safety, 85: 239–248.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T., *et al*. 2006. Uncertain Judgements: Eliciting Experts' Probabilities. Wiley, Chichester, UK. 321 pp.

Oakley, J., and O'Hagan, A. 2010. SHELF: The Sheffield Elicitation Framework (version 2.0). Sheffield, UK: School of Mathematics and Statistics, University of Sheffield, UK, 2010. Available from: http://tonyohagan.co.uk/shelf. (Accessed May 2014).

Oliver, D. M., Fish, R. D., Winter, M., Hodgson, C. J., Heathwaite, A. L., and Chadwick, D. R. 2012. Valuing local knowledge as a source of expert data: farmer engagement and the design of decision support systems. Environmental Modelling and Software, 36: 76–85.

Otway, H., and von Winterfeldt, D. 1992. Expert judgment in risk analysis and management: process, context and pitfalls. Risk Analysis, 12: 83–93.

Page, T., Heathwaite, A. L., Thompson, L. J., Pope, L., and Willows, R., 2012. Eliciting fuzzy distributions from experts for ranking conceptual risk model components. Environmental Modelling and Software, 36: 19–34.

Parent, E., and Rivot, E. 2012. Introduction to Hierarchical Bayesian Modeling for Ecological Data. Chapman and Hall/CRC, Boca Raton, FL. 427 pp.

Pauly, D. 1978. A preliminary compilation of fish length growth parameters. Berichte des Institut für Meereskunde an der Universität Kiel, No. 55. 200 pp.

Pauly, D. 1979. Gill size and temperature as governing factors in fish growth: a generalization of von Bertalanffy's growth formula. Berichte des Institut für Meereskunde an der Universität Kiel, No. 63. 156 pp.

Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters and mean environmental temperature in 175 fish stocks. Journal du Conseil International pour l'Exploration de la Mer, 39: 175–192.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R., and Torres, F. Jr. 1998. Fishing down marine food webs. Science, 279: 860–863.

Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan-Kaufman, San Mateo, CA. 552 pp.

Peterson, C. R., and Miller, A. 1964. Mode, median and mean as optimal strategies. Journal of Experimental Psychology, 68: 363–367.

Pinheiro, J. C., and Bates, D. M. 1996. Unconstrained parametrizations for variance-covariance matrices. Statistics and Computing, 6: 289–296.

Pradhan, M., Henrion, M., Provan, G., Del Favero, B., and Huang, K. 1996. The sensitivity of belief networks to imprecise probabilities: an experimental investigation. Artificial Intelligence, 85: 363–397.

Prévost, E., Parent, E., Crozier, W., Davidson, I., Dumas, J., Gudbergsson, G., Hindar, K., *et al.* 2003. Setting biological reference points for Atlantic salmon stocks: Transfer of information from data-rich to sparse-data situations by Bayesian hierarchical modelling. ICES Journal of Marine Science, 60: 1177–1193.

Pulkkinen, H., and Mäntyniemi, S. 2013. Maximum survival of eggs as the key parameter of stock–recruit meta-analysis: accounting for parameter and structural uncertainty. Canadian Journal of Fisheries and Aquatic Sciences, 70(4): 527–533.

Pulkkinen, H., Mäntyniemi, S., Kuikka, S., and Levontin, P. 2011. More knowledge with the same amount of data: advantage of accounting for parameter correlations in hierarchical meta-analyses. Marine Ecology Progress Series, 443: 29–37.

Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35–63.

Punt, A. E., Smith, D. C., and Smith, A. D. M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the "Robin Hood" approach. ICES Journal of Marine Science, 68(5): 972–981.

Quinn II, T. J., and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, New York. 560 pp.

Raiffa, H., and Schlaifer, R. 1961. Applied Statistical Decision Theory. Clinton Press, Boston. 356 pp.

Ricker, W. E. 1954. Stock and recruitment. Journal of the Fisheries Research Board of Canada, 11: 559–623.

Rivot, E., and Prévost, E. 2002. Hierarchical Bayesian analysis of capture-mark-recapture data. Canadian Journal of Fisheries and Aquatic Sciences, 59: 1768–1784.

Rivot, E., Prévost, E., Cuzol, A., Baglinière, J. L., and Parent, E. 2008. Hierarchial Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. Canadian Journal of Fisheries and Aquatic Sciences, 65: 117–133.

Robert, M., Faraj, A., McAllister, M. K., and Rivot, E. 2010. Bayesian state-space modelling of the De Lury depletion model: strengths and limitations of the method, and application to the Moroccan octopus fishery. ICES Journal of Marine Science, 67(6): 1272–1290.

Rowe, A. J., Somers, I. A., and Schutt, H. 1988. Management use of artificial intelligence. *In* Applied Expert Systems (Studies in Management Science and Systems), pp. 55–70. Ed. by E. Turban, and P. R. Watkins. North Holland Publishing, Amsterdam. 284 pp.

Rowe, A. J., and Watkins. P. R. 1992. Beyond expert systems – reasoning, judgment, and wisdom. Expert Systems with Applications, 4: 1–10.

Royle, J. A., and Dorazio, R. M. 2008. Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities. Academic Press, London. 464 pp.

Shafer, G. 1996. Probabilistic Expert Systems. CBMS-NSF Regional Conference Series in Applied Mathematics, No. 67. Society for Industrial and Applied Mathematics, Philadelphia. 80 pp.

Shafer, G., and Pearl, J. (Eds). 1990. Readings in Uncertain Reasoning. Morgan-Kaufman, San Mateo, CA. 768 pp.

Schnute, J. T., and Kronlund, A. R. 1996. A management oriented approach to stock recruitment analysis. Canadian Journal of Fisheries and Aquatic Sciences, 53: 1281–1293.

Slovic, P. 1972. From Shakespeare to Simon: speculations–and some evidence–about man's ability to process information. Oregon Research Institute Research Bulletin, 12(2): 1–29.

Spiegelhalter, D. J., Thomas, A., and Best, N. G. 1996. Computation on graphical bayesian models. *In* Bayesian Statistics 5: Proceedings of the Fifth Valencia International Meeting, June 5-9, 1994, pp 407–425. Ed. by J. M. Bernardo, J. O. Berger, A. P. David, and A. F. M. Smith. Oxford University Press, Oxford. 832 pp.

Spiegelhalter, D., Pearson, M., and Short, I. 2011. Visualizing uncertainty about the future. Science, 333(6048): 1393–1400.

Stergiou, K. I., Karachle, P. K., Apostolidis, C., Giannakaki, A., Despoti, S., Tsikliras, A., Mäntyniemi, S., *et al*. 2012. ECOKNOWS: using prior knowledge for fisheries management. Proceedings 10th Hellenic Symposium on Oceanography and Fisheries, 10: 180.

Stergiou, K. I., and Karpouzi, V. S. 2002. Feeding habits and trophic levels of Mediterranean fish. Reviews in Fish Biology and Fisheries, 11: 217–254.

Stergiou, K. I., and Moutopoulos, D. K. 2001. A review of length-weight relationships of fishes from Greek marine waters. Naga (ICLARM), 24: 23–39.

Stergiou, K. I., and Tsikliras, A. C. 2006. Underrepresentation of regional ecological research output by bibliometric indices. Ethics in Science and Environmental Politics, 2006: 15–17.

Stephenson, R. L., and Lane, D. E. 1995. Fisheries management science: a plea for conceptual change. Canadian Journal of Fisheries and Aquatic Sciences, 52: 2051–2056.

Su, Z., Peterman, R. M., and Haeseker, S. L. 2005. Spatial hierarchical bayesian models for stock-recruitment analysis of pink salmon (*Onchorhynchus gorbusha*). Canadian Journal of Fisheries and Aquatic Sciences, 62: 2471–2486.

Sutton, A. J., and Abrams, K. R. 2001. Bayesian methods in meta-analysis and evidence synthesis. Statistical Methods in Medical Research, 10(4): 277–303.

Tsikliras, A. C., and Stergiou, K. I. 2014. Size at maturity of Mediterranean marine fishes. Reviews in Fish Biology and Fisheries, 24: 219–268.

Tsikliras, A. C., Stergiou, K. I., and Froese, R. 2013. Editorial on reproductive biology of fishes. Acta Ichthyologica et Piscatoria, 43: 1–5.

Tversky, A., and Kahneman, D. 1971. Belief in the law of small numbers. Psychological Bulletin, 2: 105–110.

Tversky, A., and Kahneman, D. 1973. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 4: 207–232.

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

Tversky, A., and Kahneman, D. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90: 293–315.

Uusitalo, L., Kuikka, S., and Romakkaniemi, A. 2005. Estimation of Atlantic salmon smolt carrying capacity of rivers using expert knowledge. ICES Journal of Marine Science, 62: 708–722.

van der Graag, L. C. 1996. Bayesian belief networks: odds and ends. The Computer Journal, 39: 99–113.

Varis, O. 1994. A belief network approach to modelling of environmental change: the methodology and prospects for application. WP-94-40. International Institute for Applied Systems Analysis. Laxenburg, Austria. 18 pp.

Varis, O., and Kuikka, S. 1997. Joint use of multiple environmental assessment models by a Bayesian meta-model: the Baltic salmon case. Ecological Modeling, 10: 341–351.

Winkler, R. L. 1981. Combining probability distributions from dependent information sources. Management Science, 27: 479–488.

Wong, K. F. E., and Kwong, J. Y. Y. 2000. Is 7300 m equal to 7.3 km? Same semantics but different anchoring effects. Organizational Behavior and Human Decision Processes, 82: 314–333.

Wood, L. E., and Ford, J. M. 1993. Structuring interviews with experts during knowledge elicitation. *In* Knowledge Acquisition as Modeling, Vol. 1, pp. 71–90. Ed. by K. M. Ford, and J. M. Bradshaw. John Wiley & Sons, New York.

Yaniv, I., and Foster, D. P. 1997. Precision and accuracy of judgmental estimation. Journal of Behavioral Decision Making, 10: 21–32.

Annex 1: BUGS/JAGS code for estimation of slope at origin of Beverton–Holt stock–recruit function (Example 3.2.1)

```
model{

 for(i in 1:N_SR){
   # Number of recruits in year i follows logNormal distribution
   Rec[i] ~ dlnorm(MRec[i],tauRec)

   # Expected recruitment follows Beverton-Holt s-r-function
   muRec[i] <-K[stockSR[i]]/((K[stockSR[i]]/alpha[stockSR[i]])+Eggs[i]))*Eggs[i]
   # Expected recruitment in log-scale
   Mrec[i]<-log(muRec[i])-0.5*1/tauRec
 }

 for(i in 1:N_Stocks){
  logit(alpha[i])<-A[i]
  A[i]~dnorm(mu_a,tau_a)
  K[i]~dlnorm(0,0.00001)
 }

 ANew~dnorm(mu_a,tau_a)
 logit(alphaNew)<-ANew

 cvRec~dlnorm(-0.3,0.00001)
 tauRec<-1/log(cvRec*cvRec+1)

 # Uninformative prior for slope at logit scale
 mu_a~dnorm(-1,0.001)
 sd_a~dlnorm(-2,0.05)
 tau_a<-1/(sd_a*sd_a)

}
```

Annex 2: Derivation of priors for Beverton–Holt and Ricker stock–recruitment models using inputs to the ECOKNOWS stock–recruitment elicitation tool (Section 3.4.3.1)

Prior for α or a (slopes at the origin in the Beverton–Holt and Ricker models)

Elicitation tool inputs:

- Maximum lifetime reproductive rate $\hat{\alpha}$
- Spawner biomass per recruit (SBPR) or eggs per recruit (EPR) in the absence of fishing

The maximum lifetime reproductive rate ($\hat{\alpha}$) is the number of spawners produced by each spawner over its lifetime at very low spawner abundance (a function of longevity, fecundity, and survival). Estimates of the maximum lifetime reproductive rate (e.g. from meta-analysis) have been published for many species (e.g. Myers *et al.*, 1999), so that the distribution for $\hat{\alpha}$ can be taken from the literature if available; α (or a) can be obtained as:

$$\alpha = \frac{\hat{\alpha}}{SBPR}$$

SBPR is calculated as:

$$SBPR = \sum_{i=1}^{A-1}[w_i\eta_i \exp(-\sum_{j=0}^{i-1}M_j)] + w_A\eta_A \exp(-\sum_{j=0}^{A-1}M_j)/[1-\exp(-M_A)]$$

where w_i is weight at age i, η_i is maturity at age i, M_i is the rate of instantaneous natural mortality at age I, and A denotes the plus group. A prior for SBPR could be derived using available information on natural mortality-, weight-, and maturity-at-age, or alternatively, constructed using distributions for natural mortality, growth parameters, and maturity elicited from experts.

β (Beverton–Holt) from maximum recruitment

Elicitation tool inputs

- Maximum recruitment ($\max R$)

In the parameterization of the Beverton–Holt model used by the ECOKNOWS elicitation tool, $\max R = \frac{\alpha}{\beta}$, so that $\beta = \frac{\alpha}{\max R}$.

β (Beverton–Holt) from unfished recruitment R_0 (assumes equilibrium)

The following equations can be used to derive priors for stock–recruitment parameters corresponding to demographic equilibrium, conditional on the information provided about vital rates (used to calculate SBPR) and recruitment at demographic equilibrium.

Elicitation tool inputs

- Recruitment at equilibrium (R_0)

- Spawner biomass per recruit (SBPR) or eggs per recruit (EPR) in the absence of fishing
- Maximum lifetime reproductive rate $\hat{\alpha}$

The Beverton–Holt β parameter can be found by noting that the parameters steepness h (defined as the proportion of recruitment obtained when spawner abundance or biomass is reduced to 20% of the virgin level, relative to recruitment at unfished equilibrium) and maximum lifetime reproductive rate ($\hat{\alpha}$) are related by:

$$h = \frac{\hat{\alpha}}{4+\hat{\alpha}}$$

(Myers *et al.*, 1999) so that (using the parameterization $R = \frac{\alpha S}{1+\beta S}$):

$$\beta = \alpha \left(\frac{5h-1}{4hR_0} \right)$$

b (Ricker) from maximum recruitment

Elicitation tool inputs

- Maximum recruitment ($\max R$)

Maximum recruitment ($\max R$) is given by $\max R = \dfrac{a}{be}$, so that $b = \dfrac{a}{\max R\,e}$.

b (Ricker) from unfished recruitment R_0 (assumes equilibrium)

Elicitation tool inputs

- Recruitment at equilibrium (R_0)
- Spawner biomass per recruit (SBPR) or eggs per recruit (EPR) in the absence of fishing

Writing the Ricker equation ($R = aSe^{-bS}$) as $S_0 = \log(aSBPR)\frac{1}{b}$, where S_0 is spawning stock biomass (or some other measure of spawning potential) at equilibrium and $R_0 = \frac{S_0}{SBPR}$, the Ricker b parameter can be obtained as:

$$b = \frac{1}{R_0 SBPR / \log(aSBPR)} .$$

Annex 3: Methods to obtain priors for parameters of normal and double normal selectivity curves (Section 3.4.3.2)

Normal selectivity

The normal distribution is parameterized by its mean μ and standard deviation σ. Normal and inverse gamma priors are specified for μ and σ^2 using the elicited mode and quantile(s) for the length at which selectivity equals 0.50 and the elicited model and quantile(s) for the length at maximum selectivity (i.e. μ) (Figure 1).
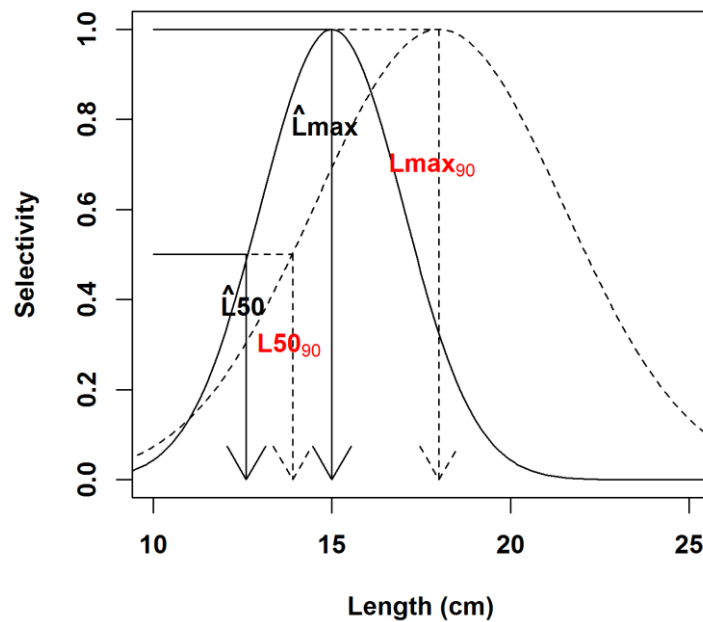

Figure 1. An example of a normal selectivity curve where the most likely (modal) length at 50% selectivity ($\hat{L}50$) is 12.5 cm, and the modal length at maximum (100%) selectivity ($\hat{L}max$) is 15 cm. The 0.90 quantile for $L50$ ($L50_{90}$) is 14 cm, and the 0.90 quantile for $Lmax$ ($Lmax_{90}$) is 18 cm.

Double normal selectivity

The procedure used to obtain priors for a double normal selectivity curve is analogous to that used for the normal curve, except that quantiles for $L50$ and $Lmax$ were elicited for both the right-hand and left-hand sides of the curve. Selectivity is then set equal to 1 between the left-hand and right-hand $Lmax$ values (Figure 2).
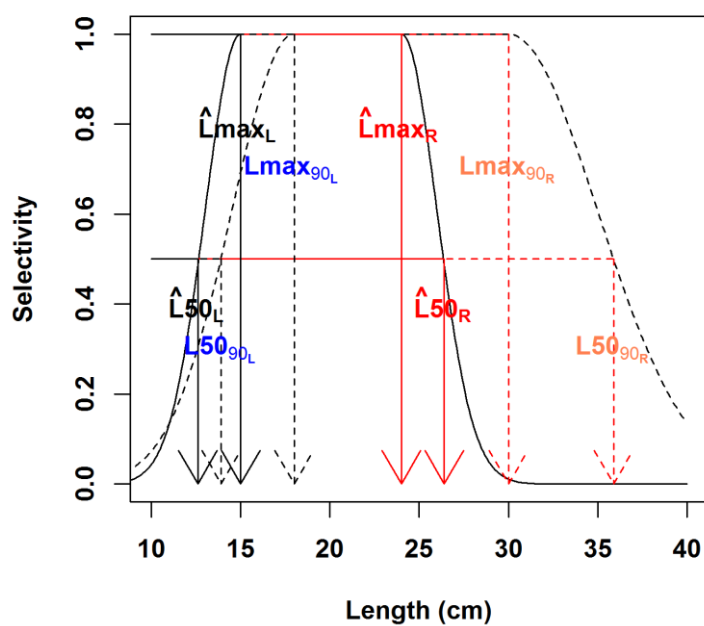
Figure 2. An example of a double normal selectivity curve where the most likely (modal) length at 50% selectivity on the left-hand side ($\hat{L}50_L$) is 12.5 cm and on the right-hand side ($\hat{L}50_R$) is 26.4 cm. The most likely range for selectivity equal to 1 is between 15 cm ($\hat{L}max_L$) and 24 cm ($\hat{L}max_R$). The 0.90 quantile for $L50_L$ ($L50_{90_L}$) is 14 cm, and the 0.90 quantile for $L50_R$ ($L50_{90_R}$) is 36 cm.

Annex 4: Equations used to derive priors using experts' input in the R selectivity elicitor tool (Section 3.4.3.2)

Normal selectivity

The mean of the prior for μ (μ_μ) is set equal to the expert's mode for μ. The standard deviation of the prior for μ is found by dividing the absolute difference between the mode for the mean and the elicited 0.10 or 0.90 quantile (or the largest absolute difference between the quantiles and the mode if both are elicited) by the 0.50 quantile of a standard normal distribution minus the 0.10 quantile:

$$\sigma_\mu = \frac{\max(L\max_{90} - \hat{L}\max, \hat{L}\max - L\max_{10})}{\phi^{-1}(0.50) - \phi^{-1}(0.10)}$$

where, e.g. $Lmax_{10}$ denotes the expert's 10th percentile for the length at maximum selectivity and ϕ^{-1} is the quantile function of the standard normal distribution. A caret denotes the mode.

The standard deviation of the normal selectivity curve (σ) can be found using the fact that the value corresponding to 50% of the probability density at the mode (and mean) in a standard normal distribution occurs at 1.177 standard deviations away from the mean. The mode of the prior for σ is thus obtained as:

$$\hat{\sigma} = \frac{\hat{L}\max - \hat{L}50}{1.177}$$

The 0.90 quantile for the prior for the standard deviation (σ_{90}) is obtained using the difference between the 0.10 quantile for $L50$ and the 0.90 quantile for $Lmax$; this is again divided by 1.177. After squaring to obtain variances, the parameters of an inverse gamma (IG) distribution for the variance are found using an iterative routine that is initialized with a starting value for the inverse gamma α (shape) parameter, computes the β (scale) parameter as:

$$\beta' = \hat{\sigma}(\alpha + 1)$$

and computes the difference between the 0.90 quantile of the resulting IG distribution and σ_{90}. In each subsequent iteration, the value of α is modified and the other steps repeated until the 0.90 quantile for σ is within 0.50 of the value obtained above. A prior for use in BUGS applications (that use precisions rather than variances by convention) can then be obtained as:

$$\frac{1}{\sigma^2} \sim \text{Gamma}(\alpha, \beta^{-1})$$

Logistic selectivity

The median of the prior for steepness (υ) can be found using the fact that $L90$ minus $L50$ times υ equals 2.2:

$$\upsilon_{50} = \frac{2.2}{L90_{50} - L50_{50}}$$

A 0.10 quantile for υ is computed as 2.2 divided by the difference between the 0.10 quantile for 50% selectivity and the 0.90 quantile for 90% selectivity:

$$v_{10} = \frac{2.2}{L90_{90} - L50_{10}}$$

Means (of $\log(x)$) of the log-normal distributions for $L50$ and v (μ_{L50} and μ_v) are then given by $\log(L50_{50})$ and $\log(v_{50})$, while the standard deviations can be found using the quantile function of the standard normal distribution (ϕ^{-1}), e.g. using the median and 0.10 quantile for v:

$$\sigma_v = \frac{\log(v_{50}) - \log(v_{10})}{\phi^{-1}(0.50) - \phi^{-1}(0.10)}$$

Annex 5: BMA example using JAGS (Section 3.4.4)

Three experts use the ECOKNOWS elicitor tool to summarize their beliefs about the proportion of tagged sea trout at age 2 released into the River Isojoki in Finland that migrate to the sea in the same year S. They each give a mode and mode weight, which are used to obtain the alpha and beta parameters of a beta distribution. In the following code, "#"denotes a comment.

```
PriorModel<-"
model{
for(i in 1:N_experts){          #Here N_experts=3
p[i]<-1/N_experts               #equal prior probabilities across experts
}


Y~dcat(p[1:N_experts])          #Indicator variable for the iᵗʰ expert's assessment. Y
                                takes values {1,2,3} with probabilities {p[1],p[2],p[3]}
P.smolt<-S[Y]                   #Combined (model averaged) distribution
#The following lines express the individual assessments of the 3 experts as Beta distri-
butions.
# Expert 1
# =================
S[1]~dbeta(64,8)
# Expert 2
# =================
S[2]~dbeta(8,4)
# Expert 3
# =================
S[3]~dbeta(37,5)
}"
```

Author contact information

Editor

Atso Romakkaniemi
Natural Resources Institute Finland
Paavo Havaksentie 3
FI-90014 University of Oulu, Finland
atso.romakkaniemi@luke.fi

Authors

Charis Apostolidis
Department of Zoology, School of Biology
Aristotle University of Thessaloniki,
UP Box 134, 541 24 Thessaloniki, Greece

Institute of Marine Biological Resources and Inland Waters
Hellenic Centre for Marine Research
Aghios Kosmas, Athens 16777, Greece
chapost@gmail.com

Guillaume Bal
Fisheries Science Services,
Marine Institute
Renville, Oranmore, Co. Galway, Ireland
guillaume.bal.pro@gmail.com

Rainer Froese
GEOMAR Helmholtz-Centre for Ocean Research
Duesternbrooker Weg 20, 24105 Kiel, Germany
rfroese@geomar.de

Juho Kopra
Department of Mathematics and Statistics
P.O. Box 35 (MaD), FI-40014 University of Jyväskylä, Finland
juho.j.kopra@jyu.fi

Sakari Kuikka
Fisheries and Environmental Management Group (FEM)
Department of Environmental Sciences
P.O. Box 65, FI-00014 University of Helsinki, Finland
sakari.kuikka@helsinki.fi

Adrian Leach
Centre for Environmental Policy
Imperial College London
Silwood Park Campus
Ascot, Berkshire SL5 7PY, United Kingdom
a.w.leach@imperial.ac.uk

Polina Levontin
Centre for Environmental Policy
Imperial College London
Silwood Park Campus
Ascot, Berkshire SL5 7PY, United Kingdom
polina.levontin02@imperial.ac.uk

Samu Mäntyniemi

Fisheries and Environmental Management Group (FEM)
Department of Environmental Sciences
P.O. Box 65, FI-00014 University of Helsinki, Finland
samu.mantyniemi@helsinki.fi

Niall Ó Maoiléidigh

Fisheries Science Services
The Marine Institute
Newport, Co. Mayo, Ireland
niall.omaoileidigh@marine.ie

John Mumford

Centre for Environmental Policy
Imperial College London
Silwood Park Campus
Ascot, Berkshire SL5 7PY, United Kingdom
j.mumford@imperial.ac.uk

Henni Pulkkinen

Natural Resources Institute Finland
Paavo Havaksentie 3, FI-90014 University of Oulu, Finland
henni.pulkkinen@luke.fi

Etienne Rivot

Agrocampus Ouest
Research Unit Ecology and Ecosystem Health, UMR 0985 ESE, Agrocampus Ouest, INRA
65, rue de Saint Brieuc, CS 84215, 35042 Rennes Cedex, France
Etienne.Rivot@agrocampus-ouest.fr

Vaishav Soni

International Council for the Exploration of the Sea (ICES)
H. C. Andersens Boulevard 44-46, 1553 Copenhagen V, Denmark
vaishav@ices.dk

Konstantinos Stergiou

Institute of Marine Biological Resources and Inland Waters
Hellenic Centre for Marine Research
Aghios Kosmas, Athens 16777, Greece

Department of Zoology, School of Biology
Aristotle University of Thessaloniki, UP Box 134, 541 24 Thessaloniki, Greece
kstergio@bio.auth.gr

Jonathan White

Fisheries Science Services
Marine Institute
Renville, Oranmore, Co. Galway, Ireland
jonathan.white@marine.ie

Rebecca Whitlock

Natural Resources Institute Finland
Itäinen pitkäkatu 3, FI-20520 Turku, Finland
rebecca.whitlock@slu.se