Brief Communication Arising re: "Quantification of modelling uncertainties in a large ensemble of climate change simulations"

Jonathan Rougier*

Department of Mathematical Sciences

University of Durham, UK

November 26, 2004

Murphy et al.[1,2] compute a probability distribution for climate sensitivity, using actual climate data and evaluations of a General Circulation Model (GCM). The purpose of this note is to clarify the inferential method they have adopted, and to point out some of the choices they have made, which might be questionable. This should make it easier for other research groups to duplicate their results, and perhaps to improve on them.

In probabilistic terms the Murphy et al. prediction is a probability distribution for climate sensitivity conditioned on the observed values for the the climate data. The starting point for this prediction is a joint distribution

*Department of Mathematical Sciences, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, UK. Tel +44(0)191 334 3331, fax +44(0)191 334 3051, email J.C.Rougier@durham.ac.uk.

over the climate data and climate sensitivity, and the purpose of the GCM is to help construct this joint distribution. To do this it is necessary to provide a statistical model that describes how the GCM and the climate are related, or, to put it another way, to show how evaluations of the GCM at different parameterisations are informative about climate itself.

Denote by z the vector of quantities for which we have climate observations, and λ the value of climate sensitivity, which we wish to predict. We can match these quantities with outputs from the GCM, denoted f: let $f_1, \ldots, f_{k-1}$ match the elements of z, and f_k match λ. The GCM parameters are denoted by the p-vector $x \in \mathcal{X}$. Therefore $f_k(x)$ denotes the climate sensitivity output of the GCM at the vector of model parameters x.

In order to relate the GCM and the climate vector (z, λ) we may adopt what is sometimes known in the statistics literature as the 'best input' approach[3]. This asserts that there is a 'best input' x^* somewhere in $\mathcal{X}$ with the property that the prediction error

$$\epsilon \triangleq (z, \lambda) - f(x^*) \tag{1}$$

is an uncertain (i.e. random) quantity that is independent of all other uncertain quantities. The k-vector ϵ is known as the simulator *discrepancy*. It is commonly assumed to have zero mean (unless the simulator is known to be biased in a certain direction). The variance of ϵ is a measure of the model's imperfection, since if we chose to set $\mathsf{Var}(\epsilon) = \mathbf{0}$ we would be asserting that somewhere in $\mathcal{X}$ is a choice for which the model output exactly matches the climate value. Typically we do not believe that our current climate models

are perfect in this sense, and so we specify a positive $\mathsf{Var}\big(\epsilon\big)$. The value of x^* is uncertain, but we have knowledge about it that can be described by a probability distribution.

In order to make inferences about λ we will want to try candidate values for x^* spanning the whole of the model parameter space $\mathcal{X}$. It is not practical to evaluate the GCM at every candidate value; for this reason Murphy et al. provide a statistical approximation for the GCM:

$$f_i(x) = \mu_i(x) + \delta_i \qquad i = 1, \ldots, k \tag{2a}$$

where δ is an error vector with mean zero. The mean function $\mu(\cdot)$ is chosen to be additive and piecewise linear in the model parameters:

$$\mu_i(x) \stackrel{\triangle}{=} \alpha_i + \sum_{j=1}^{p} \left\{ \mathbf{1}\big(x_j \geq \bar{x}_j\big)\, \beta_{ij}^{+} + \mathbf{1}\big(x_j < \bar{x}_j\big)\, \beta_{ij}^{-} \right\} \big(x_j - \bar{x}_j\big), \tag{2b}$$

where $\bar{x}_j$ is the standard value for model parameter j, and $\mathbf{1}\big(\cdot\big)$ is the indicator function. Values of α_i, β_{ij}^{+}, β_{ij}^{-} and $\mathsf{Var}\big(\delta\big)$ are determined from an ensemble of GCM evaluations. Statisticians would refer to (2) as an *emulator* of the GCM, although in practice a more general approach would be adopted[4,5,6].

Now it is possible to describe how we estimate the *prior predictive distribution* of climate sensitivity. This corresponds to the "Unweighted PDF" in Murphy et al., Fig. 3, p. 770. The prior predictive distribution is inferred from the choices we make for the distributions of x^*, δ and ϵ, since the 'best

input' approach tells us that

$$\lambda \equiv \mu_k(x^*) + \delta_k + \epsilon_k. \tag{3}$$

Note that both error components are necessary: δ_k because (2b) is not a perfect representation of the GCM, and ϵ_k because the GCM is not a perfect model for the climate. To estimate this distribution we estimate the probability of λ falling into a interval L for each $L \in \mathcal{L}$, where $\mathcal{L}$ is a partition of the real line. There are many ways to do this calculation[7,8]. The simplest is based on Monte Carlo integration:

$$\begin{aligned} \mathsf{Pr}\big(\lambda \in L\big) &= \int_{\mathcal{X}} \mathsf{Pr}\big(\lambda \in L \mid x^*\big)\, dF(x^*) \\ &\approx m^{-1} \sum_{v=1}^{m} \mathsf{Pr}\big(\lambda \in L \mid x^* = X_v\big) \qquad X_v \overset{\text{iid}}{\sim} F, \end{aligned} \tag{4}$$

where F denotes the distribution function of x^*. It is convenient to choose gaussian distributions for both δ and ϵ, because in this case

$$\mathsf{Pr}\big(\lambda \in L \mid x^*\big) = \int_L \phi\big(\lambda\, ;\, \mu_k(x^*), \sigma_k^2 + \tau_k^2\big)\, d\lambda \tag{5}$$

where $\phi(\cdot\, ;\, \cdot, \cdot)$ is the gaussian density function with given mean and variance, and $\sigma_i^2 \triangleq \mathsf{Var}\big(\delta_i\big)$ and $\tau_i^2 \triangleq \mathsf{Var}\big(\epsilon_i\big)$, for $i = 1, \ldots, k$. In other words we sample m sets of values for x^*, namely $\{X_1, \ldots, X_m\}$, compute (5) in each case, and take the arithmetic mean of the result, for each $L \in \mathcal{L}$.

How is this calculation modified if we want the *posterior* predictive distribution for λ? This corresponds to the "Weighted PDF" in Murphy et al.,

Fig. 3, p. 770. In this case we must condition λ on the climate data $\tilde{z}$, where the tilde denotes the observed value. This gives

$$\mathsf{Pr}\big(\lambda \in L \mid z = \tilde{z}\big) = c \int_{\mathcal{X}} \mathsf{Pr}\big(z = \tilde{z},\, \lambda \in L \mid x^*\big)\, dF(x^*) \tag{6a}$$

where c is the reciprocal of $\mathsf{Pr}\big(z = \tilde{z}\big)$. It is convenient if z and λ are conditionally independent given x^*, which, in the light of the gaussian choices made above, means that δ_k must be uncorrelated with $\delta_1, \ldots, \delta_{k-1}$, and similarly for ϵ. In this case

$$\begin{aligned}\mathsf{Pr}\big(\lambda \in L \mid z = \tilde{z}\big) &= c \int_{\mathcal{X}} \mathsf{Pr}\big(z = \tilde{z} \mid x^*\big)\, \mathsf{Pr}\big(\lambda \in L \mid x^*\big)\, dF(x^*) \\ &\approx \sum_{v=1}^{m} w_v\, \mathsf{Pr}\big(\lambda \in L \mid x^* = X_v\big) \qquad X_v \overset{\text{iid}}{\sim} F, \end{aligned} \tag{6b}$$

where $w_v \propto \mathsf{Pr}\big(z = \tilde{z} \mid x^* = X_v\big)$ and $\sum_{v=1}^{m} w_v = 1$. The new feature in this estimate is the use of a *weighted* rather than an arithmetic mean. The weights are proportional to the *likelihood* of x^*, i.e. they show how well each x^* in the sample explains the observed values of the climate data, $\tilde{z}$. It is convenient to treat both δ and ϵ as having uncorrelated components, i.e. choosing diagonal matrices for both $\mathsf{Var}\big(\delta\big)$ and $\mathsf{Var}\big(\epsilon\big)$, in which case

$$\mathsf{Pr}\big(z = \tilde{z} \mid x^*\big) \propto \exp\left\{ -\sum_{i=1}^{k-1} \frac{\big(\tilde{z}_i - \mu_i(x^*)\big)^2}{2\big(\sigma_i^2 + \tau_i^2\big)} \right\}. \tag{7}$$

This allows us to compute the weights straightforwardly, and from these weights to compute the posterior predictive probabilities $\mathsf{Pr}\big(\lambda \in L \mid z = \tilde{z}\big)$, for each $L \in \mathcal{L}$. Note that if the $\tilde{z}$ are observed with error, then the variance

of this error can be added to the two variances in the denominator of (7), on the assumption that the measurement error is additive and gaussian, and uncorrelated with the other uncertain quantities. Murphy et al. mention observational errors (p. 771), but choose to set the variances of these errors to zero.

Murphy et al. do not justify their calculation in probabilistic terms, as has been done here. For example, they do not explicitly state how they construct a joint distribution over (z, λ) using evaluations of their GCM. However, their calculations correspond very closely to those described above. They sample a model parameter x^* randomly from some distribution F with support $\mathcal{X}$, compute the unweighted or weighted mean of $\mathsf{Pr}\big(\lambda \in L \mid x^*\big)$ using a gaussian density function as in (5), and their weighting function—which they refer to as a likelihood—looks very similar to (7). Therefore it would be very surprising indeed if Murphy et al. had adopted a different set of choices to the 'convenient' ones described above: the 'best input' approach, simple Monte Carlo integration, and both δ and ϵ gaussian with uncorrelated components. Perhaps Murphy et al. could confirm that they had adopted these choices; or, if not, they need to describe a different probabilistic framework linking together their GCM evaluations and (z, λ) in such a way that their calculations lead to well-defined probability distributions.

We now assume that Murphy et al. concur with the probabilistic framework described above, and interpret their actions in this light. For the distribution of $x^* \in \mathcal{X}$ they choose independent and uniformly distributed components with given ranges. The τ_i^2 values in the likelihood are set according to the GCM's natural variability, as revealed on p. 3 of the supplementary

information. The f_i for $i = 1, \ldots, k-1$ are 20 year means, and these are scaled by single year natural variabilities, indicating that τ_i^2 is approximately twenty times the natural variance of model output i. There is, however, one major difference. In their likelihood, Murphy et al. appear to have ignored the contribution of the emulator error δ_i in (7), i.e. they have set $\sigma_i^2 = 0$ for $i = 1, \ldots, k-1$. This would only be appropriate if they believed that $\sigma_i^2 \ll \tau_i^2$ for all $i = 1, \ldots, k-1$. It is also important to note that in their sampling of λ they appear to have ignored the contribution of the model discrepancy ϵ_k in (3), i.e. they have set $\tau_k^2 = 0$ in (5). This means that they have computed a predictive distribution for their GCM's climate sensitivity, rather than 'true' climate sensitivity, which is not quite as stated in their introductory paragraph.

Finally, some brief remarks on how statisticians might have approached the inference differently. As already stated, the treatment of the emulator can be substantially improved: there is no need, for example, for the mean function to be restricted to being additive and piecewise linear. Second, the simple Monte Carlo estimators of (4) and (6b) are seldom used, because there are alternatives with dramatically better performance: importance sampling with a carefully tuned proposal distribution and variance reduction methods, for example, or Markov chain Monte Carlo. Third, the estimate of the diagonal of $\mathsf{Var}\big(\epsilon\big)$ seems inappropriate. The discrepancy is a property of the relationship between the model and the underlying system, while the natural variability is purely a model property. Natural variability is a component of our uncertainty, but it is more likely to feature in $\mathsf{Var}\big(\delta\big)$, since it stops us from accurately predicting $f(x')$ using $f(x)$, even when x' is in the

neighbourhood of x. Therefore there is something missing if we just scale up natural variability to estimate the variance of the discrepancy. Fourth, it is unlikely that either $\mathsf{Var}(\delta)$ or $\mathsf{Var}(\epsilon)$ have zeros off-diagonal, and there are simple ways in which we might check this. Finally, the distributional choices themselves seem a little arbitrary, which is a concern since they are the basis for the range of 2.4–5.4°C for climate sensitivity that is quoted as the main finding. A different shape for the distribution of x^* or a different way of assessing $\mathsf{Var}(\epsilon)$ might be expected to make a substantial difference to this finding, and so a sensitivity analysis would be appropriate.

Broadly speaking, however, Murphy et al. is a very welcome addition to the climate literature, showing how we may make probabilistic assessments of climate quantities by combining climate data and evaluations of a GCM. Hopefully the authors will shortly produce a sensitivity analysis for their distributional choices, and perhaps investigate suggestions from the statistical literature referenced below.

References

1. James M. Murphy, David M.H. Sexton, David N. Barnett, Gareth S. Jones, Mark J. Webb, Matthew Collins, and David A. Stainforth. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430:768–772, 2004.

2. Thomas F. Stocker. Models change their tune. *Nature*, 430:737–738, 2004.

3. M. Goldstein and J.C. Rougier. Probabilistic formulations for transferring

inferences from mathematical models to physical systems. *SIAM Journal on Scientific Computing*, 2004. In press.

4. C. Currin, T.J. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with application to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86:953–963, 1991.

5. M.C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B*, 63:425–464, 2001. With discussion.

6. P.S. Craig, M. Goldstein, J.C. Rougier, and A.H. Seheult. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96:717–729, 2001.

7. C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. New York: Springer, 1999.

8. M. Evans and T. Swartz. *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford: Oxford University Press, 2000.