Inference from a Deterministic Population Dynamics Model for Bowhead Whales: Rejoinder

Adrian E. Raftery; Geof H. Givens; Judith E. Zeh

Journal of the American Statistical Association, Vol. 90, No. 430 (Jun., 1995), 427-430.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199506%2990%3A430%3C427%3AIFADPD%3E2.0.CO%3B2-D

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Journal of the American Statistical Association is published by American Statistical Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/astata.html.

Journal of the American Statistical Association
©1995 American Statistical Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

http://www.jstor.org/
Tue Oct 14 20:29:49 2003

posterior (rather than premodel and postmodel) distributions and enjoy the usual benefits of coherence, stopping-rule independence, likelihood-principle compliance, and straightforward interpretability of Bayesian inference (see, for example, Berger and Wolpert 1988), at the cost of some additional effort in modeling and some in computation, due to the addition to the model of a single new parameter. But an alarming fourth possibility is that the expert may have suspicions or evidence that the *model is wrong*—evidence that would be ignored or misinterpreted by an automatic procedure like taking the trace of the distribution—or (closely related) that the expert may have *misjudged* what values of (θ, ϕ) were plausible and so may have given a distorted view of his or her premodel distribution (i.e., one detailed and accurate far away from the set $\mathcal{M}$, but casual and vague near $\mathcal{M}$, in the only place where (in light of the model) careful attention is called for). In either of these latter cases it will be necessary for the statistician and the elicitee to discuss in more detail the model and its implications before a satisfactory analysis will be possible.

[Received September 1994.]

ADDITIONAL REFERENCES

Berger, J. O., and Wolpert, R. L. (1988), *The Likelihood Principle* (2nd ed.), Hayward, CA: Institute of Mathematical Statistics.
Billingsley, P. (1986), *Probability and Measure* (2nd ed.), New York: John Wiley, p. 448.
Lindley, D. V., Tversky, A., and Brown, R. V. (1979), "On the Reconciliation of Probability Assessments," *Journal of the Royal Statistical Society*, Ser. A, 142, 146–180.

Rejoinder

Adrian E. RAFTERY, Geof H. GIVENS, and Judith E. ZEH

We would like to thank all five discussants for excellent comments. All have raised important points. We would like to thank Schweder and Tuljapurkar and Lee for pointing out several relevant references that predate Speed (1983).

1. THE 1994 IWC MEETING

We first submitted this article to *JASA* in June 1992, and at that point the IWC SC had reviewed the general ideas behind our approach but had not decided to use it for assessments. At the 1993 IWC SC meeting there was further methodological review, and a consensus was reached that the method was acceptable in principle.

It was the 1994 meeting, held in May 1994 in Puerta Vallarta, Mexico, that decided to use the method as the basis for setting the bowhead quota. Before that meeting, we wrote an article (Givens, Zeh, and Raftery 1994) that updated the present article by including new information in the premodel distribution and by using a slightly different PDM, which had been adopted by the IWC as the core PDM for the management of commercial whaling. We used conservative premodel distributions, leading to a lower 5% postmodel quantile for RY of 54.

Before the meeting, a paper by Butterworth and Punt (1994) (cited by Buckland) was circulated. This criticized our method on several grounds, including sensitivity to the premodel distribution and the derivation of the premodel distribution of P_0. We responded in two works (Givens, Raftery, and Zeh 1994b and Givens, Zeh, and Raftery 1994, sec. 6) both of which addressed sensitivity to the premodel distribution. Further concerns and responses were contained in a welter of working papers produced at the meeting, several of which will be published in revised form (IWC 1995).

The discussion was lively and thorough and has been summarized by Schweder. Extensive sensitivity analyses were carried out at the meeting itself, facilitated by our fast reweighting method (Givens et al. 1994a). After much discussion, it was decided to adopt our method.

This was followed by two remarkable two-hour sessions, in which the entire SC agreed on a much revised premodel distribution after a good deal of argument. This may be the first time that such a committee, consisting of about 100 scientists with conflicting views, has engaged successfully in a serious collective prior elicitation exercise. The resulting assessment led to a substantially increased lower 5% postmodel quantile for RY of 104, which was then used as the basis for setting the quota.

The premodel distributions of some other inputs and outputs also differed from those used in this article. The premodel distribution of rate of increase was revised to include the 1993 census data (Raftery, Zeh, and Givens 1995) to give a 95% premodel interval of [1.4%, 4.7%]. This was shifted and narrowed by the Bayesian synthesis approach to a postmodel distribution with 95% interval [.9%, 3.4%].

As a practical matter, once the SC did agree on the premodel distribution, the results of the assessment were available a few hours later. Hopefully, this positive experience will allay Buckland's worries. It is worth pointing out that the routines previously used by the IWC (usually referred to collectively as HITTER/FITTER) have run into numerical problems on several occasions, impeding the timely

© 1995 American Statistical Association
Journal of the American Statistical Association
June 1995, Vol. 90, No. 430, Applications and Case Studies

completion of assessments (e.g., IWC 1992, p. 63; 1993, p. 249).

2. METHODOLOGY

2.1 The Population Dynamics Model

Buckland and Tuljapurkar and Lee raised the issue of the realism and appropriateness of the PDM itself. The choice of PDM was essentially made by the IWC SC rather than by us; our methodology could be applied just as easily to other PDM's.

Buckland is concerned about the use of a model that uses the concept of an initial population size, taken to represent a fixed carrying capacity, both because the carrying capacity may vary over time and because it is hard to estimate the initial population size. He does, however, suggest that this may be more of a problem for whale stocks other than the bowhead. He recommends management on the basis of current abundance.

The big problem with this in the bowhead case is that the information in the series of ice-based censuses is too uncertain to allow reliable management. A confidence interval for the current rate of increase based on the 1978–1988 census data alone went down as far as .1%. The IWC SC believed that management could be improved by also taking into account the considerable biological and other information available, and using this does indeed lead to more precise inferences. This was the motivation for our work.

We feel that the PDM used here is broadly appropriate for bowheads. The Arctic and sub-Arctic marine environment in which they live does not seem to have changed in a major way due to human activity or climate change over the past 150 years. Also, it is important to have an idea of the extent to which the population has recovered, given the massive overexploitation of the stock in the late 19th century. Our estimate of depletion is indeed imprecise (with a 95% interval .31 to .54), but it does show clearly that the stock remains quite depleted, and the uncertainty about it is included in inference about RY. One of the main reasons for studying the bowhead is that it was the first of the overexploited large whales for which commercial hunting stopped (in 1915), and so its current depletion gives an indication of whether and how fast other whale stocks can recover. Our inferences are relatively insensitive to reasonable changes in the pre-model distribution of initial population size (Givens, Zeh, and Raftery 1994, sec. 6; Givens, Raftery, and Zeh 1994b).

If the carrying capacity has indeed been changing, Tuljapurkar and Lee indicate one way in which this could be handled, by allowing for a stochastically changing carrying capacity. In fact, this can easily be handled within the Bayesian synthesis framework; for each draw of the inputs, one simply simulates a stochastic trajectory and then resamples as before. There is no need to simulate a large number of runs for each draw of the inputs.

Buckland pointed out that in our PDM, density dependence is present only in fertility but not in mortality. Given the limited information, it seems unlikely that we could distinguish clearly between these possibilities, and it also seems unlikely that modeling density dependence in mortality and/or dependence in fertility would change our conclusions by much. But this could be assessed formally using Generalized Bayes Factors as in Section 2.4 of our article, and if the model comparison were indecisive and the results sensitive to the choice, it would be best to take account of model uncertainty explicitly by model averaging, as discussed in Section 2.5 of our article.

2.2 Catch History

Buckland suggested that by using the catch history $C = (C_1, \ldots, C_T)$ to develop the premodel distribution of P_0, and also in the model itself, the catch history data are being used twice, and that this is invalid. But, the entire procedure is conditional on the catch history data, assumed by the IWC to be largely correct. Thus the law of total probability gives

$$\pi(\psi) = \pi(\psi \mid C) = \int \pi(\psi \mid P_0, C)\pi(P_0 \mid C)\, dP_0,$$

and so the use of C to derive both the premodel distribution of P_0 and the postmodel distribution of ψ is valid.

Buckland also expressed concern about the possibility of errors in the catch history data, and this is indeed a valid concern. But it is not clear that bias in estimates from 19th century whaling records would necessarily be downward (unlike Russia in the 1960s). This concern could be addressed by allowing for measurement error in the catch data, and our method can, in principle, do this easily. C simply becomes another input parameter. We are working on ways to implement this in practice.

2.3 Computation

Schweder and Buckland pointed out that the SIR algorithm with the premodel distribution as initial sampling function is inefficient. For our application this does not matter, as the PDM runs fast enough to allow us to do the necessary 200,000 runs. But for models that take a long time to run this would be prohibitive, and it would be necessary to find a more efficient method. We discussed several other possibilities in Section 6.3 of our article, but stuck to our simple SIR algorithm because it was easy to use. To our list of possible more efficient methods, Schweder usefully adds Latin hypercube sampling.

3. BOREL PARADOX

Wolpert pointed out that different but mathematically equivalent parameterizations of the same PDM can lead to different postmodel distributions. At first sight this seems to be a fatal flaw in the method, and indeed Wolpert argues that our approach is "dangerous." But we feel that he has overstated the practical implications of his mathematical point, and that when it is reasonable to use a deterministic PDM, the sensitivity of the results to reasonable reparameterizations will be small.

The use of a deterministic model is based on the assumption that the random variation accounts for only a small amount of the overall uncertainty about quantities of interest. Wolpert himself provides a heuristic argument that when

Table 1. Postmodel Quantiles in the Simple Linear-Growth Bowhead PDM for 1915–1988

Parameter	Quantile	Parameterization 1	Parameterization 2
ρ	.05	.0089	.0088
	.50	.0179	.0175
	.95	.0269	.0269
P_{1915}	.05	1,111	1,122
	.50	2,120	2,175
	.95	4,109	4,026
P_{1988}	.05	6,638	6,718
	.50	7,764	7,849
	.95	8,947	8,997

NOTE: Parameterization 2 refers to the model with population expressed in logarithmic units.

this assumption holds, the sensitivity of the postmodel distribution to reparameterization will be small. For when the random variation is explicitly taken into account, the model becomes stochastic rather than deterministic, and then, as Wolpert points out, the postmodel distribution is just a standard Bayesian posterior distribution and so is invariant to reparameterization. Because the random variation has little effect on the final inference, the full posterior distribution should be close to the postmodel distribution (however parameterized).

As a check on this reasoning for the present application, we used the simple linear-growth PDM that Wolpert analyzed, for parameter values representative of the bowhead case. The model is $P_t = P_0 e^{\rho t}$. This does provide a reasonable first approximation to the bowhead trajectory from the end of commercial whaling in 1915 to 1988, as we have modeled it. Throughout that period, P_t/P_{1848} was probably below .54 (Fig. 6), and so the fertility rate was roughly constant at its maximum value (Fig. 1). Because mortality was assumed to be constant over time, the growth rate ρ will also have been roughly constant over that period. There are two inputs, $\theta = (\rho, P_{1915})$, and one output, $\phi = P_t$ (here we allow for the fact that the growth rate ρ is unknown).

We took ρ, P_{1915}, and P_{1988} to be premodel independent with $\rho \sim N(.031, .015^2)$ (based on the 1978–1988 censuses), $P_{1915} \sim \text{Lognormal}(7.8, .4^2)$ (based on P_{1848} and the catch history), and $P_{1988} \sim N(7{,}800, 700^2)$ (based on the 1988 census).

Then the Bayesian synthesis algorithm is

1. Draw n values of $\theta = (\rho, P_{1915})$ from its premodel distribution.
2. For each one, compute $\phi = P_{1988} = P_{1915} e^{73\rho}$ (because $1988 - 1915 = 73$).
3. Resample m values from these n with weights $w = N(P_{1988}; 7{,}800, 700^2)$ (the density of P_{1988} according to a $N(7{,}800, 700^2)$ distribution).

We used $n = 20{,}000$ and $m = 5{,}000$, obtaining about 1,500 unique points in the final sample.

As Wolpert pointed out, if instead the model is written as $\log P_{1988} = \log P_{1915} + 73\rho$, with $\theta = (\rho, \log P_{1915})$ and $\phi = \log P_{1988}$, then each weight gets multiplied by the corresponding simulated value of P_{1988}, with the algorithm being unchanged otherwise. The results are of course different, but is the difference worth worrying about?

In Table 1 we show quantiles of the postmodel distributions of the three parameters under the two different parameterizations. These are very close, and in no case do they differ by more than one-eighth of a postmodel standard deviation; even these small differences are partly due to Monte Carlo variability. In Figure 11 we show the two estimated postmodel densities of ρ; they are virtually indistinguishable.

It thus seems that sensitivity of the postmodel distribution to the PDM's parameterization is very small here. Also, in our research we did analyze several different PDM's that were parameterized differently but that modeled the same population processes, and found little sensitivity.

Wolpert is right to point out this theoretical possibility, which could become a practical concern if random variation accounted for a large part of the uncertainty and were ignored. His discussion is useful in pointing out that one way to find out if this is so is to do the analysis for several parameterizations and check that they give similar results. But we do not agree that one should *always* incorporate random variation in mechanistic models. Science and engineering abound with situations where this is more trouble than it is worth, and where deterministic models are simpler and adequate for the purpose at hand. The bowhead assessment is one such case.

4. OTHER METHODS

Buckland has suggested simulated inference as an alternative approach. This seems rather similar to our Bayesian synthesis method, but accounts explicitly for model uncertainty while not including premodel information about outputs. We applaud the incorporation of model uncertainty; this could also be incorporated in the Bayesian synthesis approach via Bayesian model averaging using the Generalized Bayes Factors, as pointed out in Section 2.5 of our paper. As a technical remark, the Bayes Information Criterion seems more likely than Akaike's Information Criterion to give a good approximation to the appropriate weights for model averaging (e.g., Kass and Raftery 1995). We feel that it is valuable to take into account premodel information about outputs (other than that expressible as a likelihood for data).

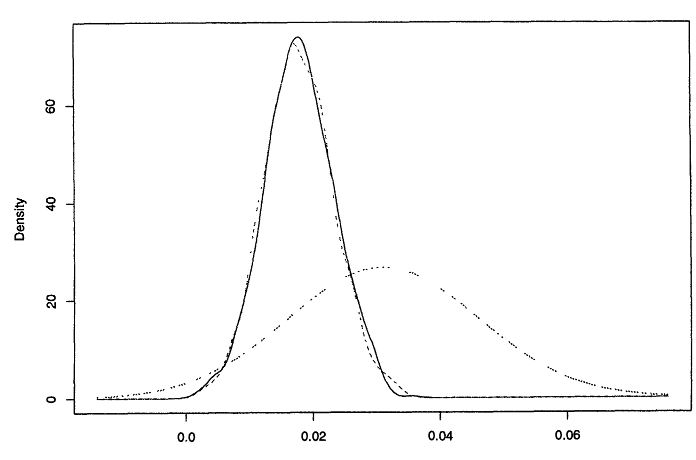

Figure 11. Postmodel Densities of ρ Under Two Parameterizations in the Linear-Growth Bowhead Model. The premodel density is also shown. ——, postmodel density of rho, parameterization 1; – – –, postmodel density of rho, parameterization 2; · · · · ·, premodel density.

Buckland said that his motivation for considering this alternative is that the Bayesian synthesis approach may not be reliably implementable on a personal computer, but all the computations using the method at the 1994 IWC meeting were done quickly and reliably on a personal computer.

Schweder has proposed and used the maximum simulated likelihood approach, and this seems to be an interesting alternative that should often give results close to those of the Bayesian synthesis. But as Schweder indicates, this method may have problems when Φ is highly nonlinear, as is often the case in PDM's.

Schweder mentions several other papers in which simulation was used for inference. But most of these are about stochastic models rather than deterministic ones, and then our approach reduces to standard Bayesian inference, for which many computational tools have been developed.

Finally, Tuljapurkar and Lee have pointed out that our approach can be viewed as a tool for parameter estimation in models with incomplete or partial information. This may well be a helpful way of thinking about the method and indicating new situations where it might be useful.

[Received November 1994.]

ADDITIONAL REFERENCES

Givens, G. H., Raftery, A. E., and Zeh, J. E. (1994a), "A Reweighting Approach for Sensitivity Analysis Within the Bayesian Synthesis Framework for Population Assessment Modeling," *Report of the International Whaling Commission,* 44, 377–384.

——— (1994b), "Response to Comments by Butterworth and Punt in SC/46/AS2 on the Bayesian Synthesis Approach," *Report of the International Whaling Commission,* 45, to appear.

Givens, G. H., Zeh, J. E., and Raftery, A. E. (1994), "Assessment of the Bering–Chukchi–Beaufort Seas Stock of Bowhead Whales Using the BALEEN II Model in a Bayesian Synthesis Framework," *Report of the International Whaling Commission,* 45, to appear.

International Whaling Commission (1992), "Report of the Scientific Committee," *Report of the International Whaling Commission,* 42, 51–79.

——— (1993), "Report of the Special Meeting of the Scientific Committee on the Assessment of Gray Whales," *Report of the International Whaling Commission,* 42, 241–253.

——— (1995), "Report of the Scientific Committee," *Report of the International Whaling Commission,* 45, to appear.

Raftery, A. E., Zeh, J. E., and Givens, G. H. (1995), "Revised Estimate of Bowhead Rate of Increase," Appendix 6 of Annex F, Report of the Scientific Committee, *Report of the International Whaling Commission,* 45, to appear.