Inference for Deterministic Simulation Models: The Bayesian Melding Approach

David Poole and Adrian E. Raftery [1]

Technical Report no. 346
Department of Statistics
University of Washington
December 1998

---

[1] Address for correspondence: Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322, USA. Tel: (206) 543-4505; Fax: (206) 685-7419; E-mail: `poole@stat.washington.edu`, `raftery@stat.washington.edu`; Web: `http://www.stat.washington.edu/raftery`.

Summary

Deterministic simulation models are used in many areas of science, engineering and policymaking. Typically, they are complex models that attempt to capture underlying mechanisms in considerable detail, and they have many user-specified inputs. The inputs are often specified by some form of trial-and-error approach in which plausible values are postulated, the corresponding outputs inspected, and the inputs modified until plausible outputs are obtained. Here we address the issue of more formal inference for such models. Raftery *et al.* (1995a) proposed the Bayesian synthesis approach in which the available information about both inputs and outputs was encoded in a probability distribution and inference was made by restricting this distribution to the submanifold specified by the model. Wolpert (1995) showed that this is subject to the Borel paradox, according to which the results can depend on the parameterization of the model. We show that this dependence is due to the presence of a prior on the outputs.

We propose a modified approach, called *Bayesian melding*, which takes full account of information and uncertainty about both inputs and outputs to the model, while avoiding the Borel paradox. This is done by recognizing the existence of two priors, one implicit and one explicit, on each input and output; these are combined via logarithmic pooling. Bayesian melding is then standard Bayesian inference with the pooled prior on inputs, and is implemented here by posterior simulation using the SIR algorithm. We develop this initially for invertible models, and then extend it to the more difficult case of noninvertible models. We also propose diagnostic checking, model validation, hypothesis testing and model selection methods, so that Bayesian melding provides a comprehensive framework for coherent statistical inference from deterministic simulation models. The situation where invariance to relabelling of inputs as outputs or vice versa is desired is approached via a further extension, called joint pooling. The methodology is illustrated using population dynamics models of varying complexity, and various open research problems are discussed. The approach was used by the International Whaling Commission in 1998 as a basis for the management of aboriginal whaling.

KEY WORDS: Bayesian inference; Borel paradox; logarithmic pooling; population dynamics models; SIR algorithm; whales.

Contents

List of Tables

List of Figures

1 Introduction

Simulation models are widely used in applied scientific disciplines. Such models are found in the study of climate, soil pollution, plant growth, epidemiology, animal populations, and other fields. Given a set of inputs, a simulation model produces a set of outputs. For example, inputs to a biological population dynamics model may include mortality and reproduction rates for various age groups, habitat parameters, and harvest information. Outputs would then typically include current population abundance and age structure. In a model of soil pollution, typical inputs include soil density, deposition velocity, and air pollutant concentration. The output would be a measure of pollutant concentration in the soil.

Simulation models are usually designed to capture some underlying mechanism or natural process, and they are often deterministic. They differ conceptually from many standard statistical models (such as linear regression) whose aim is to empirically estimate the relationships between variables. For many scientists, a mechanism is most naturally modeled using a deterministic approach. The deterministic simulation model is viewed as a useful approximation of reality that is easier to build and interpret than a stochastic model. On one hand, determinism permits ease of model construction and understanding; on the other, ignoring stochastic variation can result in a loss of modelling realism. In some cases, the positive aspects of the former are considered to outweigh the disadvantages of the latter. In other situations, the random variation is thought to account for little of the overall uncertainty, in which case a deterministic model closely approximates a stochastic counterpart. Ignoring random variation can then be thought of as a modelling assumption.

A deterministic simulation model is not necessarily a simple model. Some simulation models are extremely complicated with large numbers of inputs and outputs. The relationships between variables can be highly complex and the models are often noninvertible: a fixed single set of outputs can be generated by multiple sets of inputs. Thus, taking reliable account of parameter and model uncertainty is crucial, perhaps even more so than in standard statistical models, yet this is an area that has received little attention from statisticians. A statistician tends to automatically quantify uncertainty through the use of stochastic statistical models, and with good cause, but this may have contributed to a lack of interaction between statisticians and the applied scientists who have adopted the deterministic modelling approach.

Edwards (1967) observed that there was a general need for simulation-based methods of estimation in complex models. Hoel and Mitchell (1971) and Ross (1972) provided some early examples of simulation in stochastic models. Diggle and Gratton (1984) presented Monte Carlo methods for analyzing an implicit statistical model, stating that "we know of no other systematic investigations along the lines of the present paper." Speed (1983) was perhaps the first to point out the need for statisticians to get involved in the deterministic modelling arena, noting that scientists at the Australian national research organisation (CSIRO) were increasingly abandoning statistical models in favor of simulation models. As examples, he listed a sheep growth model, a model for predicting the nitrogen requirements of wheat crops, and a model describing the effects of light and water vapor on apple leaves. More recently, Hodges (1987, 1991) has also argued that it is vital to quantify the uncertainty in the inputs to a simulation model.

Although little attention has been paid to formally quantifying parameter uncertainty, simulation model validation has received considerable attention. Caswell (1976), Hughes (1984) and Guttorp and Walden (1987) discussed applications in ecology, the military, and geophysics respectively.

The most common way of specifying the "inputs" to such a model is a kind of *ad hoc* trial-and-error or "tuning" approach. Inputs can include parameters that quantify aspects of the underlying mechanism, initial conditions, and control parameters that specify how the simulation is to be run. The tuning approach starts with an initial guess at appropriate values of the inputs, based on professional knowledge and expertise, information in the relevant scientific literature, and so on. The model is then run with these inputs, and the outputs are examined. If the outputs seem plausible, the initial guess is used; otherwise the guess is modified. The process iterates until a satisfactory set of inputs is found, i.e. one that seems reasonable in itself and also produces plausible outputs.

While little attention tends to be paid to uncertainty about the inputs, a sensitivity analysis is often run to see if the final conclusions are sensitive to the precise values of the inputs used. This is a very good idea, but there are difficulties with it. If the conclusions turn out to be insensitive to the inputs, all is well. However, if there is some sensitivity, it is not clear what should be done except to note it in the report, and to attach a "government health warning" to the conclusions. This seems somewhat unsatisfactory, as there will often be a degree of sensitivity to the inputs of a complex model. We feel that sensitivity is a form of uncertainty, and that it should be taken into account explicitly when drawing conclusions.

The present research was initially motivated by work for the International Whaling Commission (IWC) on setting quotas for aboriginal subsistence whaling of bowhead whales, *Balaena mysticetus*, by Inuit peoples in Alaska. This has been done traditionally using a deterministic age-structured population dynamics model for the whales to assess their current natural rate of increase (= fertility − mortality). A conservative assessment of the rate of increase is then used to set an upper bound on the quota. The model uses information on the historic commercial catches and the current population size, and requires as inputs parameters that describe age-specific fertility and mortality rates, and initial population size. The results are quite sensitive to these inputs, and until 1994 the inputs were set using the *ad hoc* trial-and-error method described above, with the "conservative" assessment of natural increase rate being taken as the lowest value from a rather limited sensitivity analysis.

In 1991, the IWC decided that this approach gave too crude an assessment of uncertainty, and that the result might, for example, be too high, running the risk of damaging the stock. They called for better methods to assess uncertainty in whale population dynamics models. The Bayesian synthesis method was a first attempt to meet this call (Raftery *et al.*, 1992, 1995a; Givens, 1993; Givens *et al.*, 1995), and was used by the IWC to set the bowhead quota in 1994. It recognizes that the information about "plausibility" of input and output values that drives the tuning approach comes from a combination of quantitative and qualitative data and expert opinion, and encodes it explicitly using a probability distribution called the *pre-model distribution*. However, Wolpert (1995) showed that Bayesian synthesis was subject to the Borel paradox, so that in principle the results could depend on the model's parameterization. In their response to Wolpert, Raftery *et al.* (1995b), argued that the effect of the Borel paradox on the results was likely to be small in practice. Nevertheless, any method that is subject to the Borel paradox is unsatisfactory.

Here we describe a new approach, called *Bayesian melding*, that retains the desirable aspects of the Bayesian synthesis method but is not subject to the Borel paradox. This draws on existing work on combining expert opinions by pooling the corresponding probability distributions. The word *melding* is used because the method provides a coherent way of combining, or melding, different kinds of information (qualitative or quantitative, fragmentary or extensive, based on expert knowledge or on data) about different quantities, so long as the quantities they relate to can be linked using a deterministic model. It turns out that the numerical results for bowheads from the new approach are almost identical to those from the earlier Bayesian synthesis method, but the new approach is based on a sounder inferential footing. Earlier unpublished progress reports on this project were given by Raftery *et al.* (1996), Raftery and Poole (1997) and Poole and Raftery (1998).

In Section 2 we give background on deterministic simulation models, the Bayesian synthesis approach, and the Borel paradox. In Section 3, we review existing ideas of pooling probability distributions to combine the opinions of experts, and we build on those to develop the Bayesian melding approach, using several simple examples. In Section 4 we show how these ideas can be used for model validation and diagnostic checking, and in Section 5 we indicate how they can be extended to perform hypothesis testing and model selection, and how they can be used to take account of uncertainty about model form. In Section 6 we consider the situation where inputs can be relabelled as outputs, or vice versa, and we show how Bayesian melding can be extended to yield results that are invariant to such relabellings. In Section 7 we return to our initial motivating problem and show a Bayesian melding analysis of the BALEEN II model used for providing scientific advice on bowhead quotas. The theorems stated are proved in the Appendix.

2 Background

2.1 Deterministic Simulation Models

A deterministic model is simply a function relating a set of input variables to a set of output variables. We use the notation

$$M : \theta \to \phi, \qquad \theta \in \Theta \subseteq \Re^n, \quad \phi \in \Phi \subseteq \Re^p,$$

to denote that the deterministic model M relates a vector of input parameters θ to a vector of outputs ϕ, so that $\phi = M(\theta)$. We denote by ψ the set of quantities of interest, which may be model inputs, model outputs or functions of both, and typically will be functions of ϕ and/or θ. The model M may be non-invertible, and in many deterministic simulation applications the non-invertibility is due to the dimension of ϕ being less than that of θ, i.e. $p < n$. In these cases, a single value of the output may result from many different values of the input.

Our motivating application is the study of population dynamics models (PDMs), and our examples will be of this type. A population dynamics model relates the population at time $(t+1)$ to the population at time t. Age-structured population dynamics models relate the population aged a at time $(t+1)$ to the population at each age at time t. Here are some examples of deterministic population dynamics models, ranging from very simple to somewhat more complex.

Example 1: $P_t = RP_0$: This is an extremely simple PDM where P_0 represents the initial (year 0) population size, P_t is the size of the population in year t, and R is a growth rate parameter. P_0 and R are the 2 model inputs while P_t is the single output. Hence, $\theta = (P_0, R)$ and $\phi = P_t$. This is a simple example of a case where a given output can be generated by infinitely many values of the inputs.

Example 2: A model that underlies the commercial Revised Management Procedure of the IWC is a non-age-structured PDM of the form

$$P_{t+1} = P_t - C_t + 1.5(\text{MSYR})P_t\left(1 - (P_t/P_0)^2\right), \tag{1}$$

where P_t is the population in year t, where $t = 0$ corresponds to the baseline year before commercial hunting started (1848 in the case of bowheads), P_0 is the initial population size, MSYR is the maximum sustainable yield as a proportion of the population aged one year or more, and C_t is the number of whales killed by hunting in year t.

This model is much simpler than the BALEEN II PDM used by the IWC for bowhead assessment, but it nevertheless captures several of the major features of the bowhead population. The model is viewed as having two inputs (P_0 and MSYR) and one output (P_{1993}); using given values of the inputs, (1) is applied recursively until P_{1993} is obtained. The catch history is assumed to be known exactly, so that the C_t's are viewed as a set of constants. Although the population can be projected further to the present year, 1993 is the latest year for which independent abundance data are available. A related quantity of interest is the recent (1978-1993) rate of increase of the population.

The model (1) is simple because it is not age-structured, but it does feature density dependence, according to which the population increases more slowly when it is larger. The density dependence is introduced by the factor $(1 - (P_t/P_0)^2)$ on the right hand side of (1). This model implies the existence of a "carrying capacity", usually denoted by K, and here assumed to be equal to P_0, postulated to be the maximum number of animals that its environment could sustain. If it goes above this level, the population is assumed to decrease until it reaches the carrying capacity again.

Example 3: This is a non-density-dependent version of Example 2, defined by

$$P_{t+1} = P_t - C_t + 1.5(\mathrm{MSYR})P_t. \tag{2}$$

The difference between this and equation (1) is the absence of the density dependence factor $(1 - (P_t/P_0)^2)$.

Example 4: Here the model is the same as in Example 2, but for a different species, the eastern Pacific gray whale, *Eschrichtius robustus.* Details of population dynamics modelling for this species are given by Reilly (1984, 1992) and Wade and DeMaster (1996).

Example 5: BALEEN II : This is the full age-structured density-dependent PDM used for bowhead assessment. In addition to P_0 and MSYR, it requires inputs pertaining to survival rates, age at sexual maturity, calving intervals, and other biological information. BALEEN II is discussed in greater detail in Section 7.

In many applications, information on θ and ϕ is available independently of M. Data, statistical models and prior beliefs can yield evidence about parameters that are either inputs to or outputs of M. Such statistical models should not be confused with the simulation model M. The goal of simulation modelers is to sensibly combine the various sources of evidence to obtain the best possible inference regarding the quantities of interest.

2.2 The Bayesian Synthesis Approach

In 1991, the IWC Scientific Committee recommended that methods for taking full account of uncertainty in inputs and outputs to population dynamics models be developed. In response to this recommendation, the Bayesian synthesis method was introduced by Raftery *et al.* (1992), Givens (1993), Givens *et al.* (1993), and Raftery, Givens and Zeh (1995a), hereafter RGZ. The method was used as the basis for the IWC assessment of bowhead whales in 1994.

As we did in Section 2.1, RGZ denoted by θ the set of model inputs and by ϕ the set of model outputs about which we have information independent of the simulation model M. (RGZ used Φ rather than M to denote the deterministic model function.) They then denoted by $p(\theta, \phi)$ the joint *pre-model distribution* of θ and ϕ, which summarizes all available information about θ and ϕ except that embodied in the model itself.

The model defines a mapping from the set of possible values of θ to the set of possible values of ϕ; RGZ denoted this mapping by $\theta \longrightarrow M(\theta)$. They defined the joint distribution of θ and ϕ given the model to be simply the restriction of the pre-model distribution to the submanifold $\{(\theta, \phi) : \phi = M(\theta)\}$, namely

$$\pi(\theta, \phi) \propto \begin{cases} p(\theta, M(\theta)) & \text{if } \phi = M(\theta) \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

RGZ referred to $\pi(\theta, \phi)$ as the *post-model distribution.* The marginal post-model distribution of θ is then

$$\pi^{[\theta]}(\theta) \propto p(\theta, M(\theta)), \tag{4}$$

or, equivalently,

$$\pi^{[\theta]}(\theta) = p^{[\theta|\phi]}(\theta \mid \phi = M(\theta)). \tag{5}$$

For marginal and conditional distributions, RGZ used superscripts in square brackets to show to what the distribution applies. Thus, for example, $p^{[\theta]}(\cdot)$ denotes the marginal pre-model distribution of the inputs, $\pi^{[\phi]}(\cdot)$ denotes the marginal post-model distribution of the outputs, and $p^{[\phi|\theta]}(\cdot|\cdot)$ denotes the conditional pre-model distribution of the outputs given the inputs.

In a Bayesian context, it is useful to decompose the pre-model distribution into prior and likelihood components. Under the assumption that pre-model information about inputs is independent of that about outputs, we can decompose the pre-model distribution as follows:

$$\begin{aligned} p(\theta, \phi) &= p^{[\theta]}(\theta) p^{[\phi]}(\phi) \\ &\propto q_1(\theta) q_2(\phi) L_1(\theta) L_2(\phi), \end{aligned} \tag{6}$$

where $q_1(\cdot)$ and $q_2(\cdot)$ are the prior distributions of the inputs and outputs respectively, $L_1(\theta) = p(D_\theta|\theta)$ is the likelihood of the inputs, and $L_2(\phi) = p(D_\phi|\phi)$ is the likelihood of the outputs, where D_θ and D_ϕ represent data. Sometimes, data relating to only one of θ and ϕ, typically ϕ, are available, in which case only one likelihood is present. An example of a prior in the bowhead application (Example 5) is that of the adult mortality rate, which was based largely on similar rates for other species. An example of a likelihood is that of the recent rate of population increase (ROI), which was based directly on a time series of population indices (see RGZ).

Bayesian synthesis inference about a quantity is based on its post-model distribution. In order to draw samples from $\pi(\theta)$, RGZ used the *sampling-importance-resampling* (SIR) algorithm (Rubin, 1987, 1988). This is a Monte Carlo procedure consisting of the following steps, where the inputs and outputs are assumed to be pre-model independent:

1. Draw a large sample of size k values of θ from $p^{[\theta]}(\theta)$.
2. Compute the importance sampling weights $r_i = \frac{p(\theta_i, M(\theta_i))}{p^{[\theta]}(\theta_i)} = p^{[\phi]}(M(\theta_i))$ $(i = 1, \ldots, k)$.
3. Draw a second sample of size l from the discrete distribution with values θ_i and associated probabilities r_i.

The second sample is approximately a sample from the post-model distribution, and it is the basis for inference about inputs, outputs and quantities of interest in the Bayesian synthesis method.

2.3 The Borel Paradox

Wolpert (1995) pointed out that a conditional distribution of the form (5) is ill-defined, and as a result the Bayesian synthesis post-model distribution is subject to a phenomenon known as the Borel paradox. A consequence of Borel's paradox is that the post-model distribution depends on how the simulation model M is parameterized. Wolpert gave the example of a simple exponential growth PDM, in which the results are different if pre-model information is specified on the parameters of the model in logarithmic form than if it is specified in terms of the parameters of the model in its original form. Schweder and Hjort (1996) pointed out that the consequences of the paradox are, in theory, far-ranging: by choosing arbitrarily extreme parameterizations one can in principle obtain any density as the post-model distribution.

The Borel paradox manifests itself when a conditional distribution is defined on an arbitrary null event (or a set of probability zero). Such conditioning is indeterminate, and the resulting conditional density depends on how the space is parameterized and other irrelevant things (Wolpert, 1995). Billingsley (1986) also alludes to the phenomenon in his discussion of conditional probability. Billingsley's source was Kolmogorov (1933), who cited Borel (1909) and coined the term "Borel paradox". Borel in turn cited Bertrand (1882). It appears that the Borel paradox may have contributed to motivating Kolmogorov's (1933) development of the modern axiomatic basis of probability theory (Schweder and Hjort, 1995).

Wolpert (1995) cited a simple example of the Borel paradox. Let (x, y) be drawn uniformly on the unit square $\{(x, y) : 0 \leq x, y \leq 1\}$. We are interested in the conditional distribution of x given that (x, y) lies on the diagonal $y = x$. If we assume that "on the diagonal" means that $y - x = 0$, then the conditional distribution is also uniform (as one might intuitively expect). A standard way to show this is to reparameterize (x, y) to (w, z), where $w = x$ and $z = y - x$, find the joint density of (w, z) using the Jacobian of the transformation (which in this case is 1), and then finding $p(x|y - x = 0) = p(w|z = 0) = p(w, 0)/p(z = 0)$, where $p(z = 0)$ denotes the density of z evaluated at 0. This density turns out to be uniform on [0,1]. If, on the other hand, we take the conditioning to mean that $y/x = 1$ and proceed in the same way, the resulting conditional distribution of X is the Beta(2,1) distribution, with a density proportional to x. The paradox is that we obtain two different conditional distributions of X although it appears that we have conditioned on the same event.

The key problem with the conditioning here is exactly that indicated by Wolpert (1995). The set $\{(x, y) : y = x\}$ has probability zero and is of lower dimension than the square $\{(x, y) : 0 \leq x, y \leq 1\}$ in which it resides. Hence we are conditioning on a null event, or a *set* of points with probability zero. We should note that conditioning on a single point of probability zero with respect to a stochastic random variable does not, in general, give rise to this phenomenon; indeed, this is the definition of a standard continuous conditional density.

2.4 Effect of the Borel Paradox on the Bayesian Synthesis Approach

Inference from the Bayesian synthesis method depends on the manner in which the simulation model is parameterized. In this section, we investigate the exact nature of the noninvariance of the Bayesian synthesis to reparameterization of deterministic models.

Recall that RGZ defined a joint pre-model distribution, $p(\theta, \phi)$ on the pair $(\theta, \phi) \in (\Theta \times \Phi)$, where θ and ϕ are the inputs and outputs, respectively, of the model $M : \theta \rightarrow \phi$. Thus $p(\theta, \phi)$ has dimensionality equal to $(\dim(\Theta) + \dim(\Phi))$, the total number of inputs and outputs of M. The post-model distribution of θ was then defined to be the conditional distribution of the variable (θ, ϕ) given the model manifold $\phi = M(\theta)$, i.e. equation (5). However, since the model manifold is a lower dimensional surface in the space $(\Theta \times \Phi)$, it has probability zero in $p(\theta, \phi)$. The post-model distribution is thus conditioned on the isolated null event that the model holds, and as a result is subject to the Borel paradox. As noted in the simple illustration above, it is stochastic random variables (or σ-algebras) that must be conditioned on, not null events.

We now formulate the exact effect of model reparameterization on the Bayesian synthesis post-model distribution:

Theorem 1 : Let θ and ϕ be vector parameters related by the function $M : \theta \rightarrow \phi$, where M is not necessarily invertible. Let $p(\theta, \phi)$ represent the joint pre-model distribution of (θ, ϕ). Consider invertible transformations $\theta' = g_1(\theta)$ and $\phi' = g_2(\phi)$ so that $\theta = g_1^{-1}(\theta')$ and $\phi = g_2^{-1}(\phi')$. Then, if we re-express M in terms of θ' and ϕ' such that ϕ' is the explicit model output, the post-model distribution of the original θ, as defined in the Bayesian synthesis method, is

$$\pi^{[\theta]}(\theta) \propto p(\theta, M(\theta)) \left|\frac{\mathrm{d}g_2(\phi)}{\mathrm{d}\phi}\right|^{-1}_{\phi=M(\theta)} \tag{7}$$

□

Comparing (7) with (4), it is apparent that:

1. The reparameterization of the model causes the post-model distribution of the inputs, θ, to be multiplied by the inverse of the Jacobian of the output parameter transformation g_2, evaluated at $M(\theta)$. Thus, it is arbitrary reparameterizations of the outputs, ϕ, that result in arbitrary post-model distributions.
2. The post-model distribution is not affected by reparametrizations of the inputs, θ.

Theorem 1 does not require that the marginal pre-model distributions of θ and ϕ be independent, although most practical applications of Bayesian synthesis usually assume such independence.

Since the Bayesian synthesis approach relies on a conditional distribution that is ill-defined, it is not satisfactory in its original form. However, if it could be reformulated as a standard Bayesian procedure, then the Borel paradox would vanish. This is the motivation behind Bayesian melding.

3 Bayesian Melding: Standard Bayesian Inference via Logarithmic Pooling

3.1 Priors and Likelihoods

In considering the decomposition (6) of the pre-model distribution, it becomes apparent that the Borel paradox does not arise from the likelihood components. Likelihoods are invariant to reparameterization, as pointed out, for example, by Schweder and Hjort (1996). The problem therefore lies with the prior distributions $q_1(\theta)$ and $q_2(\phi)$.

Since θ is a random variable with density $q_1(\theta)$, $\phi = M(\theta)$ is also a random variable since it is a transformation of θ. In other words $q_1(\theta)$ and M together *induce* a distribution on ϕ, which we denote by $q_1^*(\phi)$. If M^{-1} exists, then $q_1^*(\phi) = q_1(M^{-1}(\phi))|J(\phi)|$ where $|J(\phi)|$ is the associated Jacobian. For a complex M, even when it is invertible, a functional form of $|J(\phi)|$ may be very difficult to calculate. When M is non-invertible and $\dim(\Phi)$ is less than $\dim(\Theta)$, it will be virtually impossible to obtain $q_1^*(\phi)$ analytically.

The Bayesian synthesis method does not account for the existence of $q_1^*(\phi)$. Counting this and the existing $q_2(\phi)$, there are thus two prior distributions on the same quantity, ϕ. Since these two priors are typically based on different sources of information, often not including knowledge of M, they may be different, or *incoherent*. If they could be replaced by a single prior, say $\tilde{q}^{[\phi]}(\phi)$, and then inverted to the input space to yield $\tilde{q}^{[\theta]}(\theta)$, we could then define

$$\pi^{[\theta]}(\theta) \propto \tilde{q}^{[\theta]}(\theta) L_1(\theta) L_2(M(\theta)), \tag{8}$$

which would be a standard Bayesian posterior, and standard Bayesian inference could then follow. Thus the Borel paradox would no longer arise. The concept of a *joint* pre-model distribution, as in (6), is no longer considered; we require only marginal distributions on θ and ϕ. As a result, there would no longer be a need to define the problematic conditional post-model distribution (5), and the problem would fall within the framework of standard Bayesian inference.

As with the Bayesian synthesis method, Monte Carlo methods would be required to obtain a random sample from $\pi^{[\theta]}(\theta)$, and inference would be based on the distribution of this sample. Inference about ϕ would follow by examining the distribution of $\phi = M(\theta)$ where θ is drawn from the posterior above.

3.2 Combining Probability Distributions

The Borel paradox gives rise to the ill-defined post-model distribution of the Bayesian synthesis method, and the problem can in part be attributed to the implicit presence of more than one prior distribution on the same quantity. This occurs quite naturally in the simulation modelling framework under consideration. One might be tempted to discard one prior or the other if a basis for choosing the prior to be discarded could be found. In the applications we consider, this is often equivalent to discarding one group of sources of information. A further argument against this is that one often needs a prior on an input in order to select values of the inputs for runs of the simulation model M on a computer.

A second option is to combine the two prior distributions into a single one. With just one prior, standard Bayesian analysis becomes possible, and inference is based on the posterior distribution in the usual way.

Coherizing two prior distributions on the same quantity is related to another problem: that of reaching consensus in the presence of multiple expert opinions. This topic has received considerable attention in the statistical literature; French (1985) and Genest and Zidek (1986) reviewed work in the area. There was considerable research activity on this problem up to the mid-1980s, but since then interest in it seems to have dwindled. Two pooling methodologies that have received extensive study are:

- Linear pooling : $T(q_1, ...q_k) = \sum_{i=1}^{k} \alpha_i q_i$, and
- Logarithmic pooling : $T(q_1, ...q_k) \propto \prod_{i=1}^{k} q_i^{\alpha_i}$,

where $q_1, \ldots, q_k$ are the individual priors, $\alpha_1, \ldots, \alpha_k \geq 0$, $\sum \alpha_i = 1$, and the pooling operator T represents the single combined probability distribution.

In a Bayesian framework, one needs to consider the order in which pooling and updating of a prior (given a likelihood from data) are performed. There are two possibilities:

(i) Each prior distribution is updated using Bayes's Theorem and then the combined prior distribution is formed.

(ii) The combined prior distribution is first formed, and then this is updated using Bayes's Theorem.

It seems reasonable to insist that both procedures result in the same combined posterior distribution. Madansky (1978) called this property *external Bayesianity*. Genest (1984) and Genest *et al.* (1986) showed, under mild conditions, that the logarithmic pooling operator is the only pooling operator that is externally Bayesian. This result provides an argument for using logarithmic pooling within a Bayesian analysis framework. The logarithmically pooled prior is also invariant to rescaling of individual priors (Genest and Zidek, 1986), and is usually less dispersed than other options. In addition, it possesses the property that if one of the priors assigns zero probability to a particular parameter value, the combined prior must also assign probability zero to that value. This is the so-called *zero preservation property* (ZPP). In our application, where the priors represent different sources of evidence rather than the opinions of different experts, the ZPP seems desirable because if a source of evidence rules out a value of the inputs, it seems that a combined distribution should rule it out as well.

For the particular problem at hand (i.e. combining two Bayesian prior distributions for quantities linked by a deterministic function), logarithmic pooling appears to be the most suitable option. In terms of the output ϕ, the logarithmically pooled prior distribution has the form

$$\tilde{q}^{[\phi]}(\phi) \propto q_1^*(\phi)^{\alpha} q_2(\phi)^{1-\alpha}, \tag{9}$$

where $q_1^*(\phi)$ is the prior on ϕ induced by $q_1(\theta)$ and M, $q_2(\phi)$ is the existing prior on ϕ, and α is the pooling weight. If $q_1^*(\phi)$ and $q_2(\phi)$ both exist, then $\tilde{q}^{[\phi]}(\phi)$ must also exist as a consequence of the following result.

Theorem 2 : If ϕ is continuous and $\int q_1^*(\phi)d\phi = \int q_2(\phi)d\phi = 1$, then there exists a constant k_α such that

$$k_\alpha \int q_1^*(\phi)^{\alpha} q_2(\phi)^{1-\alpha} d\phi = 1 \quad \forall \alpha \in [0,1].$$

If ϕ is discrete-valued and $\sum_j q_1^*(\phi_j) = \sum_j q_2(\phi_j) = 1$, then there exists a constant k_α such that

$$k_\alpha \sum_j q_1^*(\phi_j)^{\alpha} q_2(\phi_j)^{1-\alpha} = 1 \quad \forall \alpha \in [0,1].$$

□

A major question of interest is how to choose the pooling weight α. French (1985) argued that the choice is essentially arbitrary. Here, in the case of two priors, $\alpha_1 = \alpha_2 = \alpha = \frac{1}{2}$ assigns equal weight to each individual prior. We refer to (9) with $\alpha = \frac{1}{2}$ as *geometric* pooling because (9) then amounts to taking the geometric mean of the two prior densities.

In the bowhead whale application, we essentially have one expert (the IWC Scientific Committee) placing prior distributions on two different quantities, rather than the more common situation in which two experts provide opinions on the same quantity. This provides an informal justification for the choice $\alpha = \frac{1}{2}$ in the bowhead case. Since the two sources of information are selected by the same expert, they can be viewed as being equally reliable, and should hence be assigned equal weight. Reliability should not be confused with precision of the opinions; precision is accounted for in the variances of the respective distributions. The choice of α will be discussed further in Section 8.

3.3 Logarithmic Pooling for Noninvertible Models

So far, we have considered inference about the model outputs, ϕ. If the model is one-to-one, inference about the model inputs, θ, can be done in exactly the same way, defining the two implied prior distributions of θ and pooling them geometrically, to obtain the combined prior distribution, $\tilde{q}^{[\theta]}(\theta)$. Alternatively, inference about the model inputs, θ, can be carried out by inverting the combined prior distribution of model outputs in equation (9). These two methods give the same pooled prior, and hence the same posterior in (8). When the model is not one-to-one, the geometrically pooled prior distribution of the model outputs, ϕ, is still unambiguously defined, but that of the model inputs, θ, is not.

Here we propose a solution that seems intuitively reasonable when the model is not invertible. We start from an example of a simple noninvertible model with just one input and one output, each of which is discrete. We use the insights gained from that to propose a general solution for the situation where both inputs and outputs are discrete. We then use a limiting argument to deduce a pooled prior density of inputs in the case where inputs and outputs may be continuous. We then show how the resulting overall method can be implemented via the SIR algorithm. Finally, we illustrate the resulting methodology with a very simple continuous example and a more complex simulation model example.

3.3.1 A Simple Discrete Noninvertible Example

Consider the following simple deterministic model. It has one input and one output, each of which is discrete, and it is noninvertible. The input, θ, has three possible values (1, 2, 3), while the output, ϕ, has two possible values (1, 2). The model, M, is: $1 \to 1$, $2 \to 2$, and $3 \to 2$. It is noninvertible because 2 and 3 both map onto 2, and so $M^{-1}(2)$ is not uniquely defined, as it could be either 2 or 3. The prior densities are as follows:

θ	$q_1(\theta)$	ϕ	$q_2(\phi)$
1	.7	1	.6
2	.2	2	.4
3	.1		

The pooled prior density of ϕ is derived from (9), and, after renormalization, is as follows:

ϕ	$q_2(\phi)$	$q_1^*(\phi)$	$\tilde{q}^{[\phi]}(\phi)$
1	.6	.7	.652
2	.4	.3	.348

Now, having obtained the pooled prior density on outputs, $\tilde{q}^{[\phi]}(\phi)$, how should we invert it to obtain the pooled prior density on inputs, $\tilde{q}^{[\theta]}(\theta)$? First, it seems clear that we should have $\tilde{q}^{[\theta]}(1) = \tilde{q}^{[\phi]}(1) = .652$. This is because there is a one-to-one relationship between $\theta = 1$ and $\phi = 1$, in the sense that $M(1) = 1$ and that 1 is the only value of θ for which $M(\theta) = 1$.

By similar reasoning, it seems clear that $\tilde{q}^{[\theta]}(2) + \tilde{q}^{[\theta]}(3) = \tilde{q}^{[\phi]}(2) = .348$. The question is, how should the $\tilde{q}^{[\phi]}(2) = .348$ be split between $\theta = 2$ and $\theta = 3$? Note that any choice will give a solution that is technically an inversion of $\tilde{q}^{[\phi]}(\phi)$, so we have to decide on other grounds which is the most reasonable of the possible splits.

We propose that the split be proportional to $q_1(\theta)$. The reasoning is that prior information about ϕ tells us nothing about the relative probability of $\theta = 2$ versus $\theta = 3$, since they both map onto the same value of ϕ, and so the ratio of the prior probabilities of $\theta = 2$ and $\theta = 3$ should be determined only by the prior distribution of the input. This leads to:

$$\begin{aligned} \tilde{q}^{[\theta]}(2) &= \tilde{q}^{[\phi]}(2)\left(\frac{q_1(2)}{q_1(2)+q_1(3)}\right) \\ &= .348\left(\frac{.2}{.2+.1}\right) \\ &= .232. \end{aligned}$$

Similarly, $\tilde{q}^{[\theta]}(3) = .348(.1/(.2+.1)) = .116$.

3.3.2 Discrete Case: General Solution

The simple example in Section 3.3.1 provides the intuition behind our proposed general solution for the case where inputs and outputs are all discrete. Suppose that the possible values of θ are $A_1, A_2, \ldots$, and that those of ϕ are $B_1, B_2, \ldots$. Suppose further that $m : \mathcal{N} \to \mathcal{N}$ is a mapping induced by M such that $M(A_i) = B_{m(i)}$ $(i = 1, 2, \ldots)$, and that $C_j = M^{-1}(B_j) = \{A_i : M(A_i) = B_j\}$. Then we derive the pooled prior density of the outputs, $\tilde{q}^{[\phi]}(\phi)$, as before, using equation (9). Here we use q to denote densities with respect to a dominating discrete counting measure (i.e. probabilities).

We invert $\tilde{q}^{[\phi]}(\phi)$ using the two rules suggested by the simple example in Section 3.3.1. First, we require that the pooled prior probability of C_j be the same as that of the corresponding B_j, i.e. that

$$\tilde{Q}^{[\theta]}(C_j) = \tilde{q}^{[\phi]}(B_j), \tag{10}$$

where $\tilde{Q}^{[\theta]}(C_j) = \sum\{\tilde{q}^{[\theta]}(A_i) : A_i \in C_j\}$ is the pooled prior probability of the set C_j, or, equivalently, the pooled prior *measure* of C_j. We will use capital Q to denote a probability measure. The second rule is that if C_j has more than one member, its probability be split between its members in proportion to their prior densities, $q_1(A_i)$.

This yields the pooled prior density on outputs

$$\tilde{q}^{[\theta]}(A_i) = \tilde{q}^{[\phi]}(B_{m(i)}) \left(\frac{q_1(A_i)}{Q_1(C_{m(i)})} \right). \tag{11}$$

We also have that

$$Q_1(C_j) = q_1^*(B_j), \tag{12}$$

Combining (11) with (12) gives the alternative form

$$\tilde{q}^{[\theta]}(A_i) = \tilde{q}^{[\phi]}(B_{m(i)}) \left(\frac{q_1(A_i)}{q_1^*(B_{m(i)})} \right). \tag{13}$$

It is easily verified that (13) yields the same solution as before for the simple discrete model in Section 3.3.1. In practice, equation (13) is more convenient than equation (11) because it does not involve $C_{m(i)}$. For a given A_i it may be difficult to determine $C_{m(i)}$ whenever the model M is complex. However, it is easy to determine $B_{m(i)}$ by running the model using A_i as the input. In addition, we have the following result:

Theorem 3 : The pooled prior given by (13) is a density with respect to the dominating counting measure, i.e. $0 \le \tilde{q}^{[\theta]}(A_i) \le 1$ for all i, and $\sum_i \tilde{q}^{[\theta]}(A_i) = 1$.

□

3.3.3 The Continuous Case

We now extend the general solution for the discrete case, equation (13), to the case where the inputs and outputs are continuous. We first use an intuitive limiting argument similar to that used in defining probability density functions as derivatives of cumulative distribution functions. We then show that the heuristic derivation results in well-defined densities under certain conditions.

Suppose that A is a small hypercube contained in Θ with side of length h. Let $B = M(A) = \{M(\theta) : \theta \in A\}$ and $C = M^{-1}(B) = \{\theta : M(\theta) \in B\}$. A, B and C are shown in Figure 1. We denote measures corresponding to prior densities by writing Q in place of q.

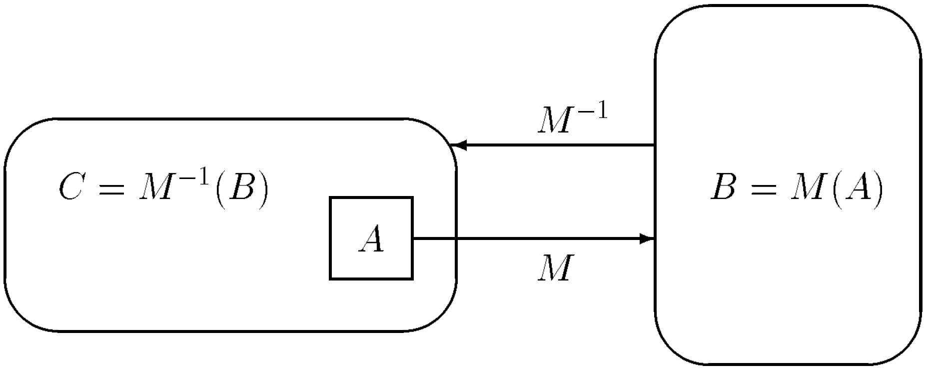


Figure 1: The sets A, B and C in the heuristic derivation of the pooled prior density of the inputs.

For $\tilde{Q}^{[\theta]}(\cdot)$ to be an inversion of $\tilde{Q}^{[\phi]}(\cdot)$, we require that

$$\tilde{Q}^{[\theta]}(C) = \tilde{Q}^{[\phi]}(B). \tag{14}$$

Also, by the definition of C and $Q_1^*(\cdot)$, we have that

$$Q_1(C) = Q_1^*(B). \tag{15}$$

The arguments we used in discussing the simple discrete example suggest that the proportion of $\tilde{Q}^{[\theta]}(C)$ attributed to A should be equal to the prior probability of A divided by that of C. This leads to

$$\begin{aligned} \tilde{Q}^{[\theta]}(A) &= \tilde{Q}^{[\theta]}(C) \left(\frac{Q_1(A)}{Q_1(C)} \right) \\ &= \tilde{Q}^{[\phi]}(B) \left(\frac{Q_1(A)}{Q_1^*(B)} \right), \end{aligned} \tag{16}$$

by equations (14) and (15). Then, we propose that the probability density function corresponding to (16), if it exists, be given by

$$\tilde{q}^{[\theta]}(\theta) = \tilde{q}^{[\phi]}(M(\theta)) \left(\frac{q_1(\theta)}{q_1^*(M(\theta))} \right). \tag{17}$$

This corresponds to letting the length h of the side of the hypercube A tend to zero.

We now identify two different situations in which (17) yields a well-defined probability density function that is also an inversion of $\tilde{q}^{[\phi]}(\phi)$. We first consider the case where the space of inputs can be divided up into disjoint sets such that within each the model, M, is invertible. A simple example would be $\phi = M(\theta) = \theta^2$, where $\Theta = \Re$ and $\Phi = \Re_+ \cup \{0\}$. There M is non-invertible, but if we split up the input space, Θ, into disjoint sets A_1 and A_2 where $A_1 = \Re_+ \cup \{0\}$ and $A_2 = \Re_-$, then $M: A_i \rightarrow \Phi$ will be invertible for each $i = 1, 2$. The general result is as follows.

Theorem 4 : Suppose that $\theta = (\theta_1, \ldots, \theta_n)$ and that $\phi = (\phi_1, \ldots, \phi_n)$, i.e. the model M has the same number of inputs as outputs. Suppose that θ and ϕ are assigned proper prior densities $q_1(\theta)$ and $q_2(\phi)$. Suppose also that $A_0, A_1, \ldots, A_k$ form a partition of Θ such that

(i) $P(\theta \in A_0) = 0$, and A_0 may be empty, and

(ii) the model map $\phi = M(\theta)$ is one-to-one from A_i onto a set B for each $i = 1, \ldots, k$, so that for each i, the inverse map can be found.

Then, for $0 \leq \alpha \leq 1$, the pooled prior on the model inputs θ given by (17) is a proper density function and an inverse of $\tilde{q}^{[\phi]}(\phi)$.

$\square$

We now consider the case where the model is noninvertible because the output ϕ is of lower dimension than the input θ.

Theorem 5 : Suppose that $\theta = (\theta_1, \ldots, \theta_n)$ and that $\phi = (\phi_1, \ldots, \phi_p)$ where $p < n$. Suppose also that there exists a transform $H : \theta \to \gamma$, where $\gamma \in \Gamma \subseteq \Re^{n-p}$, such that the n-dimensional transform

$$\begin{pmatrix} \phi \\ \gamma \end{pmatrix} = \begin{pmatrix} M(\theta) \\ H(\theta) \end{pmatrix}$$

is one-to-one. Suppose that θ and ϕ have proper prior densities $q_1(\theta)$ and $q_2(\phi)$, Then for $0 \leq \alpha \leq 1$, the pooled prior on the model inputs θ given by (17) is a proper density function and an inverse of $\tilde{q}^{[\phi]}(\phi)$.

$\square$

We conjecture that (17) holds more generally than for the cases above, where the restrictions ensure that an analytic form of the induced prior exists. In particular, we suspect that (17) is valid as long as M does not collapse onto a set of measure zero in ϕ-space, i.e. as long as $q_1^*(\phi)$ is not a point mass prior. We leave this as a matter of further research.

3.4 Implementing Logarithmic Pooling via the SIR Algorithm

To facilitate implementation, it is convenient to recast equation (17) as follows, using (9):

$$\tilde{q}^{[\theta]}(\theta) \propto q_1(\theta) \left(\frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha}, \tag{18}$$

and the posterior distribution of θ, $\pi^{[\theta]}(\theta)$, is then given by (8). The approach is implemented using the following modified SIR algorithm:

1. Draw a sample of size k values of θ from its prior distribution $q_1(\theta)$. We denote the sample by $(\theta_1, \ldots, \theta_k)$. (If $q_1(\theta)$ has a standard form, it will often be possible to do this using packaged random number generators for standard distributions).
2. For each θ_i sampled in step 1, run the model to obtain the corresponding $\phi_i = M(\theta_i)$.
3. Use nonparametric density estimation to obtain an estimate of $q_1^*(\phi)$, the resulting induced distribution of ϕ. We use kernel density estimation with a Gaussian kernel and (whenever relevant) the maximal smoothing span of Terrell (1990); this has the advantage of being easily applied in higher dimensions.
4. Form the importance sampling weights
$$w_i = \left(\frac{q_2(M(\theta_i))}{q_1^*(M(\theta_i))}\right)^{1-\alpha} L_1(\theta_i) L_2(M(\theta_i)). \tag{19}$$
5. Sample l values from the discrete distribution with values θ_i and probabilities proportional to w_i.

The result is an approximate sample from the posterior distribution $\pi^{[\theta]}(\theta)$, and it can be used to make inference about the various quantities of interest.

3.5 Example 1: A Simple Continuous Example

We now revisit Example 1 from Section 2.1, the simple PDM given by $P_t = RP_0$. Let $P_0 \sim U[0,5], R \sim U[2,4]$, and $P_t \sim U[6,9]$ be the mutually independent prior distributions of the parameters. (For simplicity in this example we will assume that there are no likelihoods available.) It follows that

$$\begin{aligned} q_1(p_0, r) &= \frac{1}{10}, \text{ for } 0 < p_0 < 5, \quad 2 < r < 4 \\ q_2(p_t) &= \frac{1}{3}, \text{ for } 6 < p_t < 9. \end{aligned}$$

Recall that the model is non-invertible since the output has lower dimension than the inputs. Indeed, each unique value of P_t can be associated with an infinite number of ordered pairs (P_0, R). This is a feature found in many of the more complicated PDMs used in real assessment situations. The simplicity of this example, however, allows an analytic solution to be found.

The change-of-variable technique yields the distribution of P_t induced by the PDM, and is given by

$$q_1^*(p_t) = \begin{cases} \frac{\log 2}{10} & \text{for } 0 < p_t < 10 \\ \frac{1}{10}\log\left(\frac{20}{p_t}\right) & \text{for } 10 < p_t < 20. \end{cases}$$

Geometric pooling of this induced density with $q_2(p_t)$ yields

$$\tilde{q}^{[\phi]}(p_t) = \frac{1}{3}, \text{ for } 6 < p_t < 9.$$

This pooled distribution on the output is uniform on the interval [6,9], and thus (in this particular case) it happens to be the same as the original $q_2(p_t)$. Applying (17), the pooled prior distribution of the inputs (P_0,R) is given by

$$\begin{aligned}\tilde{q}^{[\theta]}(p_0, r) &= \frac{\tilde{q}^{[\phi]}(p_0 r) q_1(p_0, r)}{q_1^*(p_0 r)} \\ &= \frac{\left(\frac{1}{3}\right)\left(\frac{1}{10}\right)}{\left(\frac{\log 2}{10}\right)} \\ &= \frac{1}{3 \log 2} \quad \text{for } 6 < p_0 r < 9, \quad 0 < p_0 < 5, \quad 2 < r < 4,\end{aligned}$$

and 0 elsewhere. In this simple case, the pooled prior remains uniform on a subset of the original rectangular region. Note that the density integrates to unity on its support. The marginal prior densities of P_0 and R are easily obtained from the above. For example, integrating out over R yields the marginal pooled prior of P_0 to be

$$\tilde{q}(p_0) \propto \begin{cases} 4 - 6/p_0 & \text{for } 1.5 < p_0 < 2.25 \\ 3/p_0 & \text{for } 2.25 < p_0 < 3 \\ 9/p_0 - 2 & \text{for } 3 < p_0 < 4.5. \end{cases}$$

Figure 2 shows a sample from the joint pooled prior of P_0 and R obtained using the SIR approach. The uniformity on the correct subset of the original rectangle is evident. Figure 3 shows the corresponding marginal pooled prior distributions of P_0, R and P_t. The exact marginals and the distributions simulated using SIR are shown. The SIR method provides a good approximation to the true distributions. The method also works well for the more complex examples we have examined.

3.6 Example 2: A Simple Population Dynamics Model

We now apply the pooling procedure to Example 2 from Section 2.1. This is the simple population dynamics model underlying the Revised Management Procedure of the IWC. Here we will use prior distributions that have been used in previous illustrations using this model. These are not necessarily priors agreed upon by the IWC Scientific Committee.

The model inputs are P_0 and MSYR. For P_0 we have a prior only, and we use the same prior as in Givens (1995), namely a shifted gamma distribution 6400 + Gamma (2.8085, 0.0002886). For MSYR, we also have only a prior. MSYR is the only productivity parameter in the model, and our prior information about it reflects all our prior information about natural fertility and mortality. We based this on a distribution of MSYR for the BALEEN II model reported by Punt

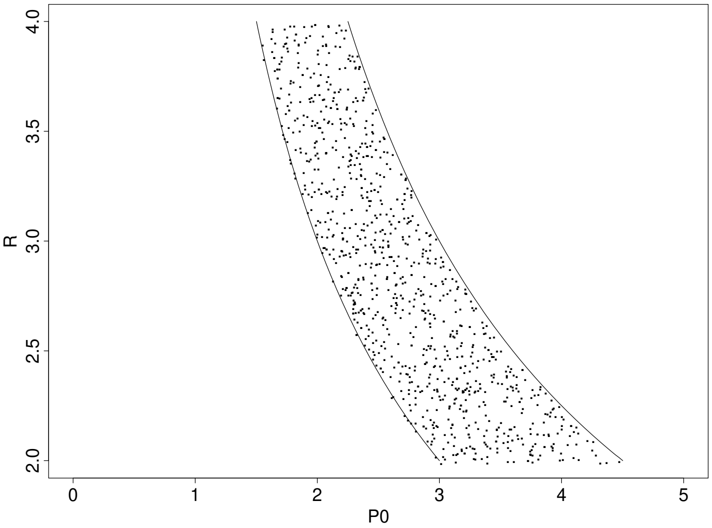
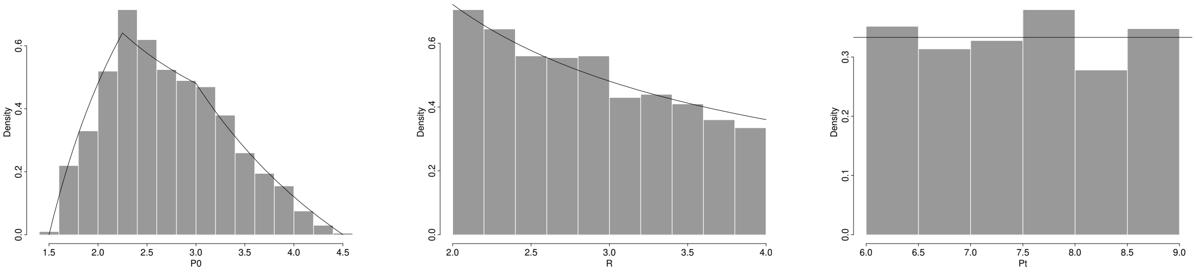


Figure 2: A sample from the pooled joint prior distribution of P0 and R in the simple example.



(a) (b) (c)

Figure 3: Exact and simulated pooled prior distributions of (a) P_0, (b) R and (c) P_t for the simple example. The simulated distributions are generated using SIR, and approximate the exact distributions well.

and Butterworth (1996), which had a mean of .0220 and 95% confidence interval $[.0082, .0335]$. We approximated this using a Gamma (8.2, 372.7) distribution, which has the same mean and the same .025 and .975 quantiles.

For the output P_{1993} we have both a prior and a likelihood. We approximated both of these by normal distributions (a good approximation), and used the results of Raftery and Zeh (1998) based on the 1993 census. The prior distribution is $N(7800, 1300^2)$, and the likelihood is $N(8293, 626^2)$. These choices correspond to a posterior distribution that is $N(8200, 564^2)$, matching the mean and variance of the Bayes empirical Bayes posterior distribution (which is nearly, but not exactly, normal).

Finally, we have a likelihood for the quantity of interest ROI, the 1978–1993 rate of increase based on the censuses from 1978 to 1993. This is defined by $P_{1993} = (1 + \mathrm{ROI})^{15} P_{1978}$. We use the likelihood derived by Raftery and Zeh (1998), which is proportional to the density of $\exp(.0302 + .0069t_8) - 1$, where t_8 is a random variable that has a t-distribution with 8 degrees of freedom. This is a likelihood only; there is no prior component.

The catch series, $\{C_t\}$, for 1848–1993, used is the one accepted by the Scientific Committee for bowhead whales.

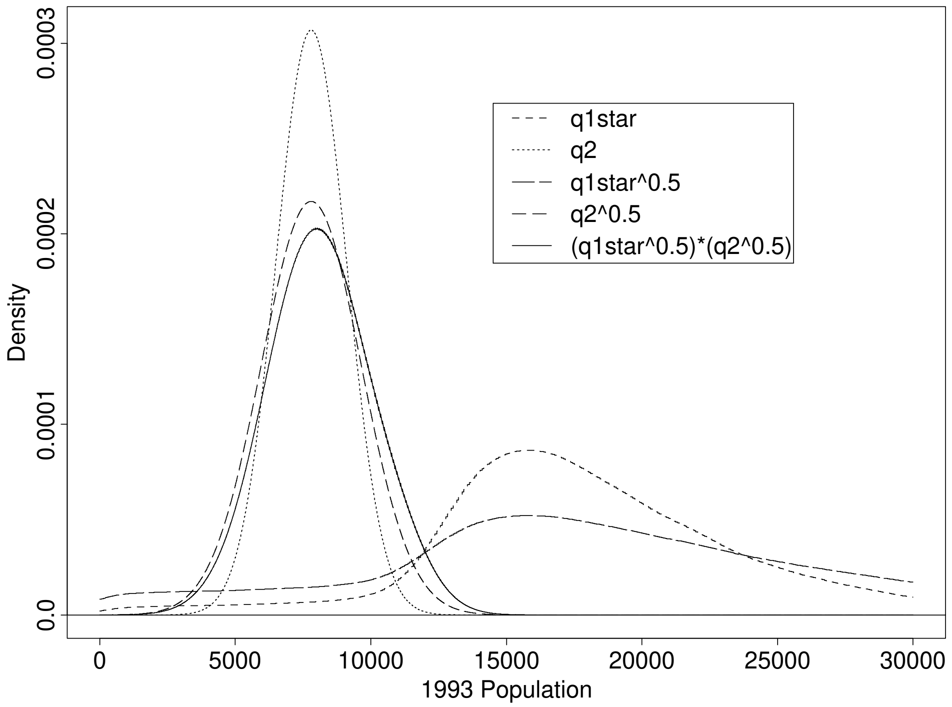


Figure 4: The components of the geometrically pooled prior distribution of the output $\phi = P_{1993}$ in the simple PDM. The final pooled prior density, $\tilde{q}^{[\phi]}(P_{1993})$ is the solid line.

Figure 4 shows the components of geometric pooling for the simple bowhead whale model we are using. It shows the different prior distributions for the output, P_{1993}, and how these are modified and put together to obtain the final geometrically pooled prior distribution. The distribution on P_{1993} induced by the model from the priors on the inputs, q_1^*, is very spread out, as one might expect. The existing prior, q_2, is far more precise. The resulting pooled prior is slightly more spread out than q_2.

Table 1 displays posterior quantiles (medians and bounds of the 95% Bayesian interval) and corresponding quantiles of the pre-model distributions. Here, as in the original Bayesian synthesis method, the term *pre-model* applies to the information (priors or likelihoods) elicited independently of the PDM. The posterior variance of P_0 is considerably lower than the prior variance; similarly

Table 1: Posterior and pre-model quantiles of inputs and outputs of the simple PDM.

Parameter	Quantile	Results	
		Posterior	Pre-model
P_0	.025	12057	8264
	.5	14346	15004
	.975	17980	30373
MSYR	.025	.0113	.0096
	.5	.0213	.0211
	.975	.0333	.0394
P_{1993}	.025	7072	7095
	.5	8196	8200
	.975	9322	9305
ROI	.025	.0105	.0146
	.5	.0204	.0307
	.975	.0318	.0469

the posterior distribution of MSYR is more precise than the prior. The posterior distribution of P_{1993} is almost identical to the pre-model Bayes empirical Bayes posterior. This is unsurprising since the likelihood is dominant; this was the best piece of pre-model information, and the priors on other parameters affect the inference very little. Finally, the melding of information leads to a lower estimate of ROI than is suggested by the pre-model likelihood. The joint posterior also reveals relationships among parameters that were not previously apparent; indeed, there is a fairly strong negative correlation between P_0 and MSYR in the posterior sample.

4 Diagnostic Checking and Model Validation

4.1 Basic Ideas

How can we assess whether the model considered is adequate for our purposes? An adequate model will allow the selection of plausible inputs that produce plausible outputs, where plausibility is measured by the priors and/or likelihoods for the inputs and outputs. Another way of saying this is that for any input, output or quantity of interest, the various sources of information relating to that quantity should imply substantially overlapping sets of values.

There may thus exist up to four sources of information, as follows:

Source of Information	Symbol	Input Information	Output Information
Prior distribution of inputs	$q_1(\theta)$	$q_1(\theta)$	$q_1^*(\phi)$
Prior distribution of outputs	$q_2(\phi)$	$q_2^*(\theta)$	$q_2(\phi)$
Likelihood of inputs	$L_1(\theta)$	$L_1(\theta)$	$L_1^*(\phi)$
Likelihood of outputs	$L_2(\phi)$	$L_2^*(\theta)$	$L_2(\phi)$

Our proposed approach to model validation consists essentially of checking, for each input, output, or quantity of interest, whether there are enough values favored by each of the four sources of information. For each source, we will compute or sample from a distribution of each input, output and quantity of interest, and check whether these overlap substantially by inspecting side-by-side boxplots.

We now consider the definition and computation of the quantities in the table above. The prior distribution of the inputs, $q_1(\theta)$, induces a distribution of the outputs, $q_1^*(\phi)$; these have already been defined. The distribution, $q_2^*(\theta)$, of inputs implied by the prior distribution of outputs, $q_2(\phi)$, is a little trickier because the model may be non-invertible. We define $q_2^*(\theta)$ to be the pooled prior, $\tilde{q}^{[\theta]}(\theta)$, defined by (18), when $\alpha = 0$. This can be thought of as the (pooled) prior on the inputs when no weight is given to the prior distribution of the inputs. When the model is one-to-one, $q_2^*(\theta)$ is induced entirely from the prior distribution of the outputs. When it is not one-to-one, the prior distribution of the inputs is used to divide the mass induced by the prior on outputs among values of the input. Since $\alpha = 0$, the prior on the inputs contributes no absolute mass to the pooled prior.

The likelihoods, L_1 and L_2, do not necessarily define distributions on inputs and outputs, but we propose the following as an informal way of determining the regions of input space and output space favored by the likelihoods:

1. Define a distribution of the inputs, $g(\theta)$, and draw a sample, $\{\theta_1, \ldots, \theta_k\}$, of values of the inputs from it. If $q_1(\theta)$ is broad enough, then $g(\theta) = q_1(\theta)$ may be a good choice. If not, $g(\theta) = q_1(\theta)^\delta$ for some $0 < \delta < 1$ may be adequate, or some very broad distribution may be preferred.

2. Define weights $w_i = L_1(\theta_i)/g(\theta_i)$ (if we are considering the likelihood of the inputs), or $w_i = L_2(M(\theta_i))/g(\theta_i)$, if we are considering the likelihood of the outputs.

3. Resample from $\{\theta_1, \ldots, \theta_k\}$ with probabilities proportional to the weights w_i.

4. For each resampled value of θ, compute the corresponding value of $\phi = M(\theta)$.

This algorithm yields four samples, corresponding to the last four distributions in the table above, namely distributions of the inputs and of the outputs corresponding to the likelihood of the inputs and the likelihood of the outputs respectively. Will these be proper distributions? $L_1(\theta)$ will define a proper distribution if $\int L_1(\theta)d\theta < \infty$. We denote by $L_1^*(\phi)$ the distribution on Φ induced by the model and by the distribution $L_1(\theta)$; this will be proper if $\int L_1^*(\phi)d\phi < \infty$. The distribution of inputs induced by L_2 will be denoted by $L_2^*(\theta)$ and will be proper if $\int L_2(M(\theta))d\theta < \infty$.

Finally, if the component likelihoods of L_2 are independent, then $L_2(\phi)$ can be written down and sampled from directly; it will be proper if $\int L_2(\phi)d\phi < \infty$. If not, however, computing $L_2(\phi)$ for a given output may not be so straightforward, and one may have to resort to the less direct importance sampling method described above. The resulting distribution will be proper if $\int L_2(\phi)d\phi < \infty$ and if $\int g^*(\phi)d\phi < \infty$, where $g^*(\phi)$ is the distribution on Φ induced by the model and $g(\theta)$. Even if the distributions induced by the likelihoods are not proper, it seems that this method could be used informally in an exploratory sense to detect conflicts between input and output likelihoods, priors and the model. It could be viewed as indicating the values favored by the likelihoods within the region to which $g(\theta)$ gives most of its probability.

At this point, we are not making specific formal recommendations for interpreting the boxplots. Rather, in the spirit of informal diagnostic checking, we suggest that users inspect the boxplots for evidence of conflict, and hence failure of the model to reconcile or accommodate the various sources of evidence. In the next subsections, we show one example where these plots are at least moderately reassuring, and another example where they clearly indicate model failure. In both of these cases, we have based our discussion purely on visual inspection of the side by side boxplots, but more formal ways of calibrating the plots might be useful for drawing more formal conclusions about model adequacy. We leave this topic for future research, however.

A conflict in these plots indicates that the model fails to reconcile the different sources of information; this could be due to model inadequacy or to poor information. The possible remedies are to improve the model, and/or to reconsider and perhaps replace or discard some of the information being used.

Here we are proposing boxplots as the way to inspect the distributions corresponding to the different sources of information. We have experimented with various ways of looking at these distributions, and the boxplots seemed the most informative for our purposes. However, there are clearly many other ways of doing this, and some of these might be thought better than boxplots for other situations, or by other data analysts. In particular, ways to look at the multivariate distributions implied by the different sources of information could be very useful, but we have not pursued this topic here.

4.2 Model Validation for the Simple Bowhead Whale Population Dynamics Model

We now apply these ideas to Example 2, the simple population dynamics model analyzed in Section 3.6. Figure 5 shows boxplots of samples from the distributions implied by the various sources of information for the inputs (P_0 and MSYR), the output (P_{1993}), and a key quantity of interest, the 1978–1993 rate of increase (ROI). There is no likelihood for the inputs, and so there are only three sources of information rather than four: the prior on inputs, the prior on the output, and the likelihood on ROI.

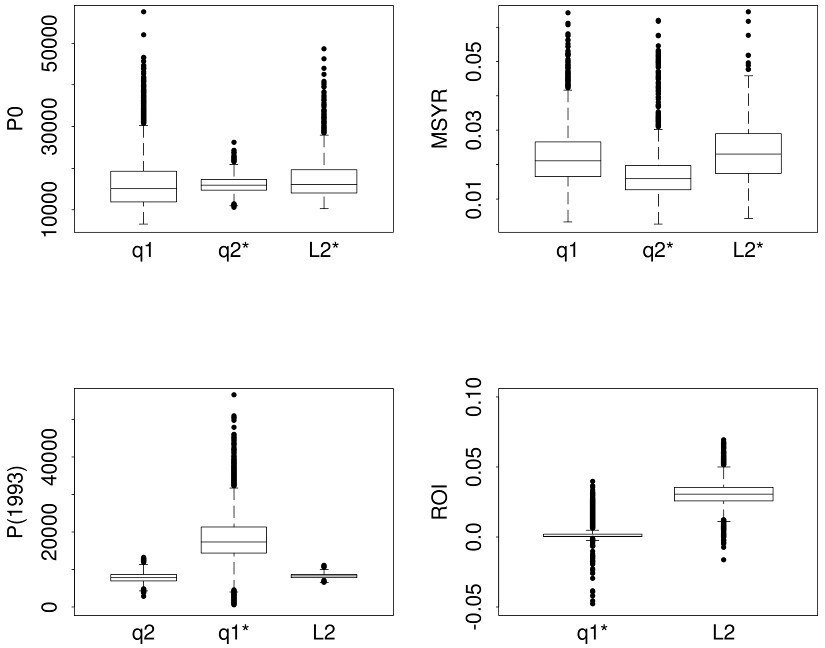


Figure 5: Boxplots of the distributions of inputs (P_0 and MSYR), output (P_{1993}) and rate of increase (ROI) implied by the prior on inputs (q_1 or q_1^*), the prior on the output (q_2 or q_2^*) and the likelihood on the output (L_2) for the simple population dynamics model for bowhead whales.

The box in the center of the boxplot represents the interquartile range, with the line in the middle showing the median. The "whiskers" are drawn out either to the minimum or maximum or to a length of 1.5 times the interquartile range, whichever is the shorter; points further out are plotted individually.

The three distributions of the inputs, P_0 and MSYR, overlap nicely. For the output, P_{1993}, the distribution implied by the prior on the inputs, $q_1^*(P_{1993})$, covers the prior on P_{1993} itself, $q_2(P_{1993})$, which in turn covers the distribution implied by the likelihood, which in this case is clearly proper. Thus the three distributions overlap; the overlap consists of the region favored by the likelihood, which is much more precise than either prior in this case. Note that the region favored by the likelihood is completely outside the interquartile range of the distribution implied by the prior on inputs, q_1^*, but this is not a concern as q_1^* is spread out enough to encompass the likelihood-favored region comfortably.

In the case of ROI, most of the region favored by the likelihood is contained within the prior upper tail. The overlap is sufficient for Bayesian melding to yield reasonable answers, and could not be characterised as a stark conflict, but it is still worth further investigation and consideration. Such an exercise could embrace all the components: the model, the priors and the likelihood. For example, one could engage in a second stage of elicitation in which this result was communicated to the Scientific Committee and they were asked whether they wished to revise their prior distributions on the inputs (which induce the prior q_1^*) so as to give more weight to values indicating a more productive stock.

Overall, while there are some concerns that the model or prior distributions may tend to favor a less productive population too much, there seems to be enough overlap for one to be able to use the model in this case. In reality, a more complex model was used for the actual assessment, in which the overlap between prior and likelihood for ROI was more satisfactory; this is described in Section 7.

4.3 Model Validation for a Simple Gray Whale Population Dynamics Model

To show how our proposed methods can diagnose clear model failure, we consider Example 4. This is the same simple population dynamics model, but for a different species, the eastern Pacific gray whale, *Eschrichtius robustus*. We use prior distributions proposed by Wade and DeMaster (1996) on the basis of the available biological information (Reilly, 1984, 1992); this may be the whale species for which abundance levels and trends are most precisely known. We use the catch history that has been agreed by the IWC, and is listed by Butterworth *et al.* (1992, Table 2). In many ways, the gray whale situation is similar to that of bowheads: the species is subject to aboriginal subsistence whaling, commercial whaling started in 1846 and rapidly depleted the stock, and recent scientific surveys started in 1968. But there are also crucial differences.

The inputs to our model are P_0, which denotes the stock size in 1846, and MSYR, while the output is P_{1988}. The priors are Uniform[0.01,0.13] for MSYR and Uniform[17000,70000] for P_0. Independent Normal likelihoods for the abundance in 16 of the years between 1968 and 1988 are also available.

The model validation boxplots are shown in Figure 6. For the input MSYR and the output P_{1988} the overlap is satisfactory. For the input P_0, however, the region favored by the likelihood is entirely concentrated at the extreme lower value of a sample from the prior for P_0; this is rather unsatisfactory. The most striking feature of the figure is that the prior distribution of ROI (induced by the prior on inputs) and the likelihood on ROI do not overlap at all, and indeed are in clear conflict. It is obvious that the model is not adequate to reconcile the various sources of information.

This lack of fit has already been noticed and discussed (Reilly, 1984; Butterworth *et al.*, 1992), but it is immediately apparent from a routine use of our model validation method. One view of

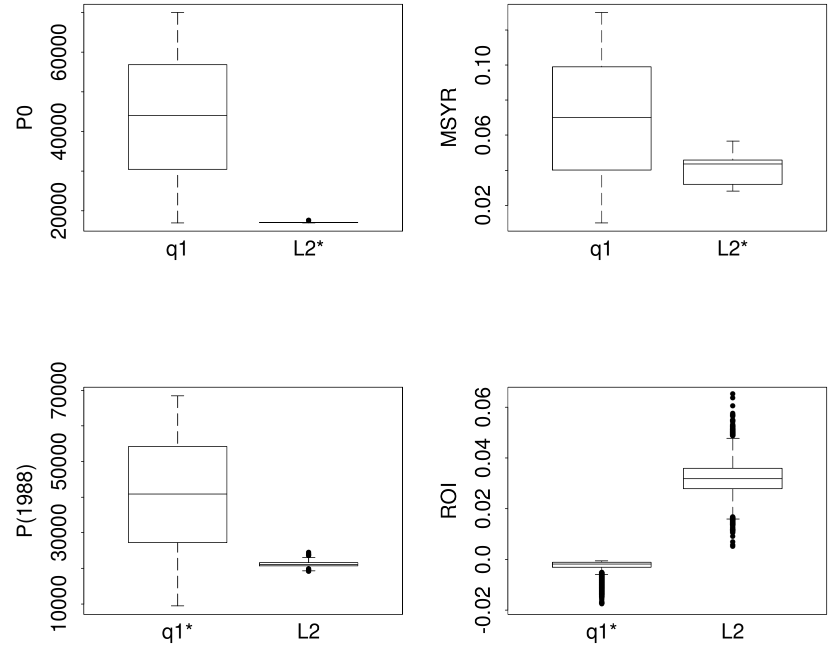


Figure 6: Boxplots of the distributions of inputs (P_0 and MSYR), output (P_{1993}) and rate of increase (ROI) implied by the prior on inputs (q_1 or q_1^*), the prior on the output (q_2 or q_2^*) and the likelihood on the output (L_2) for the simple population dynamics model for gray whales.

the conflict is that it may be caused not by failure of the model but by poor and perhaps biased data on historic catches (Wade and DeMaster, 1996). These authors kept the model but discarded the historic catch data, and when this is done the model seems to fit the more recent information quite adequately.

5 Hypothesis Testing and Model Selection

We now consider the problems of hypothesis testing and model selection. In hypothesis testing, we assume that there is both a null hypothesis and an alternative hypothesis, so that we are in a Neyman-Pearson type of testing situation rather than the kind of pure significance testing situation envisaged by R.A. Fisher. The latter problem, which is related to that of judging the overall adequacy of the model, may be addressed in part by more informal diagnostic checking methods of the kind discussed in the previous section.

We adopt a Bayesian approach to the hypothesis testing situation, based on Bayes factors (Jeffreys, 1961; Kass and Raftery, 1995). If we are testing a statistical model M_0 against an alternative model M_1, the Bayes factor for M_1 against M_0 is defined as the posterior odds for M_1

against M_0 when the prior odds are equal to one. This is

$$B_{10} = p(D|M_1)/p(D|M_0), \tag{20}$$

where $p(D|M_k)$ is the integrated likelihood of model M_k defined by

$$p(D|M_k) = \int p(D|\theta_k, M_k) p(\theta_k|M_k) d\theta_k, \tag{21}$$

D denotes the data, and θ_k is the vector of parameters for model M_k ($k = 0, 1$). Equation (21) says that the integrated likelihood is the integral over the parameter space of the ordinary likelihood, $p(D|\theta_k, M_k)$, times the prior density, $p(\theta_k|M_k)$, or, equivalently, that it is a weighted average of the likelihood values, weighted by their prior probabilities. The Bayes factor, B_{10}, is interpreted as a measure of evidence for M_1 against M_0; values below 1 indicate evidence for M_0, while values greater than 1, 3, 20 and 100 correspond respectively to weak, positive, strong and very strong evidence for the alternative hypothesis M_1. This testing framework does not require M_0 and M_1 to be nested models, unlike standard significance testing.

How can we extend this standard framework to the deterministic models we are concerned with here? We do this by defining the parameters of the model to be the inputs, the prior distribution to be the pooled prior on the inputs, and the likelihood to be the product of the likelihoods on inputs and outputs, if both are present. Thus equation (21) is replaced by

$$p(D|M_k) = \int \tilde{q}^{[\theta]}(\theta) L_1(\theta) L_2(M(\theta)) d\theta. \tag{22}$$

This can be estimated by simple Monte Carlo integration (Hammersley and Handscomb 1964); see McCulloch and Rossi (1991) and Raftery and Banfield (1991) for its application to Bayes factors. The algorithm is:

1. Generate a sample $\theta_1, \ldots, \theta_\ell$ from the pooled prior on inputs $\tilde{q}^{[\theta]}(\theta)$.
2. For each θ_i, compute $L(\theta_i) = L_1(\theta_i) L_2(M(\theta_i))$.
3. Estimate $p(D|M_k)$ as the average of the likelihoods of the sampled input values, i.e. $\ell^{-1} \sum_{i=1}^{\ell} L(\theta_i)$.

This can often be computed as an almost cost free byproduct of the algorithm for implementing the Bayesian melding analysis, which can easily produce a sample from $\tilde{q}^{[\theta]}(\theta)$ and usually computes the likelihoods of the sampled values. It may be somewhat inefficient, but it is a simple unweighted average and so its standard error is readily available; this can be used to monitor its accuracy and to decide how many samples ℓ are needed (e.g. Lewis and Raftery, 1997).

To illustrate the use of the Bayes factor, we applied it to test for the presence of density dependence in the bowhead whale case. Intuitively, density dependence implies that the rate of increase of the population decreases as the population approaches its carrying capacity, taken here to be equal to P_0. We formulated the testing problem by designating the non-density-dependent model of Example 3, given by equation (2) as the model M_0, and the density dependent model (1) of Example 2 as M_1. The inputs, outputs, priors and likelihoods for (2) were taken to be the same as for the model (1) with density dependence, and the Bayesian melding approach was implemented for (2) in the same way as before.

The resulting Bayes factor was 0.7, slightly favoring the null non-density dependence model. The evidence is very slight, however. This is not too surprising since the density-dependent analyses suggest that the bowhead population was far below its carrying capacity throughout the 1978–1993 period to which the likelihoods refer, and so it would not have been within the range where density dependence could have been detected. In the gray whale case (Example 4), the Bayes factor did provide some evidence for density dependence (Wade and DeMaster, 1996), which is less surprising since the gray whales are probably closer to their carrying capacity.

The same framework can be used for model selection. If one has several models (often corresponding to different combinations of modelling assumptions), their posterior probabilities will be proportional to the product of their integrated likelihoods with their prior model probabilities, $p(M_k)$ (the latter are often taken to be equal so as not to favour one model over another *a priori*). One natural choice would then be to select the model with the highest posterior probability.

More generally, one might wish to take account of model uncertainty when making inferences about quantities of interest. This can be done formally in the Bayesian framework using Bayesian model averaging. If Δ is a quantity of interest, such as a biological parameter (e.g. the calf mortality rate), a future observation (e.g. the estimated population in a future survey or census), or a consequence of a course of action (e.g. the number of whales in 100 years under a given management scheme), then the posterior or predictive distribution of Δ under Bayesian model averaging is

$$p(\Delta|D) = \sum_{k=1}^{K} p(\Delta|D, M_k)p(D|M_k)p(M_k) / \sum_{k=1}^{K} p(D|M_k)p(M_k). \tag{23}$$

In equation (23), $p(\Delta|D, M_k)$ is the posterior or predictive distribution of Δ under model M_k, defined by

$$p(\Delta|D, M_k) = \int p(\Delta|\theta_k, D, M_k)\pi^{[\theta]}(\theta|M_k)d\theta, \tag{24}$$

where $\pi^{[\theta]}(\theta|M_k)$ is the pooled posterior distribution defined by (8) for model M_k. In equation (24), $p(\Delta|\theta_k, D, M_k)$ may be a point mass rather than a probability distribution because the models are deterministic, but $p(\Delta|D, M_k)$ and $p(\Delta|D)$ will still be probability distributions giving valid uncertainty statements about Δ.

One interesting possible application of (23) arises in the context of exploring future scenarios by simulation. This is done by the IWC for assessing the likely consequences of different management regimes for whales. While they explore many different scenarios, they tend to condition on a single selected population dynamics model. Realistic model uncertainty could be introduced into this process by using (23) to simulate, i.e. by sampling from different population dynamics models with frequencies proportional to their posterior model probabilities. Draper (1995) has examined this possibility in a different modelling context.

6 Relabelling and Joint Pooling

6.1 Relabelling

Thus far, we have considered simulation models of the form $M : \theta \rightarrow \phi$ where θ is a vector of input parameters and ϕ is a vector of output parameters. This is a general framework in which the methods of the previous sections are applicable.

Occasionally, there may be reasons for which scientists wish to relabel some of the inputs as outputs and vice versa. In other words, there may be a desire to reformulate the model so that a quantity that was originally an input is now an output, and a corresponding quantity that was originally an output is now an input. We refer to this as *relabelling* of the model. Relabelling should not be confused with the reparameterizations of inputs and outputs that were considered in the context of the Borel paradox.

In the event of model relabelling, it is useful to refine the notation slightly. Let M_1 and M_2 be two different labellings of a model M. Denote by

λ : those parameters which are strictly inputs of M_1 and outputs of M_2,

τ : those parameters which are strictly outputs of M_1 and inputs of M_2,

γ : those parameters which are inputs to both M_1 and M_2.

and let $q_1(\lambda)$, $q_2(\gamma)$ and $q_3(\tau)$ be the independent prior distributions on λ, γ and τ respectively. The two labellings of M can then be written as

$$\begin{aligned} \tau &= M_1(\lambda, \gamma) \\ \lambda &= M_2(\tau, \gamma), \end{aligned} \tag{25}$$

and both labellings are of the form considered earlier. For example, M_1 has input $\theta = (\lambda, \gamma)$ and output $\phi = \tau$. We have assumed here that there are no quantities with priors which are outputs under both labellings, and that the large transformation $(\lambda, \gamma) \rightarrow (\tau, \gamma)$ is one-to-one. This is the case in the applications we consider.

6.2 Forwards and Backwards

An instance of relabelling arises in the context of population dynamics models that produce trajectories of abundance over time. Typically, the initial stock size P_0 is one of the inputs to the model (as in the simple PDM of Section 3.6), and the population is projected forwards in time to the present.

Alternatively, the current abundance can be thought of as an input, and the model can be "run" backwards in time to the initial year. In practice, this usually involves solving for the value of P_0 that must (for the fixed values of the other parameters) have given rise to the value of current abundance under consideration.

These two labellings are referred to as the forwards and backwards variants of the PDM. All other parameters are inputs to both variants. If we let M_1 correspond to the forwards variant, then we have P_0 (λ), current abundance (τ) and all remaining inputs (γ) as the division of parameters in (25). Forwards and backwards variants of the BALEEN II model for bowhead assessment have been discussed at length in the Scientific Committee of the IWC (Butterworth and Punt, 1995; Punt and Butterworth, 1996; Givens and Thompson, 1996; Bravington, 1996; Raftery and Poole, 1997).

6.3 Joint Pooling

Under relabelling, a little care needs to be taken in identifying the distributions that are to be pooled. If only one labelling (say M_1) is under consideration, the induced prior on the model output τ, $q_1^*(\tau)$, is pooled with the independently obtained prior $q_2(\tau)$ in the manner described earlier. However, when an alternative labelling M_2 is proposed, the *induced joint prior* of τ and γ, say $q^*(\tau, \gamma)$, needs to be compared with the existing joint prior $q(\tau, \gamma) = q_2(\tau)q_3(\gamma)$, which is the prior on inputs under M_2. In other words, M_1 induces a joint prior on all the parameters that are listed as inputs under M_2, those that are explicit outputs of M_1 and those that are inputs under both labellings. It is this joint prior that should be pooled with the existing input prior on those same parameters.

Marginalization before pooling represents a loss of the information in the joint induced prior. More importantly, if the joint induced prior is ignored (and pooling is thus performed on marginals only), the final posterior will depend on the labelling that is used, i.e. results under M_1 will differ from those under M_2. This is intuitively undesirable since the prior information (as expressed in all the marginal prior distributions) is the same under both labellings. We refer to pooling of the relevant joint prior distributions as joint pooling. Joint pooling yields a unique posterior distribution under a relabelling of the form (25). To see this, we apply (18) under M_1 and (9) under M_2 (since inputs and outputs are reversed) and verify that the result is the same. Let J be the Jacobian of the transformation $(\lambda, \gamma) \to (\tau, \gamma)$. Under labelling M_1, (18) yields

$$\begin{aligned} \tilde{q}(\lambda, \gamma) &\propto q_1(\lambda)q_2(\gamma)\left(\frac{q_2(\gamma)q_3(M_1(\lambda, \gamma))}{q^*(M_1(\lambda, \gamma), \gamma)}\right)^{1-\alpha} \\ &= q_1(\lambda)^{\alpha}q_2(\gamma)q_3(M_1(\lambda, \gamma))^{1-\alpha}|J|^{\alpha-1}_{\tau=M_1(\lambda,\gamma)} \end{aligned} \tag{26}$$

since $q^*(M_1(\lambda,\gamma),\gamma) = q_1(\lambda)q_2(\gamma)|J|_{\tau=M_1(\lambda,\gamma)}$. Under M_2, we apply (9) with pooling weight β to obtain

$$\begin{aligned}\tilde{q}(\lambda,\gamma) &\propto q^*(\lambda,\gamma)^\beta q_1(\lambda)^{1-\beta} q_2(\gamma)^{1-\beta} \\ &= q_1(\lambda)^{1-\beta} q_2(\gamma) q_3(M_1(\lambda,\gamma))^\beta |J|^{-\beta}_{\tau=M_1(\lambda,\gamma)} \qquad (27)\end{aligned}$$

which, on substituting $\beta = 1-\alpha$, is identical to (26). So as long as the weights are assigned consistently, the pooled prior is the same regardless of whether M_1 or M_2 is used as the model.

6.4 Accurate High-Dimensional Density Estimation

For joint pooling the dimensionality of the induced joint prior will be higher than when pooling is restricted to marginal distributions. The SIR algorithm of Section 3.4 involves estimation of this induced density, so it follows that an application of joint pooling may require estimation of a high-dimensional induced distribution. For example, if the labelling M_1 is used in (25), joint pooling requires estimation of $q^*(\tau,\gamma)$ instead of simply $q^*(\tau)$.

Although kernel density estimation is easily extended to higher dimensions, the estimate can become less accurate with increasing dimensionality, particularly if the PDM is complex. Fortunately, a high-dimensional estimate can be avoided in the present context.

Consider the labelling M_1 in (25). The joint prior on the inputs (λ,γ) is $q(\lambda,\gamma) = q_1(\lambda)q_2(\gamma)$ and M_1 induces a density $q^*(\tau,\gamma)$ on (τ,γ). This density will typically be complex and hence difficult to estimate accurately. However, we can express it in the form

$$q^*(\tau,\gamma) = q^*(\tau \mid \gamma) q_2(\gamma), \qquad (28)$$

where $q^*(\tau \mid \gamma)$ is the conditional density of τ given γ and $q_2(\gamma)$ is the (known) density of γ. The problem can therefore be reduced to estimation of $q^*(\tau \mid \gamma)$, a lower-dimensional distribution than $q^*(\tau,\gamma)$. If a representative sample from $q^*(\tau \mid \gamma)$ could be obtained, it could be estimated using the kernel method. For a fixed value of γ, say γ_0, we propose the following algorithm:

1. Construct a representative reference sample of size n from $q(\lambda \mid \gamma_0) = q_1(\lambda)$, since λ and γ are independent before M_1 is imposed. If λ is one-dimensional, we suggest taking ℓ quantiles of $q(\lambda \mid \gamma_0)$ corresponding to evenly-spaced probabilities to best capture the shape of the distribution. Denote this sample by $\lambda_{i,\gamma_0}, i = 1,\ldots,\ell$.

2. Obtain $\tau_i = M_1(\lambda_{i,\gamma_0},\gamma_0)$. This is a representative sample from $q^*(\tau \mid \gamma_0)$ and can be used to form a kernel estimate of the density for any given value of τ.

7 Example : Bowhead Whale Assessment

7.1 The BALEEN II Model

The BALEEN II model used by the IWC for bowhead whales and several other species (our Example 5) is a special case of the one-sex age-structured Leslie matrix population projection model (Leslie, 1945; de la Mare, 1989). It is given by

$$n_{1,t+1} = \sum_{x=1}^{\infty} b_{xt}(n_{xt} - c_{xt}), \tag{29}$$

$$n_{x+1,t+1} = r_x(n_{xt} - c_{xt}), \tag{30}$$

where n_{xt} is the number of females aged x next birthday at the start of year $(t - 1848)$, b_{xt} is the average number of female calves born in year t to a female aged x that survive to age 1, r_x is the natural survival rate of females aged x, and c_{xt} is the number of females aged x killed by hunting in year t.

We denote by $P_t = \sum_{x=1}^{\infty} n_{xt}$ the population in year t aged 1 and above. For convenience, all those aged w and above are grouped into a single age class (the choice of w does not affect the results so long as it is large enough). The initial condition is $P_0 = K$, the carrying capacity, corresponding to a pristine stock before commercial hunting started in 1848.

It is assumed that a juvenile survival rate, s_0, applies up to age a, and that thereafter a (higher) adult survival rate, s, applies, so that $r_x = s_0$ if $x = 1, \ldots, a$, and $r_x = s$ if $x \geq a + 1$, where $s_0 < s$. Fertility is assumed to be constant with respect to age beyond the age at sexual maturity (ASM), m, but to be density-dependent, so that it depends on the population at time t. This is specified as follows:

$$b_{xt} = \begin{cases} 0 & \text{if } x = 1, \ldots, m \\ f_t & \text{if } x \geq m + 1. \end{cases} \tag{31}$$

In equation (31),

$$f_t = f_0 + (f_{\max} - f_0)[1 - (P_t/K)^z], \tag{32}$$

where f_0 is the fertility rate when the population is at its carrying capacity, $f_{\max}$ is the maximum fertility, attained when the stock is near extinction, and z is the density-dependence parameter.

We have expressed the model here in terms of the eight quantities s_0, s, a, f_0, $f_{\max}$, m, z, and K. However, the IWC has reformulated it in terms of management quantities that they find more interpretable. These are based on the concept of maximum sustainable yield (MSY). A stock can indefinitely sustain any level of catch less than the MSY, and the MSY level (MSYL) is the equilibrium population corresponding to indefinite annual catches of size MSY, expressed as a proportion of K. The MSY rate (MSYR) is the natural increase (= births − deaths) when the population is at MSYL, expressed as a proportion of the population.

Punt (1998) has pointed out that of the eight quantities listed above, only seven need be specified. The Scientific Committee chose as inputs to the BALEEN II model the seven quantities $\theta = (\mathrm{MSYR}, \mathrm{MSYL}, s, a, f_{\max}, m, K)$. Once θ has been specified, the remaining quantities s_0, f_0 and z can be deduced as follows.

Equilibrium will be achieved even if the stock is subject to hunting, provided that the annual catch is no more than MSY. At equilibrium, the natural growth rate of the population, λ, defined by $P_{t+1} = \lambda P_t$, is given by

$$\lambda^{m+1} - s\lambda^m - s_0^a s^{m-1} f_t[1 - (s/\lambda)^{w-m-1}] = 0 \tag{33}$$

(Breiwick, Eberhardt and Braham, 1984). In the absence of hunting, $\lambda = 1$, and so by (33),

$$1 - s - s_0^a s^{m-a} f_0(1 - s^{w-m-1}) = 0. \tag{34}$$

Also, at MSYL, $\lambda = (1 + \mathrm{MSYR})$, and so

$$(1+\mathrm{MSYR})^{m+1} - s(1+\mathrm{MSYR})^m - s_0^a s^{m-1} f_t^{(\mathrm{MSYL})}[1 - (s/(1+\mathrm{MSYR}))^{w-m-1}] = 0, \tag{35}$$

where, by (31),

$$f_t^{(\mathrm{MSYL})} = f_0 + (f_{\max} - f_0)(1 - \mathrm{MSYL}^z). \tag{36}$$

Finally, at MSYL,

$$\mathrm{MSYL} + \mathrm{MSYR}\left[\left.\frac{\mathrm{d}\_Q(\rho)}{\mathrm{d}\_\rho}\right|_{\rho=\mathrm{MSYR}}\right] = 0, \tag{37}$$

where $Q(\rho)$ is the equilibrium population corresponding to a sustained catch equal to a proportion ρ of the population, where $\rho \leq \mathrm{MSYR}$, and $\frac{\mathrm{d}\_Q(\rho)}{\mathrm{d}\_\rho} = \lim_{h\to 0} h^{-1}[Q(\rho - h) - Q(\rho)]$ is the derivative from the left (Punt, 1998). This derivative was evaluated numerically using finite differences.

Given θ, equations (34), (35), (36) and (37) form a system of four nonlinear equations in the four unknowns s_0, f_0, z and $f_t^{(\mathrm{MSYL})}$, and this can be solved numerically. Given the inputs θ, we can thus reformulate the model in terms of the eight quantities that originally specified it, and so run the model using equations (29) and (30) above.

The formulation described above corresponds to the forwards variant of the BALEEN II model. The carrying capacity K is an input and equations (29) and (30) are used to project the population forwards through time. The backwards variant works in exactly the same way except that current abundance, P_{1993}, rather than K, is an input to the model. For generated values of current abundance and the other input parameters, a numerical method (from, for example, Press *et al.*, 1986) is used to solve for the value of K that results in the given current abundance. The 7-dimensional transform $(\mathrm{MSYR}, \mathrm{MSYL}, s, a, f_{\max}, m, P_{1993}) \to (\mathrm{MSYR}, \mathrm{MSYL}, s, a, f_{\max}, m, K)$ is one-to-one in this case, so the solution for K is unique.

7.2 Pre-model information

We apply the joint pooling method to the assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using prior distributions and likelihoods that were agreed upon by the IWC Scientific Committee at its 1998 annual meeting in Muscat, Oman. A comprehensive bowhead assessment was one of the primary tasks undertaken by the Scientific Committee at that meeting.

Table 2 lists the independent marginal prior distributions for parameters of the BALEEN II PDM. Many of the choices in Table 2 are based on previous work by the Scientific Committee (IWC 1995, p. 148). The prior distribution of the adult survival rate was agreed to be appropriate when no maximum age limit on bowheads is imposed. An alternative formulation places a maximum age bound of 100 years while removing the upper truncation point (0.995) from the prior. The two formulations give similar results.

Table 2: Prior distributions for the BALEEN II model for bowheads.

Parameter	Prior
MSYL	U[0.4, 0.8]
MSYR	U[0.01, 0.07]
Transition age (a)	Discrete U[1, 9]
Age at sexual maturity (ASM)	Grouped N(20,3^2) truncated at 13.5 and 26.5
Adult survival (s)	N(0.99,0.02^2) truncated at 0.995
Max. theoretical preg. rate (f_{max})	$1/f_{max} \sim$ U[2.5,4]
Carrying capacity (K)	$\log(K) \sim$ U[log(7000), log(31000)]
Abundance in 1993 (P_{1993})	N(7800,1300^2)

Bowhead data summarized in likelihood form are listed in Table 3. The 1993 abundance likelihood comes from Raftery and Zeh (1998) while the age data are from IWC (1995). The product of the 1993 abundance prior from Table 2 and the likelihood from Table 3 provides a very close approximation to the Bayes empirical Bayes posterior for this quantity obtained by Raftery and Zeh (1998).

Survey estimates of abundance for the years 1978-1988 (except for 1979 and 1984) are available and are listed in Punt and Butterworth (1998). These are slight modifications of estimates in Raftery and Zeh (1998). The logarithm of these estimates is assumed to follow a multivariate normal distribution. The construction of the associated covariance matrix is described in Punt and Butterworth (1998).

7.3 Implementation and Results

Since the joint pooling posterior does not depend on the variant of the PDM that is used, we could perform the analysis using either variant. However, we chose the backwards variant because it is

Table 3: Likelihoods for the BALEEN II model for bowheads.

Data	Distribution
Proportion of calves (1985-1992)	$0.0164t_5 + 0.052$
Proportion of matures (1985-1992)	$0.0286t_5 + 0.411$
Abundance in 1993 (P_{1993})	$\mathrm{N}(8293, 626^2)$
Abundances (1978-1988)	log(Abundances) $\sim$ Multivariate Normal

more efficient in terms of the SIR algorithm. The forwards variant is rather inefficient in that most of the generated points have almost zero probability of being selected in the resampling step. As a result, a small number of points are resampled multiple times.

For the backwards variant, the density estimation step requires a representative sample from the prior on P_{1993}. We found that 20 quantiles corresponding to equally-spaced probabilities adequately represent the shape of this normal curve. At each iteration of the simulation, a complete set of the seven model inputs (i.e. the six biological parameters and P_{1993}) is generated. The PDM is used to solve for the corresponding value of K. The PDM is also applied to the 20 points in the representative sample. The 20 values of K thus obtained are used to estimate the induced conditional prior, $q^*(K \mid \text{all biological parameters})$, of the true K given the generated set of biological parameters. Applying (28), the required joint induced prior is then simply

$$q^*(K, \text{biol. parameters}) = q^*(K \mid \text{biol. parameters}) q(\text{biol. parameters})$$

where $q(\text{biological parameters})$ is the product of the independent marginal prior distributions on the biological parameters in Table 2.

We chose $\alpha = \frac{1}{2}$ in the joint pooling analysis. This is equivalent to using the geometric mean of the two joint forwards and backwards priors as the joint prior. In this way, information embedded in BALEEN II is allowed to enter the prior. Model inputs are no longer *a priori* independent. The marginal prior distributions are elicited without recourse to the PDM, but the PDM is used to coherize the independent marginal beliefs into a joint prior before introduction of data via the likelihoods. Joint pooling can be viewed as a way of creating a joint prior on model inputs that allows information in the PDM to be reflected in that prior.

Table 4 shows the posterior medians and 90% Bayesian confidence bounds of various quantities. One of the key quantities of interest for management purposes is the 1998 replacement yield, RY_{1998}, the catch from the recruited stock which, if taken, would leave the recruited population at the same level at the beginning of the next season (IWC, 1988). The posterior distribution of RY_{1998} is shown in Figure 7. The 5% point of the distribution is 106 (see Table 4), and this is currently used by the IWC as a conservative upper bound in setting the annual hunting quota.

Figures 8, 9, and 10 show the marginal posterior distributions of the other model inputs, outputs,

Table 4: Posterior medians and 90% Bayesian confidence bounds for various quantities of interest.

Quantity	Quantile		
	5%	50%	95%
RY_{1998}	106	173	233
Total depletion (%)	45.9	63.5	84.1
ROI (1978-1993)	1.01	1.94	3.09
Mature proportion	0.36	0.41	0.46
Calf proportion	0.041	0.055	0.071
K	10761	13214	16813
P_{1993}	7064	7850	8616
MSYL	0.572	0.702	0.789
MSYR	0.012	0.021	0.033
s	0.970	0.984	0.994
s_0	0.807	0.921	0.974
ASM	16	20	25
f_{max}	0.256	0.309	0.388

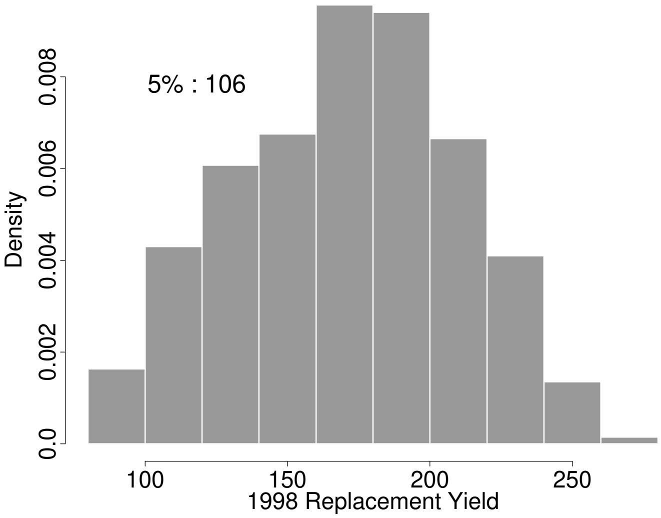


Figure 7: Posterior distribution of 1998 replacement yield.

and quantities of interest listed in Tables 4. Where applicable, the marginal pre-model distributions are shown as solid lines. In some cases, the posterior is very similar to the pre-model distribution; in others there is a substantial shift in location and/or precision. For example, there is evidence that the 1978-1993 ROI is located in the lower tail of the pre-model distribution. The variances of K and s are reduced considerably, but the PDM provides very little additional information concerning a,

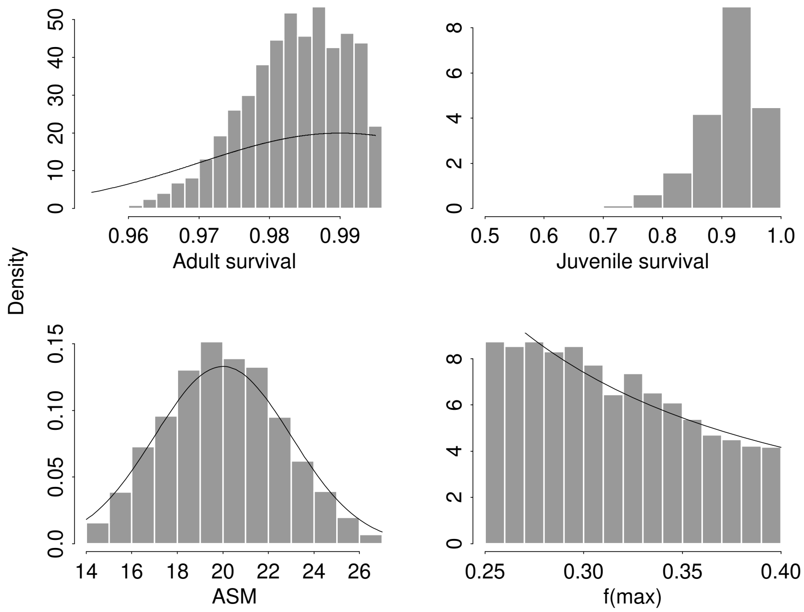


Figure 8: Posterior distributions for s, s_0, ASM, and f_{max}. Priors are shown as solid lines. s_0 does not have a prior in the reference set.

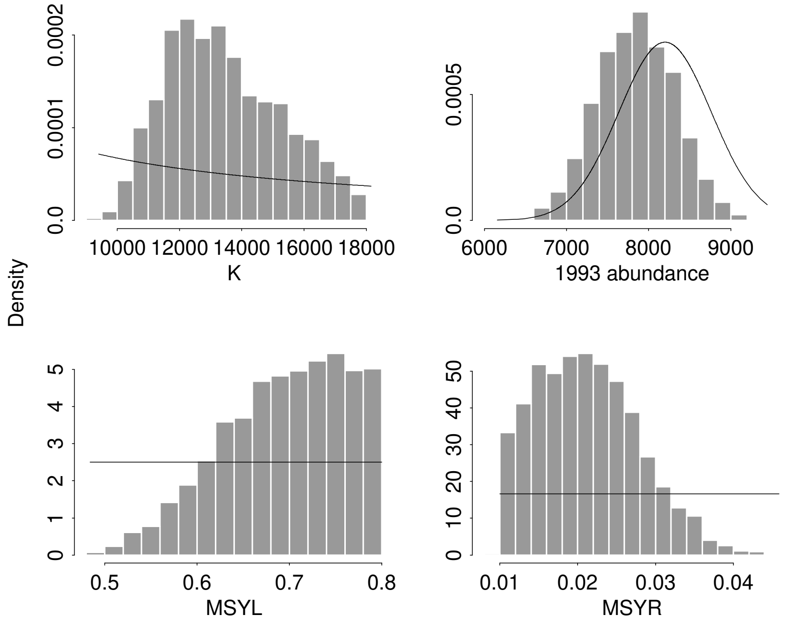


Figure 9: Posterior distributions for K, P_{1993}, MSYL, and MSYR. Pre-model distributions are shown as solid lines.

ASM, or f_{max}. The distribution of total depletion suggests that the stock is almost certainly still depleted in 1998.

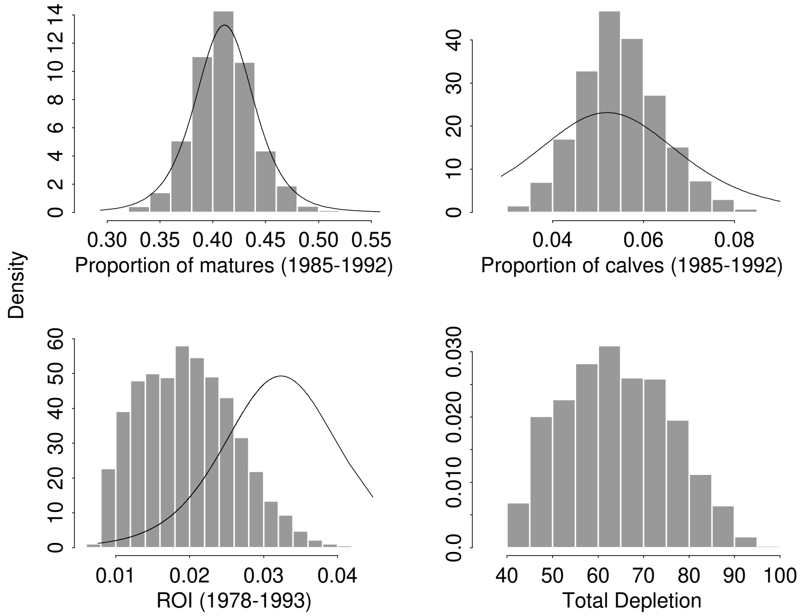


Figure 10: Posterior distributions for mature and calf proportions, ROI, and the levels of depletion. Available pre-model distributions are shown as solid lines. The pre-model distribution for ROI is a likelihood obtained by fitting a least-squares line to the survey abundance estimates.

Figure 11 shows pairwise scatterplots of K, MSYR, ROI and RY_{1998} using 1000 of the 5000 points in the posterior sample. (Using more than 1000 points simply clutters the plots without revealing any further detail.) These are quantities which exhibit high correlation. In particular, there is a strong negative correlation between K and MSYR.

In terms of model validation, we found that the various sources of information agreed better when we used the BALEEN II model than when we used the simpler PDM of Example 2. In particular, Figure 12 shows boxplots of samples from the prior on ROI (induced by the prior on inputs and the BALEEN II model) and the likelihood derived from the survey abundance estimates. Comparing Figure 12 with the corresponding boxplots in Figure 5, we see that the overlap between the sources of information is considerably larger than for the simpler model. The implementation of the BALEEN II model used in this analysis appears to fit the ROI data better than the simpler model.

8 Discussion

We have proposed a new approach to inference from deterministic simulation models, called Bayesian melding. This retains the desirable properties of the Bayesian synthesis approach of RGZ, but modifies it so as to avoid the Borel paradox to which the latter approach is subject. This is done by

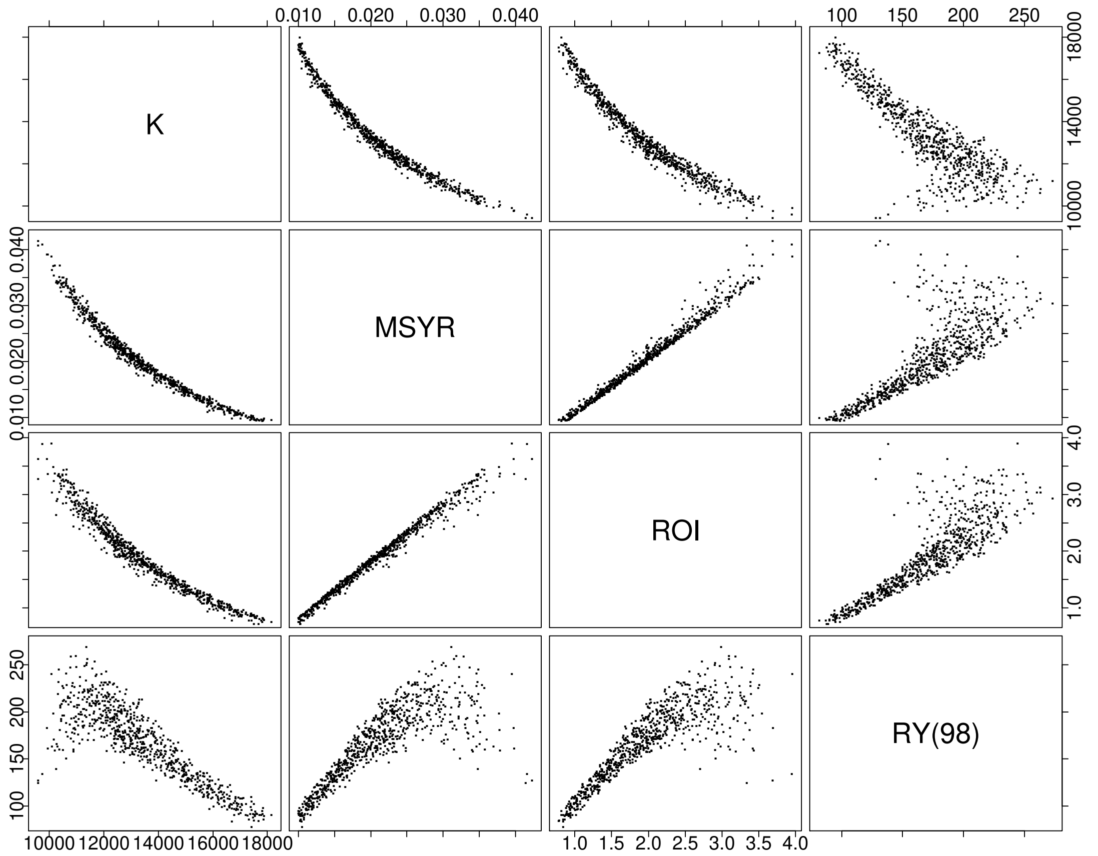


Figure 11: Pairwise scatterplots of the posterior sample of K, MSYR, ROI and RY.

recognising the presence of two different priors on the outputs, one of them explicitly specified, and the other induced by the prior on the inputs and the model; these priors are combined via logarithmic pooling. The method is extended to the difficult but common situation where the model is noninvertible. The resulting approach is formally fully Bayesian, and takes full account of evidence and uncertainty about both the inputs and outputs to the model in a coherent way; it can be viewed as a generalisation of standard Bayesian inference. It was presented to the IWC Scientific Committee in the context of giving advice about bowhead hunting quotas in its 1998 assessment of the stock (at the previous assessment in 1994, the Bayesian synthesis approach had been used).

Our exposition has focused on population dynamics models for whales, which is the context within which it was developed. However, deterministic simulation models are pervasive in science, engineering and policy research, and it seems likely that this approach could find potential applications in many areas. Green *et al.* (1998) have applied the methodology to forest management, and it has potential applications in epidemiology and environmental risk assessment.

One issue that remains unresolved is how to choose α, the parameter that says how the priors on the inputs and the outputs are to be weighted relatively to one another. In the original context

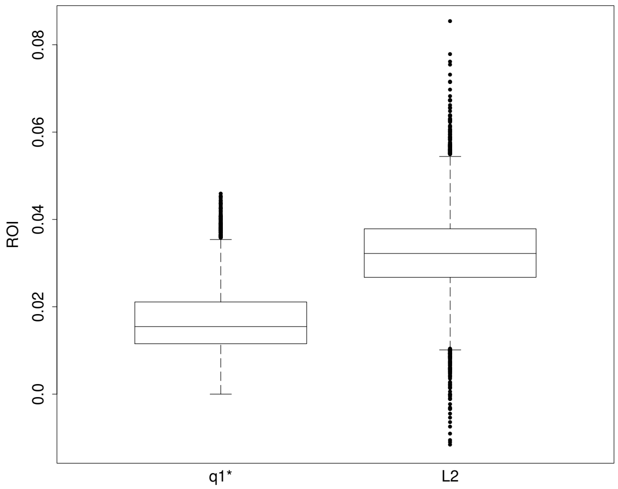


Figure 12: Boxplots of the distributions of the 1978-1993 rate of increase (ROI) implied by the priors on inputs and the BALEEN II model (q_1^*) and the likelihood (L_2) derived from the survey abundance estimates.

in which logarithmic pooling of prior distributions was developed, that of combining the opinions of two experts, α was viewed as reflecting the weight to be attached to the first expert vis-a-vis the second, although a formal basis for specifying its precise value does not seem to have been developed (Genest and Zidek, 1986). Our context is somewhat different, in that we are combining prior distributions about different parameters and based on different bodies of evidence, but assessed by the same "expert" (in our case the IWC Scientific Committee). We have used this fact as the basis for our choice of $\alpha = \frac{1}{2}$ in the examples. Note that α does not reflect the *precision* of the prior information about the inputs relative to that about the outputs, but rather its *reliability*, and it seems reasonable to assume that the information about inputs and outputs will be equally reliable, since they are both assessed and expressed quantitatively by the same expert. Figure 4 illustrates how, even when $\alpha = \frac{1}{2}$, the pooled prior will be dominated by the more precise source of information.

However, values other than $\alpha = \frac{1}{2}$ might well be reasonable. In particular, values at or near the extremes, $\alpha = 0, 1$, might be of interest. A case of particular interest arises when priors on the inputs were specified for formal reasons to make Bayesian inference possible in the absence of much prior information, but the priors on the outputs did correspond to real information. Then setting $\alpha = 0$ or close to it could allow the (uninformative) prior distribution about the inputs to be given little or no weight, while still having a proper prior on the inputs (induced by the prior distribution on the outputs).

An alternative, more subjective, approach would be to plot the pooled priors on both inputs and outputs for each of a set of values of α, and then choose the one that corresponds most closely to actual prior opinion. This could be implemented on the computer using a slider allowing the value of α to be manipulated by the user, while the resulting pooled prior distributions change interactively. The computing requirements of this approach might be an obstacle for many models, but it would seem to be a reasonable ideal.

One could also think of estimating α, for example by computing its integrated likelihood as in Section 5, and maximising this quantity with respect to α. This would provide a maximum integrated likelihood estimator of α. A further possibility would be fully Bayesian, to specify a prior for α itself, and treat it as a parameter in the usual Bayesian way, by integrating over it. In the absence of further information, a Uniform[0,1] distribution would be one possibility. A difficulty with approaches based on estimating α from the likelihoods, however, is that they will tend to give high weight to the prior distributions that least disagree with the likelihoods. This may not be desirable if the prior distributions all represent genuine prior information.

Another open issue is that of how best to do nonparametric density estimation in the present context. We have chosen to use kernel density estimation with a Gaussian kernel, selecting the smoothing parameter by Terrell's (1990) maximal smoothing approach when appropriate. This approach is easy to implement in higher dimensions. However, it is an "off-the-shelf" approach, not crafted specially for our present purpose, and so it should be possible to do better. Givens and Roback (1998) have developed an adaptive method that involves direct numerical approximation of Jacobians. Their method has led to improved accuracy and efficiency in some simple examples. Nonparametric density estimation techniques have generally been developed and evaluated with the quality of the resulting density estimate itself as the ultimate criterion, not surprisingly. However, here our ultimate criterion is the quality of the estimate of the posterior distribution after applying SIR; the quality of the intermediate density estimates is of little interest in its own right. The development of nonparametric density estimation methods specifically for this purpose would seem a worthwhile research challenge.

We have developed our methods for a family of simulation models that can be run reasonably fast on a computer, so that carrying out 50,000 runs, as we have been doing, can be done overnight with little difficulty. Many simulation models take much longer to run, however, and methods that require very large numbers of runs will remain infeasible for such models for a long time. Then efficiency will be a major issue, and methods will be required that make each run of the model count much more. One such approach is the adaptive SIR algorithm of Givens and Raftery (1996). Quadrature methods also seem promising. One possibility is the three-point iterated Gauss-Hermite quadrature of Raftery and Zeh (1993, Appendix 4); see also Levy *et al.* (1998). Another is the Bayes-Hermite quadrature of O'Hagan (1991). A different possibility is Latin hypercube sampling, mentioned in the present context by Schweder (1995).

Our approach can be viewed as involving simulating from a posterior distribution, and Markov chain Monte Carlo (MCMC) methods have been much developed in recent years for this purpose (e.g. Gilks, Richardson and Spiegelhalter, 1996). We have not used MCMC, however, instead choosing to base our methods on the SIR algorithm. The reason is that it is not clear how to implement MCMC in the present context, since it requires the availability of code to compute the (unnormalized) posterior density, which is not available here. Pooling can be implemented in a natural way via the SIR algorithm, but not, as far as we can see, with MCMC. A further difficulty with MCMC in this context is that our posterior distributions tend to be concentrated near nonlinear submanifolds of the parameter space, and it is well known that designing MCMC algorithms for this kind of situation can be extremely delicate. However, if these two obstacles could be overcome and an efficient and reliable MCMC algorithm designed for sampling from the Bayesian melding posterior distribution, this could be very useful as such an algorithm would tend to stay in or close to the high posterior density region and hence could be more efficient than our present proposal. This would seem to be a good topic for further research.

Since Wolpert (1995) pointed out that the Bayesian synthesis approach is subject to the Borel paradox, various other solutions have been proposed. As we have discussed here, the Borel paradox arises from the presence of a prior on outputs as well as inputs, and one proposed solution has been to simply discard the prior on the outputs (Butterworth and Punt, 1995; Punt and Butterworth, 1998). This may be satisfactory in some contexts, but not in others. The specification of a prior on outputs originally arose when it became apparent that plausible values of both inputs and outputs to the BALEEN II model could give rise to highly implausible oscillating trajectories (Givens *et al.*, 1995). The solution proposed at the time in the context of the Bayesian synthesis method was to specify a prior on the output, "Maximum Calf-to-Mother Ratio" (MCM), preventing it from being higher than one, which was a universally agreed biological upper limit. This was very simple and quick to implement, and worked very well to exclude the implausible oscillatory trajectories. Just ignoring this prior on the outputs would have led to unsatisfactory results.

Subsequently, Punt and Butterworth (1996) reparameterized the BALEEN II model in the form we have used here, so that maximum fertility rate became an input, f_{max}, and a prior on this input quantity was enough to avoid the oscillatory trajectories. This is a sensible solution in this case, although it was fairly time-consuming, requiring several years to reprogram the model, validate the code, and obtain approval from the Scientific Committee for the reformulated model. Essentially the same result had been achieved very quickly several years earlier by simply specifying a prior on the relevant output, MCM. With Bayesian melding, this can now be done in a way that is not subject to the Borel paradox. In other complex examples, it may not be feasible to reformulate the model in a way that eliminates the need for a prior on the outputs.

Another approach, proposed in an IWC Scientific Committee meeting by Stephen T. Buckland, is to display the different priors for inputs and outputs on a computer screen, together with the priors they induce on other inputs and outputs as well as other quantities of interest. The user would then interactively modify the priors so that the different priors on the same quantities (explicit prior and induced prior) came closer to agreement. This seems like an excellent idea, although it might be hard to zero in exactly on a fully coherized set of priors using this approach. Our suggestion above of choosing α interactively using a slider could be viewed as one version of this suggestion, that is less flexible but is also guaranteed to yield fully coherized priors.

Another option in the context of expert opinion consensus is the "supra-Bayesian" approach. Here, a group decision maker or meta-expert considers the individual expert opinions as data, and then updates his or her own prior beliefs using Bayes' rule. Lindley (1985) proposed supra-Bayesian pooling when the parameter space is finite. However, most derivations have been for a very simple scenario such as a single event. Extensions to complex simulation modelling would be challenging. Another practical consideration is the choice of the meta-expert. Givens and Roback (1998) have discussed the supra-Bayesian approach in the context of deterministic models.

Schweder and Hjort (1997) have proposed the likelihood synthesis approach, in which prior distributions are viewed as deriving from likelihoods for an implicit data set, which are then turned into "confidence distributions" for quantities of interest. One issue with this approach is how the confidence distributions are to be interpreted, as they do not seem to have either a clear frequentist interpretation or a Bayesian interpretation.

We have discussed the effect of the Borel paradox in the application of Bayesian methods to inference for deterministic models. Other paradoxes have arisen in different contexts within the general framework of Bayesian inference. For example, Dawid *et al.* (1973) examined posterior marginalization paradoxes that result from the use of improper prior distributions.

We have focused here on deterministic simulation models, but our approach could also be applied to stochastic models. For example, in our BALEEN II bowhead PDM, the numbers of births and deaths in a year could be modeled as binomial random variables rather than approximated by their expectations. We would proceed as before, except that instead of applying the model deterministically for each run, we would simulate a stochastic trajectory. Note that even if the model were made stochastic, the Bayesian synthesis method would still be subject to the Borel paradox, and so pooling would still be necessary. Bayesian melding should provide a valid solution for stochastic models as well as deterministic ones.

Acknowledgements: This research was supported by the North Slope Borough, Alaska, U.S.A., and the U.S. National Oceanic and Atmospheric Administration (through the U.S. National Marine Mammal Laboratory and the Alaska Eskimo Whaling Commission). We are grateful to Thomas F. Albert for sustained research support, and to Stephen T. Buckland, Douglas S. Butterworth, Christina Fraley, Geof H. Givens, André E. Punt, Tore Schweder, Robert L. Wolpert and Judith E. Zeh for very helpful comments and discussions. The views expressed remain those of the authors. The historical remarks about the Borel paradox in Section 2.3 are based in part on a personal communication from Stephen M. Stigler reported by Schweder and Hjort (1995).

References

Bertrand, J.L.F. (1882) *Calcul des Probabilités.* Paris: Gauthier-Villars.

Billingsley, P. (1986) *Probability and Measure* (2nd edn.), New York:John Wiley.

Borel, E.F.E.J. (1909) *Eléments de la Théorie des Probabilités.* Paris: Hermann.

Bravington, M.V. (1996) An appraisal of Bayesian synthesis, with suggested modifications and diagnostics. *Rep. int. Whal. Commn*, **46**, 509-529.

Breiwick, J.M., Eberhardt, L.L. and Braham, H.W. (1984) Population dynamics of Western Arctic bowhead whales (*Balaena mysticetus*). *Can. J. Fish. Aquat. Sci.*, **41**, 484-496.

Butterworth, D.S., Korrubel, J.L. and Punt, A.E. (1992) What is needed to make a simple density-dependent response population model consistent with data for the eastern north Pacific gray whales? Paper SC/A90/G10 presented to the IWC Scientific Committee, 1992.

Butterworth, D.S. and Punt. A.E. (1995) On the Bayesian approach suggested for the assessment of the Bering-Chukchi-Beaufort Seas stock of bohwead whales. *Rep. int. Whal. Commn*, **45**, 303-311.

Caswell, H. (1976) The validation problem. In *Systems Analysis and Simulation in Ecology* (ed. B.C. Patten), pp. 313-325. New York: Academic Press.

Dawid, A.P., Stone, M. and Zidek, J.V. (1973) Marginalization paradoxes in Bayesian and structural inference (with discussion). *J. R. Statist. Soc.* B, **35**, 189-233.

de la Mare, W.K. (1989) The model used in the Hitter and Fitter programs (Program: Fitter.SC40), Report of the Scientific Committee, Annex L, *Rep. int. Whal. Commn*, **39**, 150-151.

Diggle, P.J. and Gratton, R.J. (1984) Monte Carlo methods of inference for implicit statistical models (with discussion). *J. R. Statist. Soc.* B, **46**, 193-227.

Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *J. R. Statist. Soc.* B, **57**, 45-98.

Edwards, A.W.F. (1967) A biometric case history: studies on the distribution of sexes in families.

1889-1966. Abstract 1286. *Biometrics*, **23**, 176.

French, S. (1985) Group consensus probability distributions: a critical survey. In *Bayesian Statistics 2* (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith), pp 183-201. Amsterdam: North-Holland.

Genest, C. (1984) A characterization theorem for externally Bayesian groups. *Ann. Statist.*, **12**, 1100-1105.

Genest, C., McConway, K.J. and Schervish, M.J. (1986) Characterization of externally Bayesian pooling operators. *Ann. Statist.*, **14**, 487-501.

Genest, C. and Zidek, J.V. (1986) Combining probability distributions: a critique and an annotated bibliography (with discussion). *Statist. Sci.*, **1**, 114-148.

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

Givens, G.H. (1993) *A Bayesian framework and importance sampling methods for synthesising multiple sources of evidence and uncertainty linked by a complex mechanistic model.* Ph.D. dissertation, Department of Statistics, University of Washington, Seattle.

Givens, G.H. (1995) Confirmation of the 1994 Scientific Committee assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales, and further sensitivity trials. Paper SC/47/AS15 presented to the IWC Scientific Committee. May 1995.

Givens, G.H. and Raftery, A.E. (1996) Local adaptive importance sampling for multivariate densities with stong nonlinear relationships. *J. Amer. Statist. Ass.*, **91**, 132-141.

Givens, G.H., Raftery, A.E. and Zeh, J.E. (1993) Benefits of a Bayesian approach for synthesizing multiple sources of evidence and uncertainty linked by a deterministic model. *Rep. int. Whal. Commn*, **43**, 495-500.

Givens, G.H. and Roback, P.J. (1998) Logarithmic pooling of priors linked by a deterministic simulation model. *Submitted to J. Comput. Graph. Statist.*

Givens, G.H. and Thompson, S.E. (1996) Alternative Bayesian synthesis approaches to Bering-Chukchi-Beaufort Seas bowhead whale stock assessment: Uncertainty in historic catch and hitting with fixed MSYR. *Rep. int. Whal. Commn*, **46**, 509-529.

Givens, G.H., Zeh, J.E. and Raftery, A.E. (1995) Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using the BALEEN II model in a Bayesian synthesis framework. *Rep. int. Whal. Commn*, **45**, 345-364.

Green, E.J., McFarlane, D.W., Valentine, H.T. and Strawderman, W.E. (1998) Assessing uncertainty in a stand growth model by Bayesian synthesis. Submitted to *Forest Science.*

Guttorp, P. and Walden, A.T. (1987) On the evaluation of geophysical models. *Geophysical Journal of the Royal Astronomical Society*, **91**, 201-210.

Hammersley, J.M. and Hanscomb, D.C. (1964) *Monte Carlo Methods.* London: Methuen.

Hodges, J.S. (1987) Uncertainty, Policy analysis and statistics (with discussion). *Statistical Science*, **2**, 259-291.

Hodges, J.S. (1991) Six (or so) things you can do with a model. *Operations Research*, **39**, 355-365.

Hoel, D.G. and Mitchell, T.J. (1971) The simulation, fitting and testing of a stochastic cellular proliferation model. *Biometrics* **27**, 191-199.

Hughes, W.P. (1984) (ed.), *Military Modeling*, Alexandria, VA: Military Operations Research Society.

International Whaling Commission (1988) Report of the sub-committee on protected species and aboriginal subsistence whaling, Appendix 4. Bering-Chukchi-Beaufort sea bowhead population trend simulations. *Rep. int. Whal. Commn*, **38**, 116.

International Whaling Commission (1995) Report of the Scientific Committee, Annex F. Report of the sub-committee on aboriginal subsistence whaling. *Rep. int. Whal. Commn*, **45**, 142-164.

Jeffreys, H.S. (1961) *Theory of Probability*, (3rd edn). Oxford: Oxford Univ. Press.

Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *J. Amer. Statist. Ass.*, **90**, 773-795.

Kolmogorov, A.N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: Springer.

Leslie, P.H. (1945) On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183-212.

Levy, J., Clayton, M.K. and Chesters, G. (1998) Using an approximation of the three-point Gauss-Hermite quadrature formula for model prediction and quantification of uncertainty. *Hydrogeology J.*, **6**, to appear.

Lewis, S.M. and Raftery, A.E. (1997) Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *J. Amer. Statist. Ass.*, **92**, 648-655.

Lindley, D.V. (1985) Reconciliation of discrete probability distributions. In *Bayesian Statistics 2* (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith), pp. 375-390. Amsterdam: North-Holland.

Madansky, A. (1978) Externally Bayesian groups. Unpublished manuscript, University of Chicago.

McCulloch, R.E. and Rossi, P.E. (1991) Bayes factors for nonlinear hypotheses and likelihood distributions, *Biometrika*, **79**, 663-676.

O'Hagan, A. (1991) Bayes-Hermite quadrature. *J. Stat. Plan. and Inference*, **29**, 245-260.

Poole, D. and Raftery, A.E. (1998) Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using Bayesian and full pooling Bayesian synthesis methods. Paper SC/50/AS6 presented to the IWC Scientific Committee, May 1998.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986) *Numerical Recipes: The Art of Scientific Computing.* Cambridge: Cambridge University Press.

Punt, A.E. (1998) A full description of the standard Baleen II model and some variants thereof. Paper SC/50/AWMP1 presented to the IWC Scientific Committee, May 1998.

Punt, A.E. and Butterworth, D.S. (1996) Further remarks on the Bayesian approach for assessing the Bering-Chukchi-Beaufort Seas stock of bowhead whales. *Rep. int. Whal. Commn*, **46**, 481-491.

Punt, A.E. and Butterworth, D.S. (1998) An assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) for 1998 using a Bayesian approach. Paper SC/50/AS1 presented to the IWC Scientific Committee, May 1998.

Raftery, A.E. and Banfield, J.D. (1991) Stopping the Gibbs sampler, the use of morphology and other issues in spatial statistics. *Ann. Inst. Statist. Math.*, **43**, 32-43.

Raftery, A.E., Givens, G.H. and Zeh, J.E. (1992) Inference from a deterministic population dynamics model about bowhead whale, *Balaena mysticetus*, replacement yield. Technical Report no. 232, Department of Statistics, University of Washington, Seattle.

Raftery, A.E., Givens, G.H. and Zeh, J.E. (1995a) Inference from a deterministic population dynamics model for bowhead whales (with discussion). *J. Amer. Statist. Ass.*, **90**, 402-430.

Raftery, A.E., Givens, G.H. and Zeh, J.E. (1995b) Rejoinder. *J. Amer. Statist. Ass.*, **90**, 427-430.

Raftery, A.E. and Poole, D. (1997) Bayesian synthesis assessment methodology for bowhead whales. Paper SC/49/AS5 presented to the IWC Scientific Committee. September 1997.

Raftery, A.E., Poole, D. and Givens, G.H. (1996) The Bayesian synthesis assessment method: resolving the Borel paradox and comparing the backwards and forwards variants. Paper SC/48/AS16 presented to the IWC Scientific Committee. June 1996.

Raftery, A.E. and Zeh, J.E. (1993) Estimation of bowhead whale, *Balaena mysticetus*, population size (with discussion). In *Bayesian Statistics in Science and Technology: Case Studies* (eds. J.S. Hodges, R.E. Kass, C. Gatsonis, and N. Singpurwalla), pp. 163-240. New York: Springer-Verlag.

Raftery, A.E. and Zeh, J.E. (1998) Estimating bowhead whale population size and rate of increase from the 1993 census. *J. Amer. Statist. Ass.*, **93**, 451-463.

Reilly, S.B. (1984) Observed and maximum rates of increase in gray whales, *Eschrichtius robustus. Rep. int. Whal. Commn*, Special Issue **6**, 389-399.

Reilly, S.B. (1992) Population biology and status of eastern Pacific gray whales: recent developments. *Wildlife 2001: Populations* (eds. D.R. McCullogh and R.H. Barrett), pp. 1062-1074. London:Elsevier.

Rubin, D.B. (1987) Comment. *J. Amer. Statist. Ass.*, **82**, 543-546.

Rubin, D.B. (1988) Using the SIR Algorithm to simulate posterior distributions. In *Bayesian Statistics 3* (eds. J.M. Bernardo et al.), pp. 395-402. Oxford: Clarendon Press.

Schweder, T. (1995) Comment. *J. Amer. Statist. Ass.*, **90**, 420-423.

Schweder, T. and Hjort, N.L. (1995) Bayesian synthesis or likelihood synthesis - what does the Borel paradox say? Paper SC/47/AS3 presented to the IWC Scientific Committee meeting, Dublin, May 1995.

Schweder, T. and Hjort, N.L. (1996) Bayesian synthesis or likelihood synthesis - what does Borel's paradox say? *Rep. int. Whal. Commn*, **46**, 475-480.

Schweder, T. and Hjort, N.L. (1997) Indirect and direct likelihoods and their synthesis - with an appendix on Minke whale dynamics. Paper SC/49/AS9 presented to the IWC Scientific Committee, September 1997.

Speed, T. (1983) Modelling. *CSIRO DMS Newsletter*, **96**, 1-2.

Terrell, G.R. (1990) The maximal smoothing principle in density estimation. *J. Amer. Statist. Ass.*, **85**, 470-477.

Wade, P.R. and DeMaster, D.P. (1996) A Bayesian analysis of eastern Pacific gray whale population dynamics. Paper SC/48/AS3 presented to the IWC Scientific Committee. June 1996.

Wolpert, R.L. (1995) Comment. *J. Amer. Statist. Ass.*, **90**, 426-427.

Appendix

Proof of Theorem 1: The joint distribution of (θ', ϕ') is

$$p'(\theta', \phi') = p[g_1^{-1}(\theta'), g_2^{-1}(\phi')] \left|\frac{\mathrm{d}g_1^{-1}(\theta')}{\mathrm{d}\theta'}\right| \left|\frac{\mathrm{d}g_2^{-1}(\phi')}{\mathrm{d}\phi'}\right|$$

and rewriting the model $\phi = M(\theta)$ in terms of θ' and ϕ' with ϕ' as the output, yields

$$\phi' = g_2[M(g_1^{-1}(\theta'))] \tag{38}$$

Now consider the transformation $(\theta', \phi') \rightarrow (U, V)$ where

$$\begin{aligned} U &= \phi' - g_2[M(g_1^{-1}(\theta'))] \\ V &= \theta' \end{aligned}$$

so that

$$\begin{aligned} \theta' &= V \\ \phi' &= U + g_2[M(g_1^{-1}(V))] \end{aligned}$$

with associated Jacobian

$$\left|\begin{array}{cc} 0 & 1 \\ 1 & \frac{\partial g_2[M(g_1^{-1}(v))]}{\partial v} \end{array}\right| = 1$$

It follows that the joint distribution of U and V is

$$\begin{aligned} p_{U,V}(u, v) &= p'(v, g_2(M(g_1^{-1}(v)))).1 \\ &= p[g_1^{-1}(v), g_2^{-1}(u + g_2(M(g_1^{-1}(v))))] \left|\frac{\mathrm{d}g_1^{-1}(v)}{\mathrm{d}v}\right| \left|\frac{\mathrm{d}g_2^{-1}(\phi')}{\mathrm{d}\phi'}\right|_{\phi'=u+g_2(M(g_1^{-1}(v)))} \end{aligned}$$

Now noting that $U = 0 \Leftrightarrow \phi' = g_2(M(g_1^{-1}(v))) \Leftrightarrow \phi = M(\theta)$, we have

$$\begin{aligned} p_{V|U}(v \mid 0) &\propto p_{U,V}(0, v) \\ &= p[g_1^{-1}(v), M(g_1^{-1}(v))] \left|\frac{\mathrm{d}g_1^{-1}(v)}{\mathrm{d}v}\right| \left|\frac{\mathrm{d}g_2^{-1}(\phi')}{\mathrm{d}\phi'}\right|_{\phi'=g_2(M(g_1^{-1}(v)))} \end{aligned}$$

Since $V = \theta'$ and conditioning on $U = 0$ is equivalent to conditioning on the model (38), this is exactly the Bayesian synthesis post-model distribution of θ'. Writing in terms of θ', we have

$$\pi^{[\theta']}(\theta') \propto p[g_1^{-1}(\theta'), M(g_1^{-1}(\theta'))] \left|\frac{\mathrm{d}g_1^{-1}(\theta')}{\mathrm{d}\theta'}\right| \left|\frac{\mathrm{d}g_2^{-1}(\phi')}{\mathrm{d}\phi'}\right|_{\phi'=g_2(M(g_1^{-1}(\theta')))}$$

To express the post-model distribution in terms of θ, we transform $\theta' \rightarrow \theta$ which has inverse Jacobian $\left|\frac{\mathrm{d}g_1(\theta)}{\mathrm{d}\theta}\right|$. Noting the identities $\left|\frac{\mathrm{d}g_2^{-1}(\phi')}{\mathrm{d}\phi'}\right| = \left|\frac{\mathrm{d}g_2(\phi)}{\mathrm{d}\phi}\right|^{-1}$ and $\left|\frac{\mathrm{d}g_1^{-1}(\theta')}{\mathrm{d}\theta'}\right| = \left|\frac{\mathrm{d}g_1(\theta)}{\mathrm{d}\theta}\right|^{-1}$, we obtain

$$\pi^{[\theta]}(\theta) \propto p(\theta, M(\theta)) \left| \frac{\mathrm{d}g_2(\phi)}{\mathrm{d}\phi} \right|^{-1}_{\phi=M(\theta)}$$

as in (7).

$\square$

Proof of Theorem 2: We first consider the continuous case. For $\alpha = 0$ and $\alpha = 1$ the result is trivially true from the assumptions. For $\alpha \in (0,1)$, the function $h(x) = x^{\alpha}$ is concave on $[0,\infty)$ because $h''(x) = \alpha(\alpha-1)x^{\alpha-2} \leq 0$. Noting that the ratio $q_1^*(\phi)/q_2(\phi)$ has range $[0,\infty)$ for $\phi \in \Phi \setminus \{\phi : q_1^*(\phi) = q_2(\phi) = 0\}$, and applying Jensen's inequality, we have

$$\begin{aligned}
\int q_1^*(\phi)^{\alpha} q_2(\phi)^{1-\alpha} d\phi &= E_{q_2}\left[\left(\frac{q_1^*(\phi)}{q_2(\phi)}\right)^{\alpha}\right] \\
&\leq \left[E_{q_2}\left(\frac{q_1^*(\phi)}{q_2(\phi)}\right)\right]^{\alpha} \\
&= 1^{\alpha} = 1
\end{aligned}$$

since

$$E_{q_2}\left(\frac{q_1^*(\phi)}{q_2(\phi)}\right) = \int q_1^*(\phi) d\phi = 1.$$

So $\int q_1^*(\phi)^{\alpha} q_2(\phi)^{1-\alpha} d\phi \leq 1 < \infty$, and the normalizing constant k_{α} can be found. Note that the proof proceeds analogously in the discrete case except that the integrals are replaced by the appropriate sums.

$\square$

Proof of Theorem 3: We first note that $\bigcup A_i = \bigcup C_j$ and that the C_j are mutually exclusive because each A_i appears in only one C_j, namely $C_{m(i)}$. Since $0 \leq \tilde{q}^{[\phi]}(B_{m(i)}) \leq 1$ and $q_1(A_i) \leq q_1^*(B_{m(i)})$, it follows that $0 \leq \tilde{q}^{[\theta]}(A_i) \leq 1$ for each i. By the mutual exclusivity of the C_j,

$$\begin{aligned}
\sum_i \tilde{q}^{[\theta]}(A_i) &= \sum_j \sum_{i:A_i \in C_j} \tilde{q}^{[\theta]}(A_i) \\
&= \sum_j \sum_{i:A_i \in C_j} \tilde{q}^{[\phi]}(B_j) \left(\frac{q_1(A_i)}{Q_1(C_j)}\right) \\
&= \sum_j \sum_{i:A_i \in C_j} \tilde{q}^{[\phi]}(B_j) \left(\frac{q_1(A_i)}{\sum_{k:A_k \in C_j} q_1(A_k)}\right) \\
&= \sum_j \tilde{q}^{[\phi]}(B_j) \\
&= 1
\end{aligned}$$

using Theorem 2.

□

Proof of Theorem 4: We write $\theta = (\theta_1, \ldots, \theta_n)$ and $\phi = (\phi_1, \ldots, \phi_n)$. Denote the components of M as follows:

$$\begin{aligned}
\phi_1 &= M_1(\theta_1, \ldots, \theta_n) \\
\phi_2 &= M_2(\theta_1, \ldots, \theta_n) \\
&\vdots \\
\phi_n &= M_n(\theta_1, \ldots, \theta_n)
\end{aligned}$$

Since M is one-to-one from A_i onto B for each $i = 1, \ldots, k$, the inverse function from B to A_i can be found. Denote the ith inverse by

$$\begin{aligned}
\theta_1 &= G_{1i}(\phi_1, \ldots, \phi_n) \\
\theta_2 &= G_{2i}(\phi_1, \ldots, \phi_n) \\
&\vdots \\
\theta_n &= G_{ni}(\phi_1, \ldots, \phi_n)
\end{aligned}$$

Let J_i denote the Jacobian computed from the ith inverse, ie.

$$J_i = \begin{vmatrix} \frac{\partial\theta_1}{\partial\phi_1} & \frac{\partial\theta_1}{\partial\phi_2} & \cdots & \frac{\partial\theta_1}{\partial\phi_n} \\ \frac{\partial\theta_2}{\partial\phi_1} & \frac{\partial\theta_2}{\partial\phi_2} & \cdots & \frac{\partial\theta_2}{\partial\phi_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\theta_n}{\partial\phi_1} & \frac{\partial\theta_n}{\partial\phi_2} & \cdots & \frac{\partial\theta_n}{\partial\phi_n} \end{vmatrix}$$

From Theorem 2, $\tilde{q}^{[\phi]}(\phi)$ is a proper density function for all $\alpha \in [0,1]$. We now show that $\tilde{q}^{[\theta]}(\theta)$ in (18) is simply the distribution of a transformation of ϕ when $\phi \sim \tilde{q}^{[\phi]}(\phi)$. Fix a value of i, and consider $\theta \sim q_1(\theta)$ for $\theta \in A_i$. Since M is one-to-one from A_i to B, the induced density of $\phi = M(\theta)$ for $\theta \in A_i$ is

$$q_1^*(\phi) \propto q_1(G_{1i}(\phi), \ldots, G_{ni}(\phi)|J_i|,$$

and hence

$$\tilde{q}^{[\phi]}(\phi) \propto \{q_1(G_{1i}(\phi), \ldots, G_{ni}(\phi)|J_i|\}^\alpha q_2(\phi)^{1-\alpha}$$

The Jacobian of the inverse transformation $\theta_1 = G_{1i}(\phi), \ldots, \theta_n = G_{ni}(\phi)$ is $|J_i|^{-1}_{\phi=M(\theta)}$, so transforming the above to $\theta \in A_i$ yields

$$\begin{aligned}
\tilde{q}^{[\theta]}(\theta) &\propto \{q_1(\theta)|J_i|_{\phi=M(\theta)}\}^\alpha q_2(M(\theta))^{1-\alpha}|J_i|^{-1}_{\phi=M(\theta)} \\
&= q_1(\theta)\left(\frac{q_2(M(\theta))}{q_1(\theta)|J_i|_{\phi=M(\theta)}}\right)^{1-\alpha} \\
&= q_1(\theta)\left(\frac{q_2(M(\theta))}{q_1^*(M(\theta))}\right)^{1-\alpha}, \quad \theta \in A_i, \qquad (39)
\end{aligned}$$

since $q_1^*(M(\theta)) = q_1(\theta)|J_i|_{\phi=M(\theta)}$ for $\theta \in A_i$. Note that (39) is identical to (18), which is equivalent to (17) apart from the proportionality constant k_α. The piecewise transformation above is the same for every $i = 1, \ldots, k$, so $\tilde{q}^{[\theta]}(\theta)$ has the form (18) for every $\theta \in \Theta$. It remains to verify that if $\theta \sim \tilde{q}^{[\theta]}(\theta)$, then $\phi = M(\theta)$ has density $\tilde{q}^{[\phi]}(\phi)$. Let $\theta \sim \tilde{q}^{[\theta]}(\theta)$. From Casella and Berger (1990, p. 177), the induced distribution of ϕ is

$$\begin{aligned}
\tilde{q}^{[\phi]}(\phi) &\propto \sum_{i=1}^{k} \tilde{q}^{[\theta]}(G_{1i}(\phi), \ldots, G_{ni}(\phi))|J_i| \\
&= \sum_{i=1}^{k} \left(q_1(G_{1i}(\phi), \ldots, G_{ni}(\phi)) \left(\frac{q_2(\phi)}{q_1^*(\phi)}\right)^{1-\alpha} \right) |J_i| \\
&= \left(\frac{q_2(\phi)}{q_1^*(\phi)}\right)^{1-\alpha} \sum_{i=1}^{k} q_1(G_{1i}(\phi), \ldots, G_{ni}(\phi))|J_i| \qquad (40)
\end{aligned}$$

since the ratio does not depend on the value of i. The summation in (40) is simply $q_1^*(\phi)$, the distribution of ϕ when $\theta \sim q_1(\theta)$. Thus,

$$\begin{aligned}
\tilde{q}^{[\phi]}(\phi) &\propto \left(\frac{q_2(\phi)}{q_1^*(\phi)}\right)^{1-\alpha} q_1^*(\phi) \\
&= q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha},
\end{aligned}$$

identical to the form of $\tilde{q}^{[\phi]}(\phi)$ in (9). Therefore, $\tilde{q}^{[\theta]}(\theta)$ is a density on Θ and it induces $\tilde{q}^{[\phi]}(\phi)$ under the model transform $\phi = M(\theta)$.

□

Proof of Theorem 5: Let $\tau = (\phi, \gamma)$ and denote $\tau = G(\theta)$ so that $\theta = G^{-1}(\tau)$. By Theorem 2, there exists k_α such that

$$\int k_\alpha q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha} d\phi = 1 \quad \forall \alpha \in [0, 1]$$

Now define the pooled prior of τ to be

$$\begin{aligned}
\tilde{q}^{[\tau]}(\tau) &= \tilde{q}^{[\tau]}(\phi, \gamma) \\
&= k_\alpha q_1^*(\phi)^{\alpha-1} q_2(\phi)^{1-\alpha} q^*(\phi, \gamma)
\end{aligned}$$

where $q^*(\phi, \gamma)$ is the joint density of (ϕ, γ) induced by G and $q_1^*(\phi) = \int q^*(\phi, \gamma) d\gamma$. Note that $\tilde{q}^{[\tau]}(\tau)$ is a well-defined joint density since

$$\begin{aligned}
\int \tilde{q}^{[\tau]}(\tau) d\tau &= \int \int k_\alpha q_1^*(\phi)^{\alpha-1} q_2(\phi)^{1-\alpha} q^*(\phi, \gamma) d\gamma d\phi \\
&= \int k_\alpha q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha} d\phi \\
&= 1
\end{aligned}$$

indicating that $\tilde{q}^{[\tau]}(\tau)$ integrates to unity on its support and marginalizes (over γ) to yield $\tilde{q}^{[\phi]}(\phi)$. Now noting that

$$q^*(\phi,\gamma) = q_1(G^{-1}(\phi,\gamma))|J|$$

where $J = \left|\frac{d\theta}{d\tau}\right|$, we have

$$\tilde{q}^{[\tau]}(\tau) = k_\alpha q_1^*(\phi)^{\alpha-1} q_2(\phi)^{1-\alpha} q_1(G^{-1}(\phi,\gamma))|J|$$

Applying the transformation $\theta = G^{-1}(\tau)$ with Jacobian J^{-1} then yields

$$\begin{aligned} \tilde{q}^{[\theta]}(\theta) &= k_\alpha q_1^*(M(\theta))^{\alpha-1} q_2(M(\theta))^{1-\alpha} q_1(\theta) \\ &= k_\alpha q_1(\theta) \left(\frac{q_2(M(\theta))}{q_1^*(M(\theta))}\right)^{1-\alpha} \end{aligned}$$

which is equivalent to (17). So $\tilde{q}^{[\theta]}(\theta)$ is simply a transformation of $\tilde{q}^{[\tau]}(\tau)$ which, by nature of its construction, integrates to unity over its support and marginalizes to yield $\tilde{q}^{[\phi]}(\phi)$. Under the conditions of the proposition, $\tilde{q}^{[\theta]}(\theta)$ is a density on Θ and it induces $\tilde{q}^{[\phi]}(\phi)$ under the model transform $\phi = M(\theta)$.

□