Bayesian Analysis (2008) **3**, Number 1, pp. 39–44

Comment on article by Sansó et al.

Dave Higdon* and James Gattiker†

Sansó, Forest and Zantedeschi (SFZ) have presented an excellent analysis that combines limited computer simulations with historical observations to provide inference about key climate system parameters. The Bayesian formulation used in this analysis allows the incorporation of a number of different sources of uncertainty in the final inference. We appreciate such a collaboration requires a substantial amount of effort from all involved. There is not much to criticize here. Most of our comments, which are motivated by the case study, apply generally to analyses involving the calibration of a physics-based simulation model using physical observations.

Detailed versus reduced simulation models

This analysis purposely uses a fairly aggregated model that assumes the physical system is constant along a given latitude band. Contrast this with a very detailed global circulation model which takes months to run. Our experience is that such reduced dimension computer models are very common in a variety of physics applications and offer a number of important advantages:

- they often lead to more direct and accurate inference about aggregate entities such as trends or inventories;
- their speed makes them amenable to analyses (such as this one) that require ensembles of simulation runs be carried out;
- they can often incorporate more realistic physics models that larger-scale simulation models cannot afford;
- they are often applicable to a wider variety of data types or sources.

Relative pros and cons regarding large scale global circulation models are considered in Shackley et al. (1998).

Not only is the simulation model reduced, but so too are the historic data records. The data are temporally and spatially aggregated, as well as differenced. This choice of how to use the data is important. One would like to keep the features in the data that are important and that the simulator can accurately model, while ignoring features of the data that are either uninformative or cannot be well modeled. In particular, the use of differencing in this analysis makes the results insensitive to constant shifts in the simulation output – it is simulating the trends in the historic data that is important.

*Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM, http://www.stat.lanl.gov/staff/DHigdon/index.html

†National Oceanographic Centre, University of Southampton, Southampton, UK

 ba0009

A shortcoming of such models is their perceived (real or imagined) distance from the actual physical system of interest. How are policy makers going to treat results from an analysis based on an earth whose climate is constant along latitude bands? We certainly feel that reduced models can play a very important role in guiding decisions and policy. How can we help bridge this gap?

Connecting simulations with reality

The authors assume δ – the term for model discrepancy – is zero, so that the difference between the model $\eta(\theta)$ and the observations is controlled by the error term ξ. This term is modeled as independent $N(0, \Sigma_i)$ for the three data sources: upper air temperature changes Z_1; surface temperature change Z_2; and deep ocean temperature trend Z_3. Hence, whatever model discrepancy there is in this application is absorbed by these error terms.

Our experience at LANL has been that model inadequacies can greatly affect the resulting posterior distribution for the model parameters θ. Hence it is important to account for such a possibility in the analysis. When allowing for such a term, it is also very helpful to inform the analysis as much as possible about the error in the observed data. Our feeling is that it is important to include important terms in the model (such as discrepancy), even if we don't feel there is sufficient information in the data or priors to remove potential degeneracies. Without such terms, there is the danger of obtaining unrealistically precise posteriors for the model parameters.

As a crude attempt to explore the consistency of the climate simulation model here we carried out a set of analyses using a similar modeling framework for combining simulations and physical observations which is described in Higdon et al. (2008). The resulting posterior distribution for the model parameters is shown in Figure 2. While the posteriors appear similar to those shown in Figure 6 of SFZ, these model formulations have a number of differences which make direct comparison difficult without additional work. The point here is to show how modifying this new model formulation can lead to qualitatively different inferences.

This model, like that of SFZ, assumes a common set of parameters is appropriate for modeling the three different data sources (Figure 1). If this is true, then separate calibrations to each of the data sources should be consistent. The posteriors from three separate analyses, one conditioning on Z_1, one on Z_2, and one on Z_3 are shown in Figure 3. One could combine these data sources by using a hierarchical model (Figure 1, right side), rather than a common model. Our hierarchical model links the common parameters from each data source θ_{ij} (i indexes data source, j indexes parameter) with independent normal distributions so that

$$\theta_{ij} \sim N(\theta_{0j}, \sigma_j^2)$$

where the parameters θ_{0j} and σ_j^2 are estimated in the analysis. Figure 4 shows the resulting posterior for θ_0. Because the separate calibrations are not entirely consistent, there is substantial spread in the posterior distribution for θ_0.

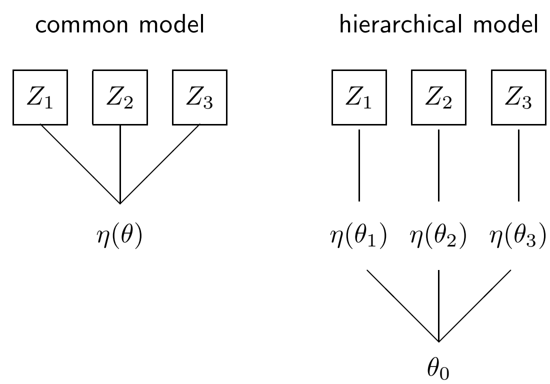

Figure 1: An alternative to constraining the model parameters θ under a common model formulation (left) used in this article is a hierarchical formulation (right). This hierarchical formulation allows a separate calibration of the model parameters for each of the three historical records – upper air temperature changes Z_1, surface temperature change Z_2 and deep ocean temperature trend Z_3 – and links these parameters with a hierarchical model. The resulting posteriors for the model parameters are given in Figures 2 – 4.

We are not making the claim that this wider posterior resulting from the hierarchical formulation is more appropriate for this application, rather we are pointing out that alternative, plausible formulations can substantially impact the resulting inference. The choice between various formulations is partly one statistics can answer, but it must also be informed by the science of the application. The form and magnitude of observation errors and discrepancies are important to consider here. See Goldstein and Rougier (2007) and the accompanying discussion for other perspectives on the topic of calibration and making the connection between computer models and reality.

In conclusion, we thank the authors for an interesting and well thought out analysis. We feel that the general field of simulation-aided inference will play an ever more prominent role in statistics, as well as in a broad range of application areas in the physical and engineering sciences. This paper makes an excellent contribution to the field.

References

Goldstein, M. and Rougier, J. C. (2007). "Reified Bayesian Modelling and Inference for Physical Systems (with discussion)." *Journal of Statistical Planning and Inference*.

Higdon, D., Gattiker, J. R., and Williams, B. J. (2008). "Computer Model Calibration using High Dimensional Output." *Journal of the American Statistical Association*.

Shackley, S., Young, P., Parkinson, S., and Wynne, B. (1998). "Uncertainty, complexity

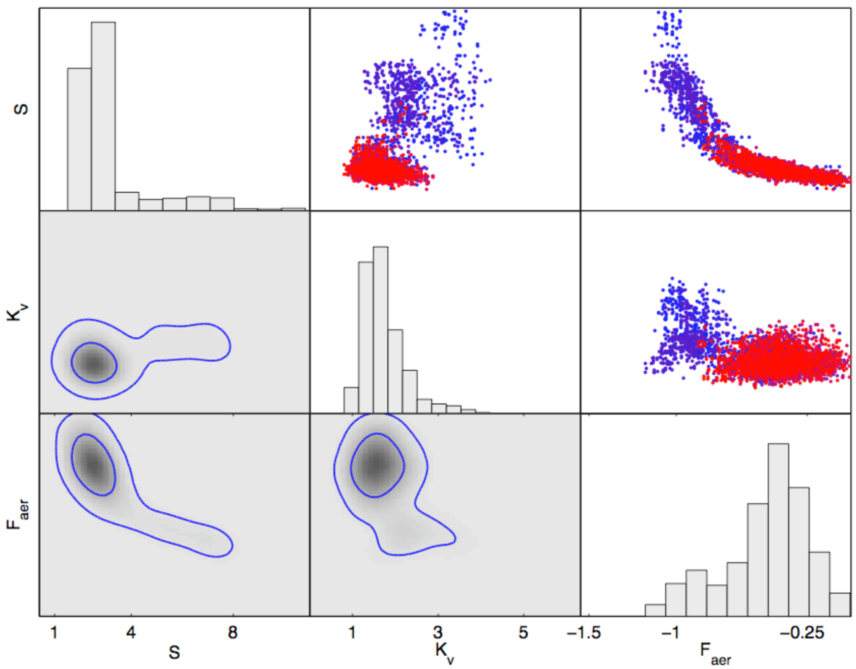

Figure 2: Posterior distribution for the calibration parameters using a common model formulation similar to the one used in SFZ.

and concepts of good science in climate change modelling: Are GCMs the best tools?" *Climatic Change*, 38: 159–205.

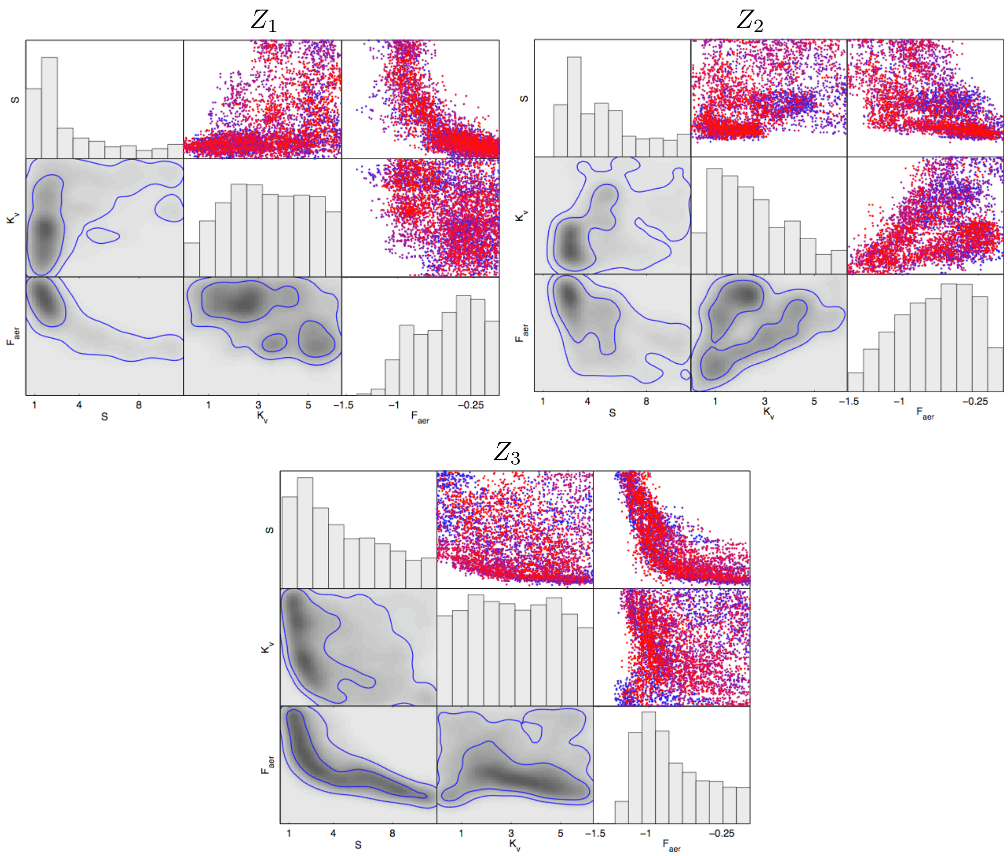

Figure 3: Posterior distributions obtained from fitting separate models to each of the three data sources – upper air temperature changes Z_1, surface temperature change Z_2 and deep ocean temperature trend Z_3.

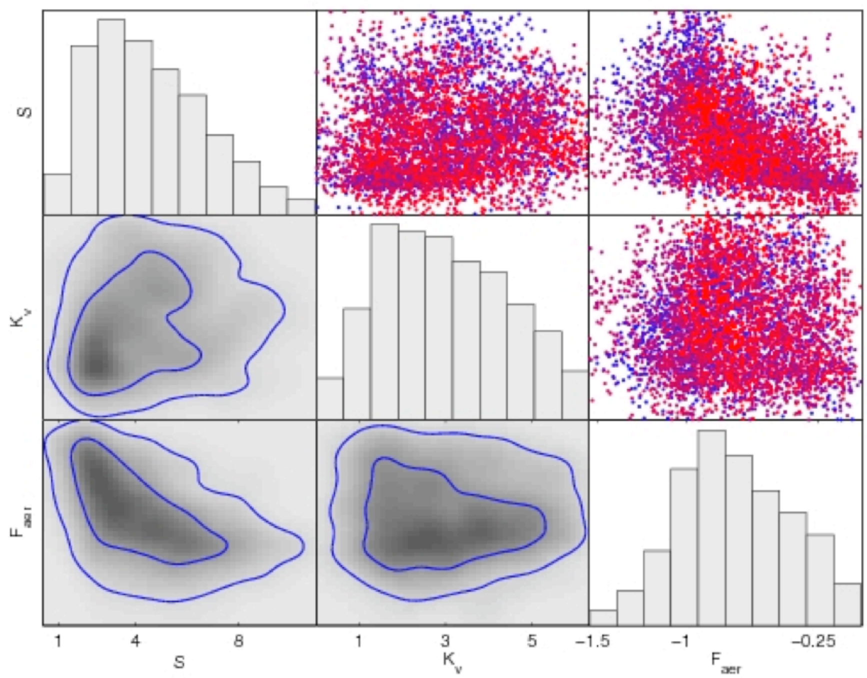

Figure 4: Posterior distribution for the common hierarchical calibration parameters θ_0 from the hierarchical model formulation. The resulting parameter uncertainty is much larger than it is under the common model formulation, which points to the possibility that the data are less informative than the original SFZ analysis suggests.