Inferring Climate System Properties Using a Computer Model

by

Bruno Sansó, Chris Forest and Daniel Zantedeschi

Four key uncertainties of future climate include: (i) Climate system response, as measured by the temperature change under a specified external perturbation; (ii) Climate variations due to natural and anthropogenic aerosols; (iii) Magnitude and character of natural climate variability; (iv) Spatio-temporal patterns of change in climate variables. An honest assessment of such uncertain climate properties is key to the support of any scientific statement about the current state of Earth's climate, and to the construction of forward-looking projections that may be used in policy decisions. In order to quantify climate system properties, computer models are run under different parameterizations. The present work is an attempt to provide a fully probabilistic quantification of the uncertainties in the climate system properties. It is the result of a multi-year collaboration between climate scientists at MIT and statisticians at UCSC.

In this work, we consider the MIT 2D climate model (Forest et al., 2006) (MIT2DCM). Such a model provides simulations of ocean, surface and upper atmospheric temperature behavior. The model uses a system of latitude and height coordinates to simulate the average state of the climate over zones defined by latitude bands. Despite the averaging over longitude, the model is sufficiently complex to match climate observations and to make similar predictions as those of full 3D atmosphere-ocean general circulation models (GCM). In the MIT2DCM, the climate system properties control the climate system response via three parameters: Climate sensitivity, denoted as $\mathcal{S}$, defined as the equilibrium global-mean surface temperature response to a doubling of CO_2; Diffusion of deep-ocean temperature anomalies, $\mathcal{K}_v$; Net anthropogenic aerosol and unmodeled forcings, $\mathcal{F}_{aer}$. Let $\boldsymbol{\theta} = (\mathcal{K}_v, \mathcal{S}, \mathcal{F}_{aer})$. In our approach, we estimate the posterior distribution of $\boldsymbol{\theta}$ using historical records and simulations from the MIT2DCM and GCMs.

Typical output from a run of the MIT2DCM consists of temperatures at 46 different latitudes, with 11 vertical layers for the 1860-1995 period, every 30 minutes. Such output is summarized in three "diagnostics": A vector of 288 components, consisting of the upper air temperature changes between the 1986-1995 and 1961-1980 periods at 36 latitudes and 8 levels; Surface temperature change, consisting of the difference between the decadal average temperatures for 1946-1995 periods and the average temperature of 1906-1995 at 4 different equal-area zonal bands, resulting in a 20 dimensional vector; Deep ocean temperature trend, calculated for the 1952-1995 period. Historical observations and GCM output are obtained in correspondence to the three diagnostics.

The MIT2DCM is run by letting $\boldsymbol{\theta}$ vary on a discrete grid. Prior information on climate properties is available from the literature on climate models. For a given diagnostic, denote $\boldsymbol{z}$ the observation and $\boldsymbol{\eta}$ the output of the model. Since $\boldsymbol{\eta}$ depends on $\boldsymbol{\theta}$, we can obtain a likelihood function assuming that the difference is Gaussian. The resulting posterior provides information for inference on the climate properties. Unfortunately, output from the MIT2DCM is available only at a few hundreds pre-specified points on an irregular grid. Even though the MIT2DCM is much faster than a GCM, the running and post-processing times needed to obtain a value for one of the diagnostics prevents us from embedding its evaluation within an iterative method. So, to fully explore the posterior distribution of $\boldsymbol{\theta}$ we create an auxiliary statistical model that provides an approximation to $\boldsymbol{\eta}$. For this purpose we use a Gaussian process. This is justified, from a Bayesian viewpoint, by the fact that for a given $\boldsymbol{\theta}$ the value of $\boldsymbol{\eta}$ is unknown, so we may consider it as a random process. In essence the setting of our problem is that of calibration of computer model parameters as described in Kennedy and O'Hagan (2001). The focus is not on prediction or data

assimilation, but on inference for the parameters that control the computer model which have a precise physical meaning. For two of the diagnostics the output is multivariate. So the calibration procedure has to incorporate information about a covariance matrix. We use GCM output to elicit a prior distribution for such matrices.

For a technical description of the model, let $\boldsymbol{t}_j \in \mathbb{R}^3, j = 1, \ldots, p$ be the available parameter configurations. $p = 499$. Let $\boldsymbol{x}_i \in \mathbb{R}^2, i = 1, \ldots, n$ be the locations of the diagnostics' components. $n = 1, 20$ or 288. Then $y_{ij} = \eta(\boldsymbol{x}_i, \boldsymbol{t}_j)$ denote the model runs and $\eta(x_i, \boldsymbol{\theta})$ denotes the value of the MIT2DCM at the 'true' value of the parameter. We assume that $z_i = \eta(x_i, \boldsymbol{\theta}) + \epsilon_i$. Here ϵ_i encompasses observational errors and model inaccuracies and biases. We let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)' \sim N(0, \sigma^2 \Sigma)$. In the climate literature, Σ is usually referred to as the unforced variability. Estimating Σ will be a byproduct of our model, but it is a problem that has an interest of its own. We then assume that $E(\eta(\boldsymbol{x}, \boldsymbol{t})) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{t})' \boldsymbol{\beta}$, where $\boldsymbol{h}$ is a polynomial in $\boldsymbol{x}$ and $\boldsymbol{t}$ and $\boldsymbol{\beta}$ a vector of coefficients of dimension q. Additionally we have that $\mathrm{cov}(\eta(\boldsymbol{x}_i, \boldsymbol{t}), \eta(\boldsymbol{x}_j, \boldsymbol{t}')) = r(\boldsymbol{t}, \boldsymbol{t}')\gamma^2 \Sigma_{ij}$. Here $r(\cdot, \cdot)$ is a correlation function. We assume that the MIT2DCM can reproduce a correlation structure compatible with the unforced variability. We obtain prior information about Σ by considering output from a GCM and assuming that it also captures the unforced variability.

The proposed model assumes a separable structure for the covariance of $\boldsymbol{\eta}$. The resulting covariance matrix can be written as a Kronecker product of two smaller matrices. This is a key modeling issue given our use of MCMC for the exploration of the posterior distribution. In fact, for the surface diagnostic the full covariance matrix is of size $(20 \cdot 500) \times (20 \cdot 500)$, for the upper air its size is $(288 \cdot 500) \times (288 \cdot 500)$, making computations, and even storage, very difficult for a non structured matrix.

Our model tackles several interesting issues for both the climatologist and the statistician. We study the properties of the climate using three parameters that are key for long-term climate projections. Such parameters provide information on how the climate will be affected by an increase in the amount of human produced emissions and the capacity of the oceans to absorb heat. We provide a comprehensive answer that incorporates scientific information, structural assumptions and different sources of data, observed and synthetic. The uncertainty is quantified using a joint probability distribution that yields a full description of the dependencies between the three parameters. The model tackles a computer calibration problem where the output is of moderate multivariate dimension. It accounts for all parameter estimation uncertainty. It provides an estimate of the unforced variability. It uses the information from different summaries of the Earth's climate that have different physical responses to the accumulation of heat.

References

Forest, C. E., P. H. Stone and A. P. Sokolov, (2006) Estimated PDFs of climate system properties including natural and anthropogenic forcings, *Geophys. Res. Let.*, **33**, L01705, doi:10.1029/2005GL023977.

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