

# Bayesian calibration of the Thermosphere-Ionosphere Electrodynamics General Circulation Model (TIE-GCM)

S. Guillas<sup>1</sup>, J. Rougier<sup>2</sup>, A. Maute<sup>3</sup>, A. D. Richmond<sup>3</sup>, and C. D. Linkletter<sup>4</sup>

<sup>1</sup>Department of Statistical Science & Aon Benfield UCL Hazard Research Centre, University College London, London, UK

<sup>2</sup>Department of Mathematics, University of Bristol, Bristol, UK

<sup>3</sup>High Altitude Observatory, National Center for Atmospheric Research, Boulder CO, USA

<sup>4</sup>Center for Statistical Sciences, Brown University, Providence, RI, USA

Received: 28 April 2009 – Published in Geosci. Model Dev. Discuss.: 12 May 2009

Revised: 1 September 2009 – Accepted: 4 September 2009 – Published: 28 September 2009

**Abstract.** In this paper, we demonstrate a procedure for calibrating a complex computer simulation model having uncertain inputs and internal parameters, with application to the NCAR Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIE-GCM). We compare simulated magnetic perturbations with observations at two ground locations for various combinations of calibration parameters. These calibration parameters are: the amplitude of the semidiurnal tidal perturbation in the height of a constant-pressure surface at the TIE-GCM lower boundary, the local time at which this maximises and the minimum night-time electron density. A fully Bayesian approach, that describes correlations in time and in the calibration input space is implemented. A Markov Chain Monte Carlo (MCMC) approach leads to potential optimal values for the amplitude and phase (within the limitations of the selected data and calibration parameters) but not for the minimum night-time electron density. The procedure can be extended to include additional data types and calibration parameters.

## 1 Introduction

The calibration of complex computer models, or simulators, of physical systems is a difficult endeavor, see Kennedy and O’Hagan (2001) and discussion therein. It consists of searching for the best combination of parameters in the simulator inputs which will produce outputs that match best the observations. There are modelling issues and heavy computational challenges. In many scientific areas, the common approach is to use parameter values that have been set through empirical

evidence or measurements of proxies. Instead, we explore here the sensitivity of the model to parameter changes in order to learn about our physical assumptions and numerical procedures. As a result of calibration through the comparison with observations, histograms of empirical posterior distributions of the parameters enable us to make a probabilistically informed choice of parameter values. To our knowledge, there are no calibration studies of simulators of the Earth’s ionosphere. Previous studies have used data to improve ionospheric model outputs, through data assimilation at regular time steps (Scherliess et al., 2006), but not to determine ionospheric model parameters. The reasons for which we want to calibrate such a simulator are: to replace tuning and fudge factors, to obtain more reliable simulations as we use more observations (ground- or space-based) under various conditions at different locations, seasons, local times, and to account for uncertainty when the model is used for system predictions. This uncertainty is a consequence of the model-parameters being underdetermined by the available observations, taking account of model limitations. Calibration also helps with code verification, in the sense that the posterior distributions ought to be physically intuitive, and if they are not then perhaps something has gone wrong en route. Given the potential gain in precision obtained through calibration of some key parameters, it is a very important step towards improving simulators. Furthermore, identifying the sensitivity to the parameters may help the modelers focus their research efforts on some selected physical phenomena. In Sect. 2 we present the simulator TIE-GCM, then in Sect. 3 we describe the observations and TIE-GCM outputs in our study. Section 4 is devoted to the Bayesian methodology and Sect. 5 to the analysis of our results. Finally in Sect. 6, we discuss potential improvements to our approach.



Correspondence to: S. Guillas  
(serge@stats.ucl.ac.uk)

## 2 The computer model TIE-GCM

The TIE-GCM simulator (Richmond et al., 1992) is designed to calculate the coupled dynamics, chemistry, energetics, and electrodynamics of the global thermosphere-ionosphere system between about 97 km and 500 km altitude. It has many input parameters to be specified at the lower and upper boundaries, as well as a number of uncertain internal parameters. There are also many output quantities from the TIE-GCM simulator (densities, winds, airglow emissions, geomagnetic perturbations, etc.) that can be compared with observations. For this study we explore the response of the magnetic-eastward ( $D$ ) magnetic perturbation at the ground, MAGGRD-D in [nT] (nano Tesla), at two locations to variations in just three inputs: two that help describe atmospheric tides at the TIE-GCM lower boundary, and one that constrains the minimum night-time electron density. MAGGRD-D varies from hour to hour during the day, but also with season, solar cycle and location of the observation. It is caused by electric currents flowing in the ionosphere, primarily on the day side of the Earth where solar extreme-ultraviolet radiation partially ionizes the upper atmosphere, rendering it electrically conducting. Winds move the conducting medium through the Earth's magnetic field, generating electric fields and currents by an electrodynamic effect in the so-called dynamo region, at heights of approximately 90–200 km. Observations of MAGGRD-D also indicate a considerable amount of day-to-day variability not captured by the TIE-GCM when driven by inputs that remain the same from one day to the next. The observations were therefore averaged over several days of quiet geomagnetic conditions.

Atmospheric tides are global waves with periods that are harmonics of 24 h. They comprise a major portion of the winds in the dynamo region. They are generated at lower atmospheric levels, and they are modulated by variable background winds as they propagate to the upper atmosphere. They are difficult to define since observations are limited and the tides vary not only with geographic location, local time and season, but also in a somewhat irregular manner from one day to the next. Modelling the tidal propagation through the atmosphere, and accurately determining their distribution at the TIE-GCM lower boundary, remains a challenge. For this study, we include fixed diurnal (24 h period) and semidiurnal (12 h period) migrating (Sun-synchronous) tidal components at the TIE-GCM lower boundary, taken from the physical model of Hagan and Forbes (2002, 2003), plus an additional variable tidal forcing (migrating (2,2) mode) which is known to be important for the electrodynamics (Fesen et al., 2000). The amplitude of the perturbation in the height of a constant-pressure surface at the TIE-GCM lower boundary,  $AMP \in [0, 36\,000]$  cm, and the local time at which this maximises,  $PHZ \in [0, 12]$  h, are two of the three inputs we explore.

At night, the ionospheric electron density below 200 km is small and difficult to measure, but nonetheless has an important influence on the night-time electric field. Our third simulator input is the minimum night-time electron number density in  $\text{cm}^{-3}$ ,  $EDN \in [1000, 10\,000]$ . All other input parameters in the TIE-GCM simulator are held constant for our experiments. The simulations are done for equinox, at low solar and geomagnetic activity. For each evaluation, the TIE-GCM simulator is initially spun-up to get a diurnally reproducible state.

## 3 Observations and computer runs

Over one hundred magnetometers around the globe provide geomagnetic variation data. The TIE-GCM can simulate the magnetic perturbations for any site. Here we analyse the sites marginally, disregarding shared information that might be available from sites that are proximate, by using data from only one site at a time. Therefore the simulator output for each evaluation comprises points on a periodic function of time for some pre-specified site. We concentrate here on the MAGGRD-D at two locations: Apia (API, 13.81° S, 171.77° E) and Odessa (ODE, 46.78° N, 30.88° E.)

Note that PHZ is a periodic input, so that  $f(AMP, 0, EDN) = f(AMP, 12, EDN)$  for all AMP and EDN. We consider here an alternative parameterisation of AMP and PHZ,  $\theta_1 = AMP \cos(\pi PHZ/6)$  and  $\theta_2 = AMP \sin(\pi PHZ/6)$ , which accommodates the periodicity. The EDN input is not reparameterized and is denoted  $\theta_3$ . We would expect that there is a strong correlation between the tidal input at the lower boundary (amplitude and phase) and the magnetic perturbation during the day at low and mid latitudes.

In our initial comparisons of model predictions with observations it was found that the TIE-GCM underpredicted the amplitudes of MAGGRD-D, owing to E-region electron densities that were too low. This resulted in low conductivities, therefore low current, and low magnetic perturbations, since these are the results of the current flowing overhead. Fang et al. (2008) considered the need to increase the ionospheric electron density in order to get the TIE-GCM to agree with electron-density observations, and in order to get magnetic perturbations compatible with observations. In their case, they noted that the TIE-GCM electron density at the peak of the ionospheric E-layer, around 110 km altitude, was about 37% too small, meaning it needed to be increased by a factor of 1.58. To quickly fix this before calibration, we multiply the TIE-GCM outputs by an empirical factor of 1.4. (This adjustment was made before the final results of Fang et al. (2008) were available.) This adjustment is sufficient for demonstrating the capabilities of our method. Versions of the TIE-GCM currently under development are expected to eliminate the need for such adjustments in the future.

In order to increase the influence of small-amplitude MAGGRD-D values relative to large-amplitude values in the calibration, we transformed the observations and simulator outputs as follows:  $y \rightarrow \text{sgn}(y) \log(1 + |y|)$ . The transformation also enables us to satisfy the assumption that the variability is constant in time so that the methodology in the next section can be applied. The transformed measurements at the two locations API and ODE have different features, see Fig. 2.

#### 4 Bayesian methodology for calibration

For the calibration of TIE-GCM, we follow here a Bayesian approach (Kennedy and O'Hagan, 2001). It consists of putting distributional assumptions (prior distributions or simply priors) on the calibration (also called tuning) parameters  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  before comparing with observations and letting the information contained in the data update this a priori assumption to get as a result a posterior distribution of the calibration parameters. The advantage of such a Bayesian analysis over standard estimation of parameters (e.g. by minimizing the differences between observations and simulator outputs) lies mainly in the ability to retrieve a full description of the uncertainties about the parameters and consequently about the simulator outputs. Moreover, the possibility for the modelers to express their – uncertain – scientific beliefs in terms of priors on the parameters enables a natural integration of scientific knowledge and evidence given by measurements. Since magnetic variations in the two locations API and ODE are different, independent calibrations that would give consistent results for each of these locations may be deemed reliable.

The complete set of inputs  $x=(t, \theta)$  consists of parameters divided into two categories: the known parameters (controllable parameter time  $t$  in  $[0, 24)$ ) and the unknown calibration parameters  $\theta = (\theta_1, \theta_2, \theta_3)$ . We denote by  $\eta(x)$  the output of the computer model which depends on the complete set of parameters  $x=(t, \theta)$ . The computer code output  $\eta(x)$  is an approximation of the reality  $y^R(t)$ . The notation used emphasizes that physical observations are only made at values of the observable parameter,  $t$ . To learn about the values of the calibration parameters, TIE-GCM is run at inputs  $x$  in a design (i.e. choice of values)  $D^M$ . Field data (i.e. observations)  $y^F(t)$  are collected at a number of inputs  $t$  in a design  $D^F$ .

The design  $D^F$  (i.e. only the time points of observations) is given by the 24 hourly observation times: around 00:14, ..., 23:14 magnetic local time (MLT) for API and around 00:43, ..., 23:43 MLT for ODE. (MLT is defined as the magnetic longitude difference between the point in question and the anti-solar point on the Earth, multiplied by 24 h/360°. Magnetic longitude is referenced to the geometry of the geomagnetic field instead of to geographic coordinates.) Local time is different for the two locations. How-

ever, since the study is done for each location separately, it does not matter as we focus on covariances. The part of our design of experiments  $D^M$  corresponding to the calibration parameters is a maximin Latin Hypercube Design (Williams et al., 2000). With this design we try to cover as much space as possible in the three-dimensional space of the calibration parameters with only  $n=30$  runs. Two-dimensional projections of this design are shown in Fig. 1. This is not a perfect design, but seems satisfactory for our study. For the time component of the computer design  $D^M$  we choose 12 points (every other hour) to maximize the amount of information obtained through these time points under the constraint of the computing time necessary to perform the Bayesian calibration, see Fig. 2. The time points in  $D^F$  and  $D^M$  are different, but the methodology accommodates such variation. Note that time is an input parameter, but a so-called controllable one, which is included in the design but on which we do not do inference.

The following equations describe the bias between the computer simulator and the physical observations at the time design points, denoted  $\delta(t)$ , and observation error  $\epsilon(t)$  (Kennedy and O'Hagan, 2001):

$$y^R(t) = \eta(t, \theta^*) + \delta(t) \quad (1)$$

$$y^F(t) = y^R(t) + \epsilon(t) \quad (2)$$

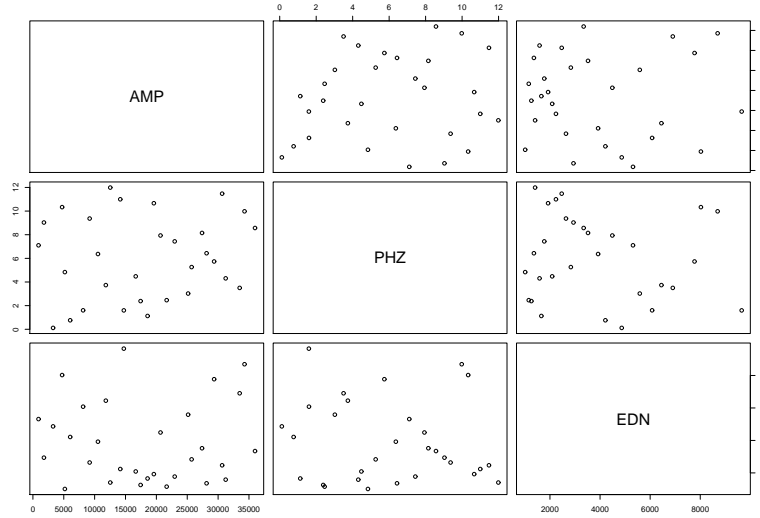
Here,  $\theta^*$  is used to represent the true (unknown) values of the calibration parameters. These equations suggest that even if the computer simulator was run at the true values of the calibration parameters, it would still be a biased representation of reality. Note that we do not include a regression parameter that generalizes further the analysis by multiplying the computer outputs by a constant (Kennedy and O'Hagan, 2001). We took into account this scale issue as explained in Sect. 3. Hence we effectively removed an additional statistical parameter from the Bayesian analysis and saved computer time since this might have led to more identification problems and longer convergence.

Because the simulator output  $\eta(\cdot)$  is unknown except at the design points  $D^M$ , we assume that the unknown function follows a Gaussian stochastic process (GASP) distribution. That is, we model the  $12n$  observed simulator responses  $\eta(x)$ ,  $x \in \mathbb{R}^p$  (here  $p=4$  since  $D^M$  is over a range of  $t$ ,  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  values), as coming from a multivariate normal distribution with a constant mean function  $\mu$  and an  $12n \times 12n$  variance-covariance function  $\Sigma$ , with density:

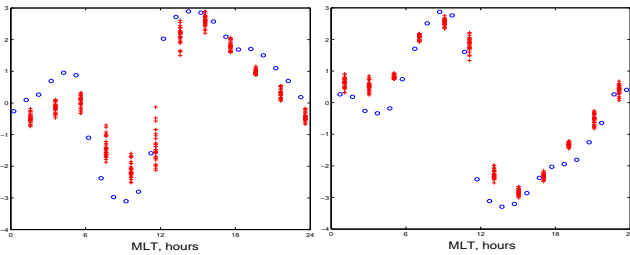
$$f(\eta(x)) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}. \quad (3)$$

Thus, we approximate the computer simulator by specifying a distribution of functions that interpolate the response  $\eta(x)$  in between the design points  $x$  in  $D^M$ . The random function is certain at the design points, and uncertain at untried points.

After inspection of the transformed outputs, it appears that the normal assumption is reasonable. To specify  $\Sigma$  according to the calibration parameters we use a product Gaussian



**Fig. 1.** Two-dimensional projections of the design of experiments for the initial calibration parameters of TIE-GCM. The 30 combinations are based on a maximin Latin Hypercube Design approach. Units: cm for AMP, hours for PHZ, and  $\text{cm}^{-3}$  for EDN.



**Fig. 2.** Measurements (circles) and computer outputs (crosses) for API (left panel) and ODE (right panel). Transformation of original data in [nT] (nano Tesla), see text for details.

variance-covariance. For the time dimension, we allowed for a periodic correlation structure by representing it in terms of an angle, so that the values at the end of the day are correlated with the values at the beginning of the day. After we rescale the time onto the interval  $[0, 1]$ , we choose a valid (i.e. positive-definite) isotropic correlation function on the circle  $[0, 1]$  (Gneiting, 1999). Thus, the  $(i, j)$ -th element of  $\Sigma$  is

$$\begin{aligned} \Sigma_{i,j} &= \text{cov}(\eta(x_i), \eta(x_j)) \\ &= \frac{1}{\lambda_\eta} \exp\left(-\sum_{k=1}^3 \beta_k |\theta_{ik} - \theta_{jk}|^2\right) \\ &\quad \times \exp\left(-\beta_4 (\sin(\angle(t_i, t_j)/2))^2\right) \end{aligned} \quad (4)$$

The notation  $\theta_{ik}$  denotes the  $i$ -th design point in  $D^M$  for  $\theta_k$ , and  $\angle(t_i, t_j)$  is the angle between  $t_i$  and  $t_j$  (i.e. minimum distance between  $t_i$  and  $t_j$  on the circle  $[0, 1]$  rescaled to  $[0, 2\pi]$ ). The hyperparameters  $\mu$ ,  $\lambda_\eta$  (the precision of the GASP model),  $\beta_k$  (which we call ‘‘correlation hyperparameters’’) are to be estimated from the model output and the observations as described below.

The unknown bias function  $\delta(t)$  is also modeled as a GASP random function with mean 0 and periodic correlation matrix with precision  $\lambda_\delta$  and correlation parameter  $\beta_5$ . Finally, the random error component is modeled as independent  $\epsilon(t) \sim N(0, 1/\lambda_\epsilon)$ . For estimation of the calibration and hyperparameters, we make use of the Markov Chain Monte Carlo (MCMC) approach (Gilks et al., 1996). The chains are dependent random samples that ought to be distributed in the long run as the so-called posterior distributions of the parameters of interest, which is a combination of prior uncertainty about the values of these parameters and the information about the parameters provided by the data. Of particular interest, we retrieve the posterior distributions of the various calibration parameters, which allows us to make inferences and quantify our uncertainty about the true values of these unknown quantities.

For ease of implementation of the MCMC algorithm, we initially standardize the entire set of responses (simulator and observed) by the mean and standard deviation of the simulator responses, so  $\mu$  can be assumed to be 0 without loss of generality and the variability in the simulator ( $1/\lambda_\eta$ ) is approximately 1. The design space on the calibration parameters is also scaled to be  $[0, 1]^3$ , and the time dimension of the design space is scaled to a circle  $[0, 1]$  as we assume periodicity.

All the unknowns in the model (i.e. the calibration parameters and the hyperparameters) require specified prior distributions which represent uncertainty about the values of these parameters before any data is collected. The following choices are made for the priors:

- To represent vague prior information about the true calibration parameter values, we specify a uniform prior

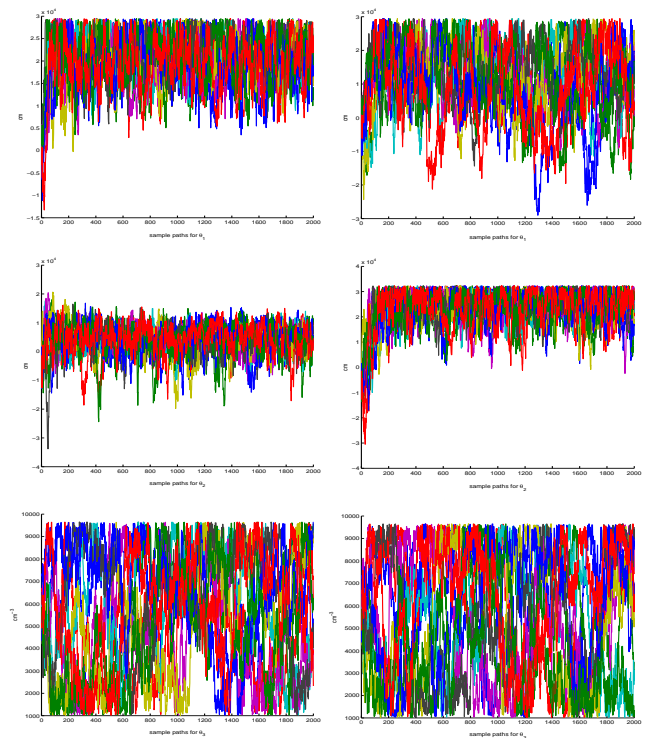
distribution over an interval twice as wide as the interval on which they were sampled for simulator runs.

- To model the correlation hyperparameters in  $\Sigma$ , we reparameterize using  $\rho_k = \exp(-\beta_k/4)$ . Because  $\beta_k > 0$ , this yields  $0 < \rho_k < 1$ . Thus, for  $\rho_k$ , a *Beta*(1, .5) distribution is used, which conservatively places most of its prior mass on values of  $\rho_k$  near 1 (indicating an insignificant effect). Similarly, even more conservative *Beta*(1, .4) priors were used for reparameterized correlation hyperparameters in the GASP model for the bias function.
- Gamma prior distributions were used for each of the precision (i.e. inverse of the variance) hyperparameters  $\lambda_\eta$ ,  $\lambda_\delta$  and  $\lambda_\epsilon$ . Specifically, we use priors  $\lambda_\eta \sim \text{GAM}(10, 10)$  (with expectation 1 due to standardization of the responses),  $\lambda_\delta \sim \text{GAM}(10, .3)$  (with expectation around 20% of standard deviation of the standardized responses), and  $\lambda_\epsilon \sim \text{GAM}(10, .03)$  (with expectation around 5% of standard deviation of the standardized responses).

Because our choice of priors make the full conditional distributions of the unknowns difficult to sample from in the MCMC chain, we implement a Metropolis-Hastings algorithm to explore the multidimensional space of parameters. This eventually yields draws from the posterior distribution by repeatedly accepting and rejecting a choice of move in the parameter space. We used multiple chains to confirm the convergence towards a stationary posterior distribution (after an initial burn-in period), saving wall-clock time by running the chains in parallel.

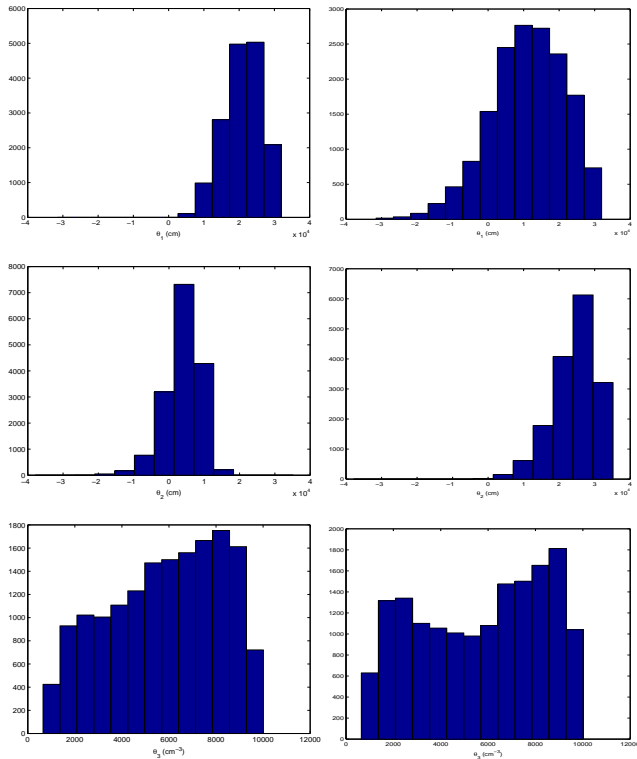
### 5 Results

Figure 3 shows the sample paths for 10 chains, with 2000 iterations, corresponding respectively to the calibration parameters  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ . From the visual inspection of these chains, it seems that for the parameters  $\theta_1$  and  $\theta_2$ , convergence occurs after roughly 400 iterations. These first 400 values will be dropped for the rest of the analysis as they are considered to be in the so-called burn-in period. Unfortunately, the convergence of the chains can not be established for the calibration parameter  $\theta_3$ , even by running the chains longer. Some parameters paths are cut off at certain values because Metropolis-Hastings algorithm rejects jumps beyond these values. It seems that these values correspond to the limits of the intervals we used in the design and there is little information there. Not surprisingly, our method is not able to tune this parameter. The minimum night-time electron density represented by EDN does affect ionospheric electrodynamics at night, including the ionospheric drift velocities, but is too small to increase the electrical conductivity to anywhere near the daytime values. The geomagnetic

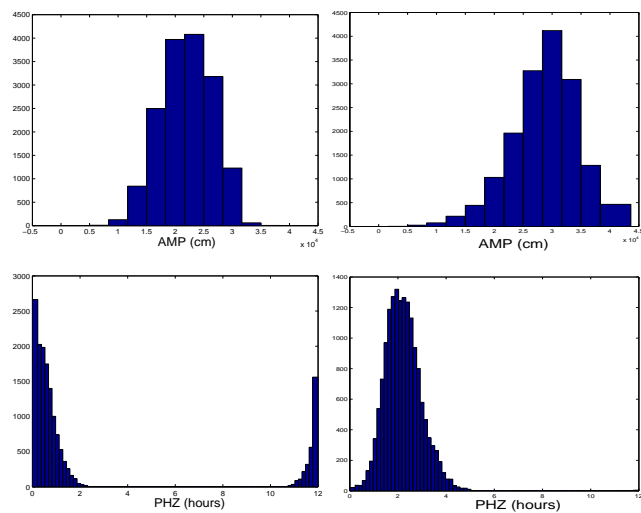


**Fig. 3.** Calibration for parameter  $\theta$ . MCMC sample paths for 10 chains. Left panels: API. Right panels: ODE.

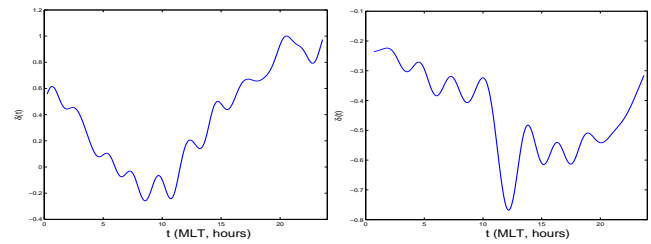
perturbations of MAGGRD-D are dominated by the much larger currents in the day-side ionosphere, and are very insensitive to EDN. Future studies that add ionospheric drift data to the observation set should be able to constrain EDN much better than the present study (Fesen et al., 2000). The inclusion of EDN here serves the purpose of testing how well the method works when insensitive parameters are included. The histograms of the empirical posterior densities are displayed in Fig. 4 for the calibration parameters  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ . The resulting histograms for the parameters AMP and PHZ are displayed in Fig. 5. Note that  $\theta_3$  is the parameter EDN and does not need to be transformed. The peaks of the histograms for AMP and PHZ for the two sites are reasonably consistent since they overlap, but show distinctive features. This discrepancy may be explained by the fact that these calibration parameters may compensate for other factors or parameters. We derive best values of approximately  $2 \times 10^4$  cm and 00:00 MLT respectively for AMP and PHZ at API. For ODE, the best values are respectively  $3 \times 10^4$  cm and 02:00 MLT. For these respective values, TIE-GCM outputs are closer to the observations at the two locations of interest than for other combinations of AMP and PHZ in the range of values we considered (though a bias is still present). Note that Fig. 2 shows that the observations are outside the evaluations. This might be because there is a large systematic discrepancy, but it might also be because the best choice for the parameters is in a region of the parameter space that our ensemble did not



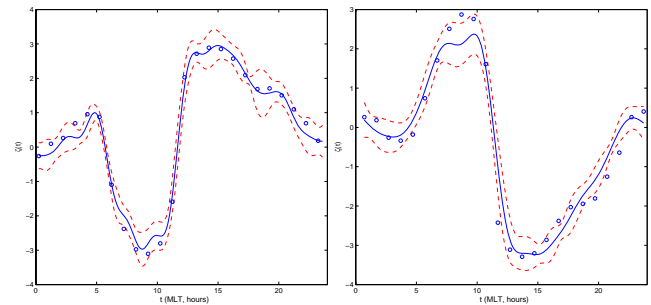
**Fig. 4.** Calibration for parameter  $\theta$ . Histograms of posterior distributions based on MCMC sample paths for 10 chains, with first 400 values dropped as they are considered to be in the burn-in period. Left panels: API. Right panels: ODE.



**Fig. 5.** Calibration for parameter AMP and PHZ at location API. Histograms of posterior distribution based on MCMC sample paths for 10 chains, with first 400 values dropped as they are considered to be in the burn-in period. Left panels: API. Right panels: ODE.



**Fig. 6.** Posterior mean model biases  $\delta$ . Left panel: API. Right panel: ODE. Posterior distributions are based on MCMC sample paths for 10 chains, with first 400 values dropped as they are considered to be in the burn-in period.



**Fig. 7.** Observed (circles) and predicted MAGDDR-D (lines) with associated 95% confidence bands ( $\pm 1.96$  times the standard errors assuming a Normal distribution). Left panel: API. Right panel: ODE. Posterior distributions are based on MCMC sample paths for 10 chains, with first 400 values dropped as they are considered to be in the burn-in period.

explore. Indeed, the number of point in the design for which AMP is in between  $2 \times 10^4$  and  $3 \times 10^4$  cm and PHZ is near 0 is indeed empty, see center upper panel in Fig. 1. Finally, there is no indication that EDN has a significant impact on the outputs through our analysis.

Mean posterior biases are shown in Fig. 6. These biases estimates are indeed compensating for some of the differences between calibrated model outputs and observations. The values are not small since they represent a variation of the order of one unit compared to variations of the order of 3 units in the transformed observations. The high-frequency variability displayed is due to the various time locations at which sources of information (observations and model outputs) are collected, and could be smoothed. We believe that the inclusion of more calibration parameters may help reduce these biases. The Bayesian calibration, through the propagation of uncertainties, also provides distributions of the posteriors for the predictions of the real values  $y^R$  at any time, conditional on the observed data. This statistical surrogate for the computer model is called an emulator. Figure 7 compares our emulator with observations at API and ODE. Our

emulator performs well, though it is only doing a prediction within sample. In our companion paper (Rougier et al., 2009) we considered a direct emulator of TIE-GCM that gives even better results, but in a different setting: the focus was on the upward  $\mathbf{E} \times \mathbf{B}$  drift at one location. This emulator is an Outer Product Emulator (Rougier, 2008) that utilizes expert knowledge by being tailored to the problem. In that framework, out-of-sample emulations are more reliable and much less computationally intensive.

## 6 Conclusions

Linkletter et al. (2006) and Welch et al. (1992) addressed the choice of calibration parameters. They identify the inputs that most impact the system so that these factors can be investigated further, dropping the others. We could use such a methodology in the calibration of TIE-GCM, when considering more than three calibration parameters. Furthermore, in the situation where the input space is large (for instance with many calibration parameters), a so-called sequential design (Williams et al., 2000; Kleijnen and van Beers, 2004) may help reduce the computational effort and focus on areas of interest in the input space. Our method readily accommodates larger calibration inputs, but the computing time will increase.

Since the outputs of TIE-GCM are effectively continuous (though discretized) quantities distributed in space and time, to carry out the calibration, we could have followed recent functional approaches (Bayarri et al., 2007; Higdon et al., 2008) by decomposing in wavelets bases or according to the first few principal components. We could have used periodic Fourier bases as we did for the direct emulation (Rougier, 2008), since they worked well there. However, since we chose a fully Bayesian method for which we did not want to impose too many constraints, and the dimension of the problem was reasonable for computational purposes, we did not resort to functional approaches here.

To improve further the calibration of TIE-GCM, we could consider more locations and more output types. We aim to obtain single estimates of parameters like AMP, PHZ, EDN, based on the combined data sets. However, including multiple sites requires us to parameterise the discrepancy function by location, to account for spatially systematic model biases; in this way we borrow strength across multiple locations, but we do not over-count proximate locations, because we appreciate that they share error. Furthermore, we chose only a single (geomagnetic) type of data, only two magnetometer locations out of over a hundred available, and only one ( $D$ ) component. For instance, Fesen et al. (2000) carried out a sensitivity analysis of TIE-GCM to EDN using vertical ion drifts; they concluded that EDN should not be below  $4000 \text{ cm}^{-3}$  to represent the important pre-reversal enhancement.

*Acknowledgements.* We thank the referees for their comments that improved the presentation of the paper. We thank Tomoko Matsuo for helpful comments on an earlier draft. The research described herein has been funded by the US National Science Foundation through a New Researcher Fellowship for S. Guillas at the Statistical and Applied Mathematical Sciences Institute (SAMSI) in 2006–2007. This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. The National Center for Atmospheric Research is also sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Edited by: J. Annan

## References

- Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., Sacks, J., and Walsh, D.: Computer model validation with functional output, *Ann. Stat.*, 35, 1874–1906, 2007.
- Fang, T., Richmond, A., Liu, J., Maute, A., Lin, C., Chen, C., and Harper, B.: Model simulation of the equatorial electrojet in the Peruvian and Philippine sectors, *J. Atmos. Sol.-Terr. Phys.*, 70, 2203–2211, 2008.
- Fesen, C. G., Crowley, G., Roble, R. G., Richmond, A. D., and Fejer, B. G.: Simulation of the pre-reversal enhancement in the low latitude vertical ion drifts, *Geophys. Res. Lett.*, 27(13), 1851–1854, 2000.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (Eds.): *Markov chain Monte Carlo in practice*, Interdisciplinary Statistics, Chapman & Hall, London, 1996.
- Gneiting, T.: Correlation functions for atmospheric data analysis, *Q. J. Roy. Meteor. Soc.*, 125, 2449–2464, 1999.
- Hagan, M. and Forbes, J.: Migrating and nonmigrating diurnal tides in the middle and upper atmosphere excited by tropospheric latent heat release, *J. Geophys. Res.*, 107(D24), 4754, doi:10.1029/2001JD001236, 2002.
- Hagan, M. and Forbes, J.: Migrating and nonmigrating semidiurnal tides in the middle and upper atmosphere excited by tropospheric latent heat release, *J. Geophys. Res.*, 108(A2), 1062, doi: 10.1029/2002JA009466, 2003.
- Higdon, D., Gattiker, J., Williams, B., and Rightley, M.: Computer model calibration using high-dimensional output, *J. Am. Stat. Assoc.*, 103, 570–583, 2008.
- Kennedy, M. C. and O’Hagan, A.: Bayesian calibration of computer models, *J. Roy. Stat. Soc. B Met.*, 63, 425–464, 2001.
- Kleijnen, J. and van Beers, W.: Application-driven sequential designs for simulation experiments: Kriging metamodelling, *J. Oper. Res. Soc.*, 55, 876–883, doi:10.1057/palgrave.jors.2601747, 2004.
- Linkletter, C., Bingham, D., Hengartner, N., Higdon, D., and Ye, K. Q.: Variable selection for Gaussian process models in computer experiments, *Technometrics*, 48, 478–490, 2006.
- Richmond, A., Ridley, E., and Roble, R.: A Thermosphere/Ionosphere general circulation model with coupled electrodynamics, *Geophys. Res. Lett.*, 19, 601–604, 1992.

- Rougier, J.: Efficient Emulators for Multivariate Deterministic Functions, *J. Comput. Graph. Stat.*, 17, 827–843, doi:10.1198/106186008X384032, 2008.
- Rougier, J., Guillas, S., Maute, A., and Richmond, A.: Expert Knowledge and Multivariate Emulation: The Thermosphere-Ionosphere Electrodynamics General Circulation Model (TIE-GCM), *Technometrics*, revised, 2009.
- Scherliess, L., Schunk, R. W., Sojka, J. J., Thompson, D. C., and Zhu, L.: Utah State University Global Assimilation of Ionospheric Measurements Gauss-Markov Kalman Filter model of the ionosphere: Model description and validation, *J. Geophys. Res.*, 111, A11315, doi:10.1029/2006JA011712, 2006.
- Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D.: Screening, predicting, and computer experiments, *Technometrics*, 34, 15–25, 1992.
- Williams, B. J., Santner, T. J., and Notz, W.: Sequential design of computer experiments to minimize integrated response functions, *Stat. Sinica*, 10, 1133–1152, 2000.