Inference for deterministic simulation models: The Bayesian melding approach
David Poole; Adrian E Raftery
Journal of the American Statistical Association; Dec 2000; 95, 452; ABI/INFORM Global
pg. 1244

Inference for Deterministic Simulation Models: The Bayesian Melding Approach

David POOLE and Adrian E. RAFTERY

Deterministic simulation models are used in many areas of science, engineering, and policy making. Typically, these 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. A probabilistic approach, called Bayesian synthesis, was shown to suffer from the Borel paradox, according to which the results can depend on the parameterization of the model. We propose a modified approach, called *Bayesian melding,* which takes into full account 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 sampling-importance-resampling (SIR) algorithm. We develop this initially for invertible models, and then extend it to the more difficult and more common case of noninvertible models. We illustrate the methodology using a number of examples. Simulation studies show that the method outperforms a simpler Bayesian approach in terms of mean squared error. A number of open research problems are discussed.

KEY WORDS: Bayesian inference; Borel paradox; Logarithmic pooling; Population dynamics model; Sampling-importance-resampling algorithm; Whales.

1. INTRODUCTION

Deterministic 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.

Simulation models are usually designed to capture some underlying mechanism or natural process. They differ conceptually from many standard statistical models (such as linear regression) whose aim is to empirically estimate the relationships between variables. The deterministic model is viewed as a useful approximation of reality that is easier to build and interpret than a stochastic model. However, such models can be extremely complicated with large numbers of inputs and outputs, and they 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 for standard statistical models, yet this is an area that has received little attention from statisticians.

Edwards (1967) observed that there was a general need for simulation-based methods of estimation in complex models. Speed (1983) was perhaps the first to point out the need for statisticians to get involved in the deterministic modeling arena, noting that scientists at the Australian national research organization (CSIRO) were increasingly abandoning statistical models in favor of simulation models. He gave a model for sheep growth as an example.

Although little attention has been paid to formally quantifying parameter uncertainty, simulation model validation has received considerable attention. Caswell (1976), Guttorp and Walden (1987), and Sampson and Guttorp (1999) discussed applications in ecology, geophysics, and atmospheric science. Hodges (1987, 1991) discussed model uncertainty in the context of simulation models.

Deterministic simulations have received attention in the statistical literature under the general topic of computer experiments (Currin, Mitchell, Morris, and Ylvisaker 1991; Morris, Mitchell, and Ylvisaker 1993; Sacks, Welch, Mitchell, and Wynn 1989). Computer experiments are similar to the models discussed earlier in that they are typically mathematical simulations of complex systems, and they also require a number of inputs. Often the experimenter wishes to fit a cheap predictor of the output (using some stochastic process) because the model itself is computationally expensive to run. Another common objective is to choose values of inputs that optimize some function of the output. Sacks et al. (1989) discussed experimental design issues and the choice of the stochastic process, and they provided some examples. The objectives of computer experiments are somewhat different than those on which we focus here. In our ap-

David Poole is Senior Technical Staff Member, AT&T Labs—Research, Florham Park, NJ 07932 (E-mail: *poole@research.att.com*). Adrian E. Raftery is Professor of Statistics and Sociology, University of Washington, Seattle, WA 98195 (E-mail: *raftery@stat.washington.edu*). This research was supported in part by the North Slope Borough, Alaska, and the U.S. National Oceanic and Atmospheric Administration (through the U.S. National Marine Mammal Laboratory and the Alaska Eskimo Whaling Commission), and in part by the United States Environmental Protection Agency through agreement CR825173-01-0 to the University of Washington. However, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. The authors are grateful to Thomas F. Albert for sustained research support, and to Christina Fraley, Geof H. Givens, and Judith E. Zeh for very helpful comments and discussions. The authors would also like to thank the editor, the associate editor, and three referees for very useful comments that improved the article, and in particular a reviewer for the comment regarding the Kullback-Leibler distance in Section 3.

© 2000 American Statistical Association
Journal of the American Statistical Association
December 2000, Vol. 95, No. 452, Theory and Methods

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

plications, the chief objective is to quantify and reduce the uncertainty about key quantities, particularly the model inputs and outputs and functions thereof, using all the sources of evidence available.

A common way of specifying the inputs to a simulation model is an ad hoc trial-and-error, or "tuning," approach. This begins with an initial educated guess at appropriate values of the inputs. If the corresponding outputs are plausible, then this initial guess is used; otherwise, new values are tried until a satisfactory set of outputs is obtained. Thereafter, a sensitivity analysis is often run to see whether the final conclusions are sensitive to the precise values of the inputs used. If the conclusions turn out to be insensitive to the inputs, then all is well. However, if there is some sensitivity, then it is not clear what should be done. One option is to note it in the report and to attach a warning to the conclusions. Another is to assume values that are conservative or that cater to the "worst-case" scenario. 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. 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 earlier, with the conservative assessment of natural increase rate taken as the lowest value from a rather limited sensitivity analysis. Because the key quantities of interest here depend crucially on model inputs, it is important to reduce the uncertainty about the model inputs as much as possible. In other simulation exercises (e.g. combat modeling), the simulation is intended primarily as a means of obtaining inference about the outputs.

In 1991, the IWC decided that sensitivity analysis 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 (Givens 1993; Raftery, Givens, and Zeh 1992, 1995—hereafter, RGZ), and was used by the IWC to set the bowhead quota in 1994. However, Wolpert (1995) showed that Bayesian synthesis was subject to the Borel paradox, so that the results could depend on the model's parameterization. A 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. The word "melding" is used because the method provides a way of combining different kinds of information (qualitative or quantitative, fragmentary or extensive, based on expert knowledge or on data) about different quantities, as long as the quantities to which they relate can be linked using a deterministic model.

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 build on those to develop the Bayesian melding approach, using a pair of simple examples. In Section 4 we illustrate the method using a population dynamics model for bowhead whales, and in Section 5 we compare the method to a simpler standard Bayesian approach by means of simulation. We prove selected theorems 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

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

to denote the situation in which the deterministic model M relates a vector of input parameters θ to a vector of outputs ϕ, so that $\phi = M(\theta)$. We let ψ denote 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 noninvertible, and in many applications the noninvertibility is due to the dimension of ϕ being less than that of θ; that is, $p < n$. In these cases, a single value of the output vector may result from many different values of the input vector.

Our motivating application is the study of population dynamics models (PDMs); Example 2 is of this type. A PDM typically relates the population at time $(t+1)$ to the population at time t. More complex age-structured PDMs relate the population aged a at time $(t+1)$ to the population at each age at time t. For the purposes of this article, we consider the following two deterministic functions.

2.1.1 Example 1: $\boldsymbol{Z = Y/X}$. This is an extremely simple illustrative example. X and Y are the two inputs to the model, whereas Z is the single output. Hence $\theta = (X, Y)$ and $\phi = Z$. This is an example of a case where a given output can be generated by infinitely many values of the inputs.

2.1.2 Example 2. A model that underlies the commercial revised management procedure of the IWC is a non–age-structured PDM of the form

$$P_{t+1} = P_t - C_t + 1.5(\text{MSYR})P_t(1 - (P_t/P_0)^2). \quad (1)$$

In (1), 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 rate, and C_t is the number of whales killed by hunting in year t. MSYR is defined as the natural rate of increase when the population is at an equilibrium known as the maximum sus-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tainable yield level. At this equilibrium, the population can indefinitely sustain an annual catch known as the maximum sustainable yield.

This model is somewhat simpler than the PDM actually used by the IWC for bowhead assessment, which is called BALEEN II, 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 population increase.

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 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 the size of the population goes above this level, then the population is assumed to decrease until it reaches the carrying capacity again.

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.

2.2 Bayesian Synthesis and the Borel Paradox

RGZ (1995) described an approach to the bowhead whale problem that became known as the Bayesian synthesis method. As we did in Section 2.1, RGZ let θ denote the set of model inputs and let ϕ denote 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 *premodel 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, denoted by $\theta \to M(\theta)$. RGZ defined the joint distribution of θ and ϕ given the model to be simply the restriction of the premodel distribution to the submanifold $\{(\theta, \phi) : \phi = M(\theta)\}$, namely

$$\pi(\theta, \phi) \propto \begin{cases} p(\theta, M(\theta)) & \text{if } \phi = M(\theta) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

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

$$\pi^{[\theta]}(\theta) \propto p(\theta, M(\theta)) \tag{3}$$

or, equivalently,

$$\pi^{[\theta]}(\theta) = p^{[\theta|\phi]}\ (\theta|\phi = M(\theta)). \tag{4}$$

Wolpert (1995), in a discussion of RGZ, pointed out that a conditional distribution of the form (4) is ill-defined, and as a result the Bayesian synthesis postmodel distribution is subject to a phenomenon known as the Borel paradox. One consequence of the Borel paradox is that the postmodel distribution depends on how the simulation model M is parameterized. Schweder and Hjort (1996) pointed out that the consequences of the paradox are 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 alluded to the phenomenon in his discussion of conditional probability. Because the Bayesian synthesis approach relies on a conditional distribution that is ill-defined, it is not satisfactory. 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 a Bayesian context, it is useful to decompose the premodel information into prior and likelihood components. Under the assumption that premodel information about inputs is independent of that about outputs, we can decompose the Bayesian synthesis premodel distribution as

$$\begin{aligned} p(\theta, \phi) &= p^{[\theta]}(\theta) p^{[\phi]}(\phi) \\ &\propto q_1(\theta) q_2(\phi) L_1(\theta) L_2(\phi), \end{aligned} \tag{5}$$

where $q_1(\cdot)$ and $q_2(\cdot)$ are the prior distributions of the inputs and outputs, $L_1(\theta) = p(D_\theta|\theta)$ 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. In considering the decomposition (5), it becomes apparent that the Borel paradox does not arise from the likelihood components. Likelihoods are invariant to reparameterization, as pointed out by, for example, Schweder and Hjort (1996). Thus the problem lies with the prior distributions $q_1(\theta)$ and $q_2(\phi)$.

Because θ is a random variable with density $q_1(\theta), \phi = M(\theta)$ is also a random variable, because 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 we can write

$$q_1^*(\phi) = q_1(M^{-1}(\phi))|J(\phi)|, \tag{6}$$

where $J(\phi) = |d\theta/d\phi|$ is the associated Jacobian of the transformation. For a complicated M, even when it is invertible, the functional form of $J(\phi)$ may be very difficult to calculate. When M is noninvertible and $\dim(\Phi)$ is less

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

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 $q_1^*(\phi)$. Allowing for this induced distribution and the pre-existing $q_2(\phi)$, there thus are two prior distributions on the same quantity, ϕ. Because 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 $\tilde{q}^{[\theta]}(\theta)$, then we would be in a position to define

$$\pi^{[\theta]}(\theta) \propto \tilde{q}^{[\theta]}(\theta) L_1(\theta) L_2(M(\theta)), \tag{7}$$

which is a standard Bayesian posterior distribution, and standard Bayesian inference could then follow. The Borel paradox would no longer arise. The concept of a *joint* pre-model distribution, as in (5), is no longer considered; we require only marginal prior distributions for θ and ϕ; prior, that is, to the likelihood components $L_1(\cdot)$ and $L_2(\cdot)$.

As with the Bayesian synthesis method, Monte Carlo methods would be used 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 in (7).

3.2 Combining Probability Distributions

The presence of two priors on one quantity occurs quite naturally in a simulation modeling framework. One possibility is to combine the two prior distributions into a single one as in Section 3.1. We assume here that the two distributions are not completely incompatible, and thus that there exists a region on which both have support. (If the two priors place mass on completely conflicting regions of the parameter space, this would indicate a problem with the model or with at least one of the priors, and one would need to place both the model and the sources of evidence under careful review before proceeding.)

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), Genest and Zidek (1986), and Givens and Roback (1999) reviewed work in the area. Two pooling methodologies that have received extensive study are

linear pooling: $T(q_1, \ldots q_k) = \sum_{i=1}^k \alpha_i q_i$,
and
logarithmic pooling: $T(q_1, \ldots q_k) \propto \prod_{i=1}^k q_i^{\alpha}$,

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 must consider the order in which pooling and updating of a prior (given a likelihood from data) are performed. There are two possibilities. First, each prior distribution is updated using Bayes's theorem, and then the combined prior distribution is formed. Alternatively, 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, McConway, and Schervish (1986) showed that under mild conditions, 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. In terms of the output ϕ, a logarithmically pooled prior distribution has the form

$$\tilde{q}^{[\phi]}(\phi) \propto q_1^*(\phi)^{\alpha} q_2(\phi)^{1-\alpha}, \tag{8}$$

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 1. If ϕ is expressed on a continuous scale 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, $\alpha_1 = \alpha_2 = \alpha = .5$ assigns equal weight to each individual prior. We refer to (8) with $\alpha = .5$ as *geometric* pooling because (8) 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 originally envisaged situation in which two experts provide opinions on the same quantity. This provides an informal justification for the choice $\alpha = .5$ in the bowhead case. Because the two sources of prior information (although independent) are selected by the same expert, they can be viewed as equally reliable and hence should be assigned equal weight. Recall that it is only the prior distributions that are pooled in this way; likelihoods enter the posterior with full weight as in (7). We discuss the choice of α in more detail in Section 6.

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, then inference about the model inputs, θ, can be obtained by simple inversion of the combined prior distribution of model outputs in (8). We denote the resulting pooled prior on inputs by $\tilde{q}^{[\theta]}(\theta)$. When the model is not one-to-one, the geometrically pooled prior distribution of the model outputs is still unambiguously defined, but that of the model inputs is not. Here we propose a solution for the general noninvertible case.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.3.1 A Simple Discrete Noninvertible Example. Consider the following very simple deterministic model. It has one input and one output, each of which is discrete. The input, θ, has three possible values (1, 2, 3), whereas 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 (8), 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 so as 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 use other grounds to decide 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 $\theta = 3$, because 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 (8). 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; that is,

$$\tilde{Q}^{[\theta]}(C_j) = \tilde{q}^{[\phi]}(B_j), \tag{9}$$

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 use capital Q to denote a probability measure. The second rule is that if C_j has more than one member, then its probability is split between its members in proportion to their prior densities, $q_1(A_i)$.

This yields the pooled prior density on inputs

$$\tilde{q}^{[\theta]}(A_i) = \tilde{q}^{[\phi]}(B_{m(i)})\left(\frac{q_1(A_i)}{Q_1(C_{m(i)})}\right). \tag{10}$$

We also have that

$$Q_1(C_j) = q_1^*(B_j). \tag{11}$$

Combining (10) with (11) gives the alternative form

$$\tilde{q}^{[\theta]}(A_i) = \tilde{q}^{[\phi]}(B_{m(i)})\left(\frac{q_1(A_i)}{q_1^*(B_{m(i)})}\right). \tag{12}$$

It is easily verified that (12) yields the same solution as before for the simple discrete model in Section 3.3.1. In practice, (12) is more convenient than (10) 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 2. The pooled prior given by (12) is a discrete probability function with respect to the dominating counting measure; that is, $0 \le \tilde{q}^{[\theta]}(A_i) \le 1$ for all i, and $\sum_i \tilde{q}^{[\theta]}(A_i) = 1$.

3.3.4 The Continuous Case. We now extend the general solution for the discrete case, (12), 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.

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

$$\tilde{Q}^{[\theta]}(C) = \tilde{Q}^{[\phi]}(B). \tag{13}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

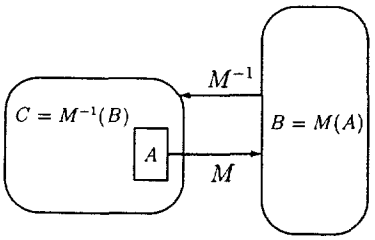

Figure 1. The Sets A, B, and C in the Heuristic Derivation of the Pooled Prior Density of the Inputs.

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

$$Q_1(C) = Q_1^*(B). \tag{14}$$

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

$$\begin{aligned} \tilde{Q}^{[\theta]}(A) &= \tilde{Q}^{[\theta]}(C)\left(\frac{Q_1(A)}{Q_1(C)}\right) \\ &= \tilde{Q}^{[\phi]}(B)\left(\frac{Q_1(A)}{Q_1^*(B)}\right), \end{aligned} \tag{15}$$

by (13) and (14). Then we propose that the probability density function corresponding to (15), if it exists, be given by

$$\begin{aligned} \tilde{q}^{[\theta]}(\theta) &= \tilde{q}^{[\phi]}(M(\theta))\left(\frac{q_1(\theta)}{q_1^*(M(\theta))}\right) \\ &= k_\alpha q_1(\theta)\left(\frac{q_2(M(\theta))}{q_1^*(M(\theta))}\right)^{1-\alpha}. \end{aligned} \tag{16}$$

This corresponds to letting the length h of the side of the hypercube A tend to 0. Looking at (16), we observe that $\tilde{q}^{[\theta]}(\cdot)$ is simply the original prior, $q_1(\cdot)$, weighted by the ratio of two (possibly lower-dimensional) densities in ϕ space, $q_2(\cdot)$ and $q_1^*(\cdot)$, evaluated at $M(\theta)$. The magnitude of the weight is dictated by the value of α. Note also that the ratio is never evaluated at arbitrary values of ϕ, only at $M(\theta)$ for a given value of $\theta \in \Theta$.

We now identify two different situations in which (16) yields a well-defined probability density function that is also an inversion of $\tilde{q}^{[\phi]}(\phi)$. We first consider the case in which 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 noninvertible, 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 $M : A_i \to \Phi$ will be invertible for each $i = 1, 2$. The general result is as follows.

Theorem 3. Suppose that $\theta = (\theta_1, \ldots, \theta_n)$ and that $\phi = (\phi_1, \ldots, \phi_n)$; that is, 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 (a) $P(\theta \in A_0) = 0$, and A_0 may be empty, and (b) 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 $0 \le \alpha \le 1$, the pooled prior on the model inputs θ given by (16) is a proper density function and an inverse of $\tilde{q}^{[\phi]}(\phi)$.

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

Theorem 4. 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 $0 \le \alpha \le 1$, the pooled prior on the model inputs θ given by (16) is a proper density function and an inverse of $\tilde{q}^{[\phi]}(\phi)$.

Recall from the simple discrete example that we chose to use $q_1(\cdot)$ as the basis for inverting $\tilde{q}^{[\phi]}(\cdot)$ for each given value of ϕ. We picked a member of a class of priors $\mathcal{P}$ that all induce $\tilde{q}^{[\phi]}(\cdot)$; specifically, we chose $p \in \mathcal{P}$ such that $p(\cdot|\phi) = q_1(\cdot|\phi)$. This choice minimizes the Kullback–Leibler distance to $q_1(\cdot)$. For any $p \in \mathcal{P}$, we have

$$D(q_1, p) = ED(q_1(\cdot|\phi), p(\cdot|\phi)) + D(q_1^*, p^*),$$

where the asterisks denote distributions induced through the model as before. Now the second term is constant for every $p \in \mathcal{P}$, and the first term is made to vanish by making our specific choice.

3.4 Implementing Logarithmic Pooling via the Sampling-Importance-Resampling Algorithm

Using the pooled prior in (16), the posterior distribution of $\theta, \pi^{[\theta]}(\theta)$, is given by (7). A sample from this posterior can be obtained via the following modified sampling-importance-resampling (SIR) algorithm (Rubin 1987, 1988):

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, then it often will 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

$$w_i = \left(\frac{q_2(M(\theta_i))}{q_1^*(M(\theta_i))}\right)^{1-\alpha} L_1(\theta_i) L_2(M(\theta_i)). \tag{17}$$

5. Sample l values from the discrete distribution with values θ_i and probabilities proportional to w_i.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

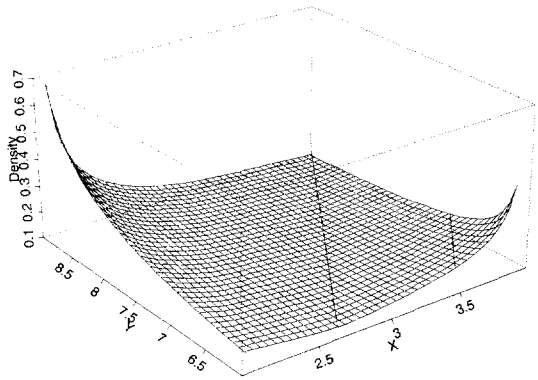

Figure 2. The Pooled Joint Prior Distribution of X and Y in the Simple Example. The shape reflects prior information on Z projected through the deterministic relationship. The two superimposed solid lines represent sets of points in the (X, Y) plane that map to two single points in Z space. The prior on Z says that each z value is equally likely, and the pooled prior accounts for this by increasing the density where such lines are shorter.

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

3.5 Example: A Simple Continuous Noninvertible Model

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

$$q_1(x,y) = \frac{1}{6}, \quad \text{for} \quad 2 < x < 4, \qquad 6 < y < 9$$

and

$$q_2(z) = \frac{1}{5}, \quad \text{for} \quad 0 < z < 5.$$

Recall that the model is noninvertible, because the output has lower dimension than the inputs. Indeed, each unique value of Z can be associated with an infinite number of ordered pairs (X, Y). This is a feature found in many of the more complicated models used in real situations. The simplicity of this example, however, allows an analytic solution to be found.

The change-of-variable method can be used to find $q_1^*(z)$, the distribution of Z induced by the model. Geometric pooling of $q_1^*(z)$ and $q_2(z)$ yields

$$\tilde{q}^{[\phi]}(z) = \begin{cases} \frac{k\sqrt{4z^2-9}}{\sqrt{15}z} & : 1.5 < z < 2.25 \\ \frac{k\sqrt{3}}{2z} & : 2.25 < z < 3 \\ \frac{k\sqrt{81-4z^2}}{2\sqrt{15}z} & : 3 < z < 4.5. \end{cases}$$

where $k \approx 1.4$ is the appropriate normalizing constant. To obtain the corresponding pooled prior on the inputs (X, Y), we apply (16) to obtain

$$\tilde{q}^{[\theta]}(x,y) = \frac{\tilde{q}^{[\phi]}(y/x)q_1(x,y)}{q_1^*(y/x)}$$
$$= \begin{cases} \frac{ky}{2\sqrt{15(4y^2-9x^2)}} & : y < 2.25x \\ \frac{k\sqrt{3}y}{45x} & : 2.25x < y < 3x \\ \frac{ky}{\sqrt{15(81x^2-4y^2)}} & : y > 3x \end{cases}$$

on $\Theta = \{(x,y) : 2 < x < 4, 6 < y < 9\}$, and 0 elsewhere. This density integrates to unity on its support and is shown graphically in Figure 2. The original flat prior on the inputs has been modified to reflect the additional marginal prior information on the output. Figure 3 shows the corresponding marginal pooled prior distributions of X, Y, and Z. The exact marginals and the distributions simulated using SIR are shown. The SIR method provides a good approximation to the true distributions.

4. EXAMPLE: A POPULATION DYNAMICS MODEL FOR BOWHEAD WHALES

We now apply the Bayesian melding 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 use prior distributions that have been used in previous illustrations using this model. These are not necessarily priors agreed to 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 used by Givens (1995), namely a shifted gamma distribution $6,400+$ gamma (2.8085, .0002886), where the mean is given by $2.8085/.0002886 = 9,731$. 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

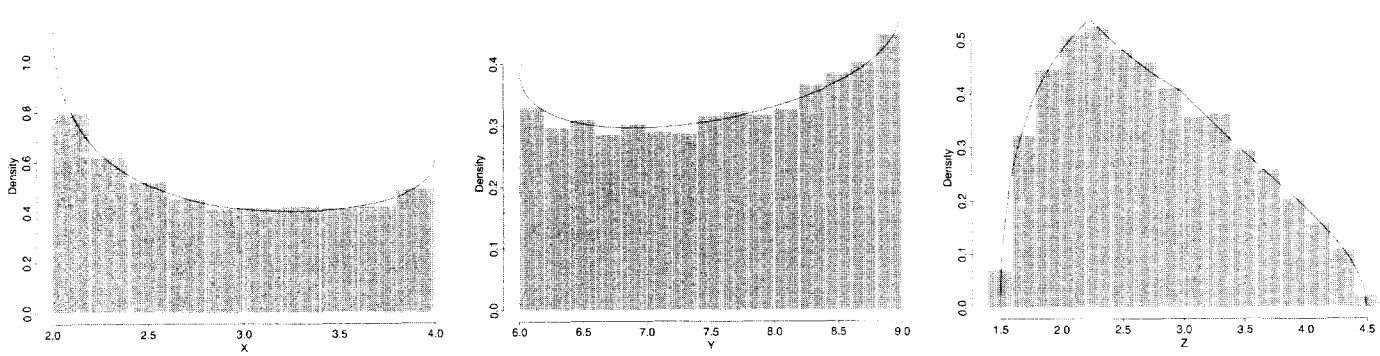

Figure 3. Exact and Simulated Pooled Prior Distributions of (a) X, (b) Y, and (c) Z for the Simple Example. The simulated distributions are generated using SIR, and approximate the exact distributions well.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

prior information about natural fertility and mortality. We based this on a distribution of MSYR for bowheads reported by Punt 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. Punt and Butterworth (1996) obtained the distribution from the BALEEN II model as a collection of values of MSYR, which (in combination with values of other parameters of the larger model) were not rejected by the model as being biologically unreasonable.

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 (q_2) is $N(7,800, 1,300^2)$, and the likelihood (L_2) is $N(8,293, 626^2)$. These choices correspond to a posterior distribution that is $N(8,200, 564^2)$, matching the mean and variance of the Bayes-empirical Bayes posterior distribution (which is nearly, but not exactly, normal). The prior for 1993 (q_2) is based on data from a previous visual and acoustic census carried out in 1988. Raftery and Zeh (1993) described the 1988 census in detail. Because this prior is based on a previous recent census, it is independent of the priors on the inputs P_0 and MSYR. Neither of these priors uses census information in any way.

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 $P_{1993} = (1+\text{ROI})^{15}P_{1978}$. 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.

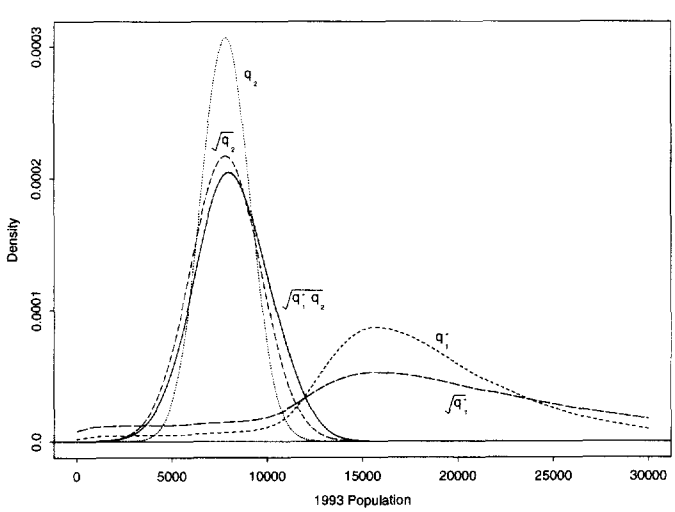

Figure 4. The Components of the Geometrically Pooled Prior Distribution of the Output $\phi = P_{1993}$ in the Simple PDM. These components are as follows: - - - q_1^*; $\cdots$ q_2; — — $\sqrt{q_1^*}$; - - $\sqrt{q_2}$; —— $\sqrt{q_1^* q_2}$. All of the densities shown have been renormalized to integrate to unity. 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 that 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 for P_{1993} induced by the model from the priors on the inputs, q_1^*, is very spread out. It has a median of 17,400 and a 95% probability interval of (5,340, 32,300). The existing prior, q_2, is far more precise. The resulting pooled prior is similar to, but 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 premodel distributions described earlier in this section. The premodel distributions of P_0 and MSYR are the gamma priors, and the premodel distribution of ROI is the transformed t-likelihood only. For P_{1993}, the premodel distribution is the normal approximation to the Bayes empirical Bayes posterior from the 1993 census. In Table 1, the posterior distribution of P_{1993} is almost identical to the premodel distribution. This is unsurprising, because the likelihood (L_2) is dominant; this was the best piece of premodel information, and the priors on other parameters affect the inference very little. The posterior is very slightly more spread out than the premodel distribution; this is due to the influence of q_1^*, the diffuse induced prior for P_{1993}.

For P_0, the posterior variance is considerably lower than the prior variance; similarly, the posterior distribution of MSYR is more precise than the prior. For ROI, the melding of information leads to a lower estimate than that suggested by the premodel likelihood. The model maps each draw from the priors on inputs to values of P_{1978} and P_{1993}, and hence to a value of ROI. The distribution of ROI obtained from these draws suggests a slightly lower rate of increase than the premodel likelihood. As a result, when these draws are combined with the likelihood, we obtain the observed posterior distribution. Histograms of the posterior samples are shown in Figure 5. 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.

Table 1. Posterior and Premodel Quantiles of Inputs and Outputs of the Simple PDM

		Results	
Parameter	*Quantile*	*Posterior*	*Premodel*
P_{1993}	.025	7,072	7,095
	.5	8,196	8,200
	.975	9,322	9,305
ROI	.025	.0105	.0146
	.5	.0204	.0307
	.975	.0318	.0469
P_0	.025	12,057	8,264
	.5	14,346	15,004
	.975	17,980	30,373
MSYR	.025	.0113	.0096
	.5	.0213	.0211
	.975	.0333	.0394

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

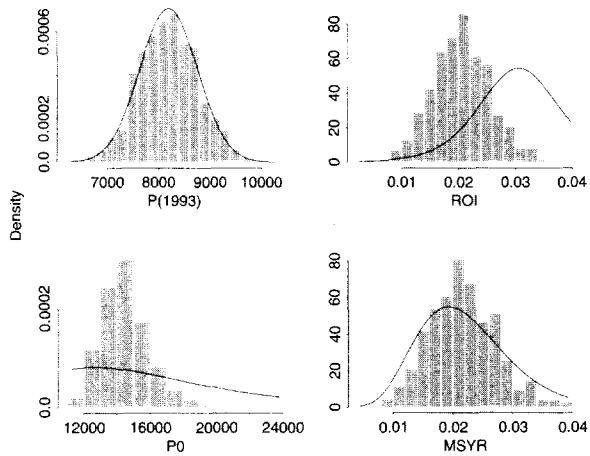

Figure 5. Histograms of the Bayesian Melding Posterior Distributions for Parameters of the PDM for Bowhead Whales. The premodel distributions are shown as solid lines.

5. COMPARISON VIA SIMULATION

We performed some simulations using the bowhead whale model to compare the Bayesian melding approach to a simpler Bayesian alternative where we placed priors on the inputs only. With melding, the posterior distribution of the inputs $\theta = (P_0, \text{MSYR})$ is

$$\pi^{[\theta]}(\theta) \propto \tilde{q}^{[\theta]}(\theta) L(M(\theta)),$$

where $\tilde{q}^{[\theta]}(\cdot)$ is the pooled prior on inputs, and $L(\cdot)$ is the joint likelihood for the output $M(\theta) = (P_{1993}, \text{ROI})$. For the simpler alternative, we used only the prior on the inputs P_0 and MSYR and ignored the existing prior (q_2) on the output P_{1993}. This eliminates the need for logarithmic pooling. Thus the corresponding posterior distribution for the simpler alternative ("no melding") is

$$\pi^{[\theta]}(\theta) \propto q_1(\theta) L(M(\theta)),$$

where $q_1(\cdot)$ is the original joint prior on inputs. The likelihood is included in the same way in both cases.

We generated 100 true values of the triple (P_0, MSYR, P_{1993}) from the posterior distribution in Table 1. For each of these 100 truths, we generated 100 values of P_{1993} (pseudodata) from a normal likelihood centered at the true P_{1993} and standard deviation equal to 626, the standard deviation of the real Bayes–empirical Bayes likelihood. We then applied the two estimation procedures to the pseudodata by plugging in each pseudodata point as the center of the likelihood with standard deviation again equal to 626. The point estimates of parameters (the posterior medians) were compared to the true values using the relative mean squared error (MSE),

$$\text{relative MSE} = \frac{1}{(100)^2} \sum_{i=1}^{100} \sum_{j=1}^{100} \left(\frac{\hat{\theta}_{ij}}{\theta_i} - 1 \right)^2,$$

where θ_i is the true value of the parameter θ in truth set i and $\hat{\theta}_{ij}$ is the estimate of θ_i using the jth simulated data set under truth set i.

The total relative MSEs for each parameter using the two approaches are shown in Table 2. (The constant $1/(100)^2$ was omitted for convenience.) For P_{1993}, the melding approach outperforms the simpler alternative by about 23% in total relative MSE. Although the likelihood is strongly dominant in determining the posterior distribution for this parameter in both cases, there appears to be considerable gain in including the extra prior in the analysis. In other words, inclusion of the existing prior for P_{1993} improves the estimates of this parameter across various data that may have been observed. For the other parameters, the methods perform very similarly.

Table 2. Relative MSEs for Each of the Four Parameters in the PDM Using Two Bayesian Estimation Procedures on Simulated Data

	Total relative MSE		
Parameter	Melding	No melding	Ratio
P_{1993}	21.6	27.9	.77
ROI	405.1	414.7	.98
P_0	120.4	119.2	1.01
MSYR	415.1	410.9	1.01

Figure 6 shows the ratio MSE (melding)/MSE (no melding) for each of the simulated true values of each parameter. Ratios less than 1.0 indicate superior performance for the melding approach. The performance of melding for P_{1993} is noticeably better than that of the simpler procedure.

These simulations are significant in that they represent situations where the quality of the data differ (i.e., the observed P_{1993} could be close to or far from the true value) across a number of true states of the whale stock (from highly productive to endangered). We are interested in how the estimation methods perform in all of these situations, because any one of them could represent the truth.

6. DISCUSSION

We have proposed a new approach to inference from deterministic simulation models, called Bayesian melding. This retains the desirable properties of the Bayesian syn-

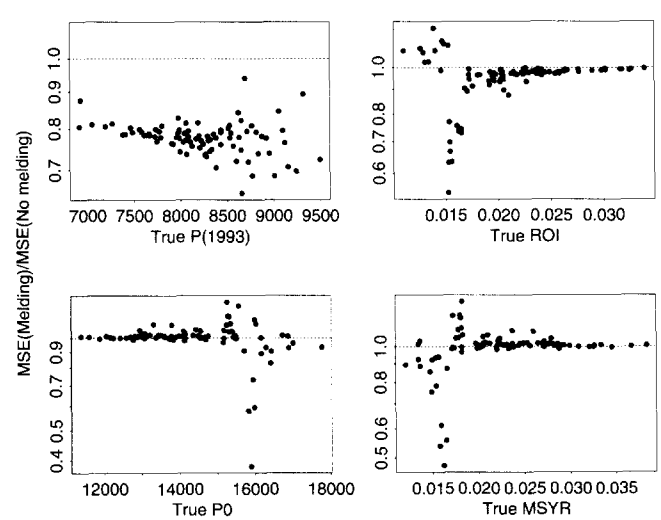

Figure 6. MSE (Melding)/MSE (No Melding), the Ratio of Relative MSEs, Across a Range of Simulated True Values for Each Parameter of the PDM. Ratios below 1.0 indicate superior performance for the melding approach. The y-axis is calibrated on the log scale.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

thesis approach of RGZ, but modifies it so as to avoid the Borel paradox to which the latter approach is subject. This exposition has focused on a PDM for whales, which is the context within which it was developed. For ease of presentation here, we used a model simpler than the complex BALEEN II model used by the IWC for bowhead assessment. Bayesian melding has also been applied to the larger model, and 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.) The method has also been applied to a model for tree growth by Green, McFarlane, Valentine, and Strawderman (1999), and it has potential applications in epidemiology and environmental risk assessment.

6.1 Model Uncertainty

We have assumed the form of the model M to be fixed or agreed on. However, there may be uncertainty about this, and taking it into account can be important. Although it is not the main topic of this article, we indicate some possible approaches.

A significant form of model uncertainty arises when there are several model formulations, which cannot be distinguished between decisively on the basis of the available information, but which may lead to different predictions about quantities of interest. In the full BALEEN II model used by the IWC, for example, density dependence is assumed to operate on fertility, but it might equally well operate on mortality rather than on fertility, or on both.

One approach to this problem is again Bayesian. It starts by computing the posterior probabilities of the models considered; this is fairly easy to do using the SIR algorithm (e.g. Kass and Raftery, 1995, eq. 10). Bayesian model averaging is then used to compute the posterior density of the quantity of interest, ψ, taking into account model uncertainty; this is a weighted average of the posterior densities of ψ under the different models, weighted by their posterior model probabilities. (See Hoeting, Madigan, Raftery, and Volinsky 1999 for a review of Bayesian model averaging.) If one model has a posterior model probability that is close to 1, or if the posterior distributions of ψ are similar under all the models with substantial posterior probabilities, then choosing a single model rather than averaging over the models will give a good approximation to the full Bayesian model-averaging solution.

An application of this approach arises in a policy- or decision-making context when the best decision depends on a model assumption about which there is disagreement. Then the weighted average of the resulting posterior distributions from Bayesian model averaging may give a result that is acceptable to both sides in the debate. An example from the IWC was given by Raftery and Schweder (1993) in the contentious context of the assessment of North Atlantic minke whales.

A second form of model uncertainty flows from the more general concern that the approach here assumes the model M to be true (at least approximately), and this may not be the case. Perhaps the most common approach to this problem in statistics is to check the model in various ways and change it if there seem to be systematic discrepancies (e.g., Box and Jenkins 1976). We discussed some approaches to model checking for deterministic simulation models in earlier work (Poole and Raftery 1998). Another approach is via model expansion, in which a neighborhood of the current model is defined, and uncertainty about model form within this neighborhood is propagated through to final inference about quantities of interest in the standard Bayesian way, via integration (e.g., Draper 1995; Perrichi, Sansó, and Smith 1993; Smith 1983).

A third, little discussed, source of model uncertainty is the possibility that the model itself, the algebra specifying it, or the computer implementation is in serious error without the user being aware of it. A formal way of taking this into account may be available, based on the following thought experiment, modified from a proposal of RGZ.

Imagine first that we did not know about the model. Then our conclusions about the inputs, θ, would be based only on the prior and likelihood for θ alone, and would be summarized by the distribution

$$r^{[\theta]}(\theta) \propto q_1(\theta)L_1(\theta),$$

which we call the *premodel posterior distribution* of θ. Now suppose instead that we did know about the model, but that just before we had to make a decision we learned that the model was seriously flawed, with no time to revise the model or build a new one. Then it might well be reasonable to revert to the premodel posterior distribution and base decisions on that. If such errors occur with probability ε, then this suggests using the "hyperposterior distribution,"

$$\pi^{[\theta]^*}(\theta) = (1-\varepsilon)\pi^{[\theta]}(\theta) + \varepsilon r^{[\theta]}(\theta),$$

a mixture of the premodel and postmodel posterior distributions of the inputs, where $\pi^{[\theta]}(\theta)$ comes from (7). Inference about the quantities of interest, ψ, would then be made by using the induced distribution of ψ corresponding to the hyperposterior distribution of the inputs, $\pi^{[\theta]^*}(\theta)$.

In the bowhead case, the issue of formally accounting for model uncertainty was not a critical one at the IWC. This is because agreement that the model used was reasonable for management purposes had developed over a period of perhaps 20 years of studies, sensitivity analyses, presented papers, and Scientific Committee debates.

6.2 Choice of α

An issue that remains unresolved is how best to choose α, the pooling weight. In the original context 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 (Gen-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

est 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). As noted earlier, we have used this fact as the basis for our choice of $\alpha = .5$ 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;* precision is accounted for by the variances of the respective distributions. It seems reasonable to assume that the information about inputs and outputs will be equally reliable, because they are both assessed and expressed quantitatively by the same expert. Figure 4 illustrates how even when $\alpha = .5$, the pooled prior will be dominated by the more precise source of information.

However, values other than $\alpha = .5$ 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). One could also think of estimating α as a parameter of the model in the usual Bayesian way.

6.3 Density Estimation and Computation

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. This 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 (1999) have developed an adaptive method that involves direct numerical approximation of Jacobians. Their method has led to improved accuracy and efficiency in some examples.

We have developed our methods for a family of simulation models that can be run reasonably fast on a computer. 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. Efficiency is a major issue, and methods that make each run of the model count much more will be required. 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, app. 4) (see also Levy, Clayton, and Chesters 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 involves 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 because it requires the availability of code to compute the (unnormalized) analytic 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 known that designing MCMC algorithms for this kind of situation can be extremely delicate. The design of an algorithm that overcomes these obstacles would be a worthwhile topic for further research.

APPENDIX: PROOFS

Here we give the proofs of Theorems 1 and 4. Proofs of the other results have been given in an earlier work (Poole 1999).

Proof of Theorem 1

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 $h''(x) = \alpha(\alpha - 1)x^{\alpha-2} \le 0$. 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

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

because

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

So $\int q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha}\, d\phi \le 1 < \infty$, 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.

Proof of Theorem 4

Let $\tau = (\phi, \gamma)$ and denote $\tau = G(\theta)$ so that $\theta = G^{-1}(\tau)$. Consider the integral of $\tilde{q}^{[\theta]}(\theta)$ in (16) over Θ,

$$\begin{aligned}\int_\Theta \tilde{q}^{[\theta]}(\theta)\, d\theta &= k_\alpha \int_\Theta q_1(\theta)\left(\frac{q_2(M(\theta))}{q_1^*(M(\theta))}\right)^{1-\alpha} d\theta \\ &= k_\alpha \int_\Phi \int_\Gamma q_1(G^{-1}(\phi,\gamma))\left(\frac{q_2(\phi)}{q_1^*(\phi)}\right)^{1-\alpha} |J|\, d\gamma\, d\phi\end{aligned}$$

by a change of variable within the integral, where $J = |d\theta/d\tau|$. Let $q^*(\phi, \gamma)$ denote the joint density of (ϕ, γ) induced by $q_1(\theta)$ and G, and observe that

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

Noting that $q_1^*(\phi) = \int q^*(\phi, \gamma)\, d\gamma$, substitution into the integral above yields

$$\begin{aligned}\int_\Theta \tilde{q}^{[\theta]}(\theta)\, d\theta &= k_\alpha \int_\Phi \int_\Gamma q^*(\phi,\gamma)\left(\frac{q_2(\phi)}{q_1^*(\phi)}\right)^{1-\alpha} d\gamma\, d\phi \\ &= k_\alpha \int_\Phi q_1^*(\phi)\left(\frac{q_2(\phi)}{q_1^*(\phi)}\right)^{1-\alpha} d\phi \\ &= k_\alpha \int_\Phi q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha}\, d\phi \\ &= 1\end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

by Theorem 1. 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 $\phi = M(\theta)$.

[Received April 1999. Revised February 2000.]

REFERENCES

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

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control* (2nd ed.), San Francisco: Holden-Day.

Caswell, H. (1976), "The Validation Problem," in *Systems Analysis and Simulation in Ecology*, ed. B. C. Patten, New York: Academic Press, pp. 313–325.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, 86, 953–963.

Draper, D. (1995), "Assessment and Propagation of Model Uncertainty" (with discussion), *Journal of the Royal Statistical Society*, Ser. 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, Amsterdam: North-Holland, pp. 183–201.

Genest, C. (1984), "A Characterization Theorem for Externally Bayesian Groups," *The Annals of Statistics*, 12, 1100–1105.

Genest, C., McConway, K. J., and Schervish, M. J. (1986), "Characterization of Externally Bayesian Pooling Operators," *The Annals of Statistics*, 14, 487–501.

Genest, C., and Zidek, J. V. (1986), "Combining Probability Distributions: A Critique and an Annotated Bibliography" (with discussion), *Statistical Science*, 1, 114–148.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996), *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," doctoral dissertation, University of Washington, Dept. of Statistics.

——— (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 Strong Nonlinear Relationships," *Journal of the American Statistical Association*, 91, 132–141.

Givens, G. H., and Roback, P. J. (1999), "Logarithmic Pooling of Priors Linked by a Deterministic Simulation Model," *Journal of Computational and Graphical Statistics*, 8, 452–478.

Green, E. J., McFarlane, D. W., Valentine, H. T., and Strawderman, W. E. (1999), "Assessing Uncertainty in a Stand Growth Model by Bayesian Synthesis," *Forest Science*, 45, 528–538.

Guttorp, P., and Walden, A. T. (1987), "On the Evaluation of Geophysical Models," *Geophysical Journal of the Royal Astronomical Society*, 91, 201–210.

Hodges, J. S. (1987), "Uncertainty, Policy Analysis and Statistics" (with discussion), *Statistical Science*, 2, 259–291.

——— (1991), "Six (or so) Things You can do With a Model," *Operations Research*, 39, 355–365.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial" (with discussion), *Statistical Science*, 14, 382–401.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

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 Journal*, 6, 457–468.

Madansky, A. (1978), "Externally Bayesian Groups," unpublished manuscript, University of Chicago.

Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993), "Bayesian Design and Analysis of Computer Experiments—Use of Derivatives in Surface Prediction," *Technometrics*, 35, 243–255.

O'Hagan, A. (1991), "Bayes-Hermite Quadrature," *Journal of Statistical Planning and Inference*, 29, 245–260.

Perrichi, L. R., Sansó, B., and Smith, A. F. M. (1993), "Posterior Cumulant Relationships in Bayesian Inference Involving the Exponential Family," *Journal of the American Statistical Association*, 88, 1419–1426.

Poole, D. (1999), "Bayesian Inference for Noninvertible Deterministic Simulation Models, With Application to Bowhead Whale Assessment," doctoral dissertation, University of Washington, Dept. of Statistics.

Poole, D., and Raftery, A. E. (1998), "Inference for Deterministic Simulation Models: The Bayesian Melding Approach," Technical Report 346, University of Washington, Dept. of Statistics, ⟨http://www.stat.washington.edu/tech.reports/tr346.ps⟩

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," *Report of the International Whaling Commission*, 46, 481–491.

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 232, University of Washington, Dept. of Statistics.

——— (1995), "Inference From a Deterministic Population Dynamics Model for Bowhead Whales" (with discussion), *Journal of the American Statistical Association*, 90, 402–430.

Raftery, A. E., and Schweder, T. (1993), "Inference About the Ratio of Two Parameters, With Application to Whale Censusing," *The American Statistician*, 47, 259–264.

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, New York: Springer-Verlag, pp. 163–240.

——— (1998), "Estimating Bowhead Whale Population Size and Rate of Increase From the 1993 Census," *Journal of the American Statistical Association*, 93, 451–463.

Rubin, D. B. (1987), Comment, on "The Calculation of Posterior Distributions by Data Augmentation," by M. Tanner and W. H. Wang, *Journal of the American Statistical Association*, 82, 543–546.

——— (1988), "Using the SIR Algorithm to Simulate Posterior Distributions," in *Bayesian Statistics 3*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 395–402.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments" (with discussion), *Statistical Science*, 4, 409–435.

Sampson, P. D., and Guttorp, P. (1999), "Operational Evaluation of Air Quality Models," in *Environmental Statistics: Analyzing Data for Environmental Policy*, Chichester, U.K.: Wiley, pp. 33–51.

Schweder, T. (1995), Comment, on "Inference from a Deterministic Population Dynamics Model for Bowhead Whales," by A. E. Raftery, G. H. Givens, and J. E. Zeh, *Journal of the American Statistical Association*, 90, 420–423.

Schweder, T., and Hjort, N. L. (1996), "Bayesian Synthesis or Likelihood Synthesis—What Does Borel's Paradox Say?," *Report of the International Whaling Commission*, 46, 475–480.

Smith, A. F. M. (1983), "Bayesian Approaches to Outliers and Robustness," in *Specifying Statistical Models*, eds. J. P. Florens, M. Mouchart, J. P. Raoult, L. Simar, and A. F. M. Smith, Berlin: Springer-Verlag, pp. 13–55.

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

Terrell, G. R. (1990), "The Maximal Smoothing Principle in Density Estimation," *Journal of the American Statistical Association*, 85, 470–477.

Wolpert, R. L. (1995), Comment, on "Inference from a Deterministic Population Dynamics Model for Bowhead Whales," by A. E. Raftery, G. H. Givens, and J. E. Zeh, *Journal of the American Statistical Association*, 90, 426–427.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.