ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2015)
64, *Part* 1, *pp.* 93–113

Emulating and calibrating the multiple-fidelity Lyon–Fedder–Mobarry magnetosphere–ionosphere coupled computer model

Matthew J. Heaton,

Brigham Young University, Provo, USA

William Kleiber

University of Colorado, Boulder, USA

and Stephan R. Sain and Michael Wiltberger

National Center for Atmospheric Research, Boulder, USA

[Received April 2013. Final revision December 2013]

Summary. The Lyon–Fedder–Mobarry global magnetosphere–ionosphere coupled model LFM-MIX is used to study Sun–Earth interactions by simulating geomagnetic storms. This work focuses on relating the multifidelity output from LFM-MIX to field observations of ionospheric conductance. Given a set of input values and solar wind data, LFM-MIX numerically solves the magnetohydrodynamic equations and outputs a bivariate spatiotemporal field of ionospheric energy and flux. Of particular interest here are LFM-MIX input settings required to match corresponding output with field observations. To estimate these input settings, a multivariate spatiotemporal statistical LFM-MIX emulator is constructed. The statistical emulator leverages the multiple fidelities such that the less computationally demanding yet lower fidelity LFM-MIX is used to provide estimates of the higher fidelity output. The higher fidelity LFM-MIX output is then used for calibration by using additive and non-linear discrepancy functions.

Keywords: Dimension reduction; Multivariate computer model; Multivariate emulator; Non-additive discrepancy

1. Introduction

Coronal mass ejections are a highly energetic body of plasma and magnetic fields released from eruptions near complex magnetic field configurations on the Sun. When coronal mass ejections propagate out from the Sun and reach the Earth, the magnetosphere (the area of space that is controlled by Earth's magnetic field) prevents most of the particles from hitting the Earth. However, the interaction of the plasma with the Earth's magnetosphere can lead to significant disruption of the near Earth space environment. Such disruptions are referred to as geomagnetic storms. Severe storms can result in interferences to satellite operations and radio communications as well as impact the Federal Aviation Administration's 'Wide area augmentation system'. The electric currents driven in these storms can become so intense that they disrupt power grids, the prime example being the March 1989 'Quebec blackout' that was caused by two large

Address for correspondence: Matthew J. Heaton, Department of Statistics, Brigham Young University, 204 TMCB, Provo, UT 84602, USA.
E-mail: mheaton@stat.byu.edu

 0035–9254/15/64093

geomagnetic storms (see National Research Council Committee on the Societal and Economic Impacts of Severe Space Weather Events (2008) and Sparks *et al.* (2013)). As recently as spring 2012, massive coronal mass ejections dumped enough energy into the Earth's upper atmosphere to power every residence in New York City for 2 years (Phillips, 2012).

1.1. The LFM-MIX computer model

The Lyon–Fedder–Mobarry global magnetosphere–ionosphere coupled model LFM-MIX is a computer model that is used at the Center for Integrated Space Weather Modeling (CISWM) to study geomagnetic storms and is an important piece in an on-going effort to build a space weather forecasting system. Particularly, LFM-MIX is used to simulate interactions between the magnetosphere and the ionized portion of the Earth's upper atmosphere called the ionosphere (see Wiltberger *et al.* (2009) for a complete description of LFM-MIX). LFM-MIX uses solar wind observations to drive a magnetohydrodynamic simulation of the magnetosphere (the 'LFM' piece) and couples it with an electrostatic model of the ionosphere (the 'MIX' piece). The output is a bivariate spatiotemporal field of the ionospheric precipitation energy (in kiloelectron volts) and flux (per square centimetre second).

Beyond solar wind data, to run the magnetohydrodynamic simulation, LFM-MIX requires three spatially constant inputs (or initial conditions) given by

$$E_0 = x_1 \xi^2, \tag{1}$$

$$F_0 = x_2 \omega \sqrt{E_0}, \tag{2}$$

$$\mathrm{ED}_0 = \frac{x_3(\mathrm{FAC})\sqrt{E_0}}{\omega} \tag{3}$$

where E_0, F_0 and ED_0 represent initial precipitation energy, flux and potential energy difference of electrons flowing from the magnetosphere into the ionosphere respectively. CISWM scientists treat the values of ξ, ω and FAC as known constants because these are computed within LFM-MIX. The values for x_1, x_2 and x_3, in contrast, are unknown and are used to 'tune' LFM-MIX so that output roughly matches observations, i.e. scientists run LFM-MIX at various values of $\mathbf{x} = (x_1, x_2, x_3)$ and choose a value so that resulting output best mimics observations.

The LFM-MIX coupled model is a 'multiple-fidelity' computer model (see Qian and Wu (2008) and Gratiet (2013)), i.e. LFM-MIX can be run at different settings (which we choose to call 'fidelity levels') so that running the model at higher fidelity results in output that has more realistic spatiotemporal features. However, this added fidelity comes at the cost of computation—higher fidelity models require more computing time. For this study, we consider only two fidelity levels which we refer to as 'double' and 'single' fidelity where the double fidelity is preferred to the single fidelity. The choice of fidelity only impacts the modelling of the magnetosphere, ideally improving the modelling in the LFM component of the coupled LFM-MIX model.

1.2. Research objectives and challenges

To facilitate proper use of LFM-MIX to study Sun–Earth interactions, the primary goal of this analysis is to use field data to determine the 'best' value of $\mathbf{x} = (x_1, x_2, x_3)'$ such that the LFM-MIX output of ionospheric energy most accurately mimics observational data. Let $\boldsymbol{\chi} = (\chi_1, \chi_2, \chi_3)$ represent this best value for $\mathbf{x}$ where a Greek letter is used to emphasize the fact that this setting is unknown. Within the statistical computer modelling literature this problem is referred to as a computer model inverse problem (Tarantola, 2005) and is closely related to

'calibration' problems (see Kennedy and O'Hagan (2001) and Santner *et al.* (2003), chapter 2). Here, the approach is to develop a hierarchical model with two distinct layers:

(a) a layer that predicts LFM-MIX at untried values of $\mathbf{x}$ and
(b) a layer that links LFM-MIX output to observations at the calibrated value $\boldsymbol{\chi}$.

The first or 'emulator' layer of the model is used to predict LFM-MIX output at untried initial conditions (Rougier, 2008). Owing to the computational requirements of running LFM-MIX, the Gaussian process response surface methodology of Sacks *et al.* (1989) and Kennedy and O'Hagan (2001) is used to emulate computer model output. However, because the LFM-MIX output that we use here is 18 time points of a 46×36 spatial grid (a total of 29808 output points per run of LFM-MIX), the matrix inversion and determinant calculation that is required for a full Gaussian process implementation is computationally taxing. To deal with the issue of spatial dimensionality, we use Gaussian predictive processes (Banerjee *et al.*, 2008; Finley *et al.*, 2009) to obtain a low rank Gaussian process LFM-MIX model emulator. The choice to use predictive processes is primarily from the interpretability of the predictive process as a kriged spatial surface. Note, however, that the model that is outlined below could be easily adapted to use other low rank processes including fixed rank kriging (Cressie and Johannesson, 2008; Kang and Cressie, 2011) or discrete kernel convolutions (Higdon *et al.*, 1999; Higdon, 2002; Bhat *et al.*, 2012; Calder *et al.*, 2011). See Sun *et al.* (2011) for an overview of dimension reduction methods for spatial problems. Beyond dimension reduction techniques, compactly supported covariance functions (Kaufman *et al.*, 2011) or functional methods (Higdon *et al.*, 2008; Fricker *et al.*, 2013; Paulo *et al.*, 2012) could also be considered.

Because LFM-MIX is a multiple-fidelity computer model, our LFM-MIX emulator needs to be able to emulate both single and double fidelity. Building a multiple-fidelity emulator, although more complex, is advantageous and can greatly facilitate calibration, i.e., by positing a statistical emulator which links the double- and single-fidelity outputs (i.e. utilizes correlations between the two fidelities), the computationally cheap single-fidelity LFM-MIX can be run at more initial conditions and used to predict the double-fidelity output at those same initial conditions (saving computation time).

The second or 'validation' layer of the model links observations and computer model output (see Bayarri *et al.* (2007)). The traditional approach to linking model output to observations follows Kennedy and O'Hagan (2001) who modelled

$$\text{observations} = \text{high fidelity output at } \boldsymbol{\chi} + \text{discrepancy} + \text{observational error}$$

where the discrepancy term describes deviations of the computer model from reality and acknowledges the possibility that the computer model does not capture all aspects of the physical process being modelled. Inclusion of a discrepancy presents several issues such as a high correlation between the calibration parameter $\boldsymbol{\chi}$ and the discrepancy term which leads to confounding (see Bayarri *et al.* (2007, 2009) and Liu *et al.* (2009) for more discussion). In spite of this confounding effect, use of a discrepancy term is necessary because LFM-MIX, even if $\boldsymbol{\chi}$ were known, is not expected to capture all dynamics of the physical process. To alleviate confounding issues partially, we assume *a priori* independence between the three components above. Furthermore, we model the discrepancy term by using a low rank Gaussian process. By using a low rank Gaussian process, the discrepancy term is identified by using relatively few parameters, leaving more information (degrees of freedom) to estimate $\boldsymbol{\chi}$.

1.3. Contributions and outline of the paper

To conclude and summarize this section, the main contributions of this paper are as follows:

(a) posit low rank and non-additive methods for dealing with discrepancies between computer model output and observations
(b) utilize a low rank Gaussian process response surface methodology to implement a fast computer model emulator and
(c) describe a low rank method for leveraging multiple-fidelity computer model output.

This analysis builds on and extends the work of Kleiber *et al.* (2013). To contrast this work with Kleiber *et al.* (2013), we consider both the energy and the flux LFM-MIX output fields whereas Kleiber *et al.* (2013) considered only energy. Kleiber *et al.* (2013) also did not include discrepancy between the calibrated LFM-MIX output and the observations. Here, we show that accounting for this discrepancy is necessary to account for deficiencies in LFM-MIX.

The remainder of this paper is structured as follows. Section 2 describes the output of LFM-MIX, observations and the experimental design points for initially exploring the space of input settings. Section 3 describes the statistical model to achieve calibration of LFM-MIX. Section 4 describes the results of a cross-validation study to validate the statistical emulator of Section 3. Section 5 describes the calibration results and Section 6 summarizes lessons learned and outlines future research directions.

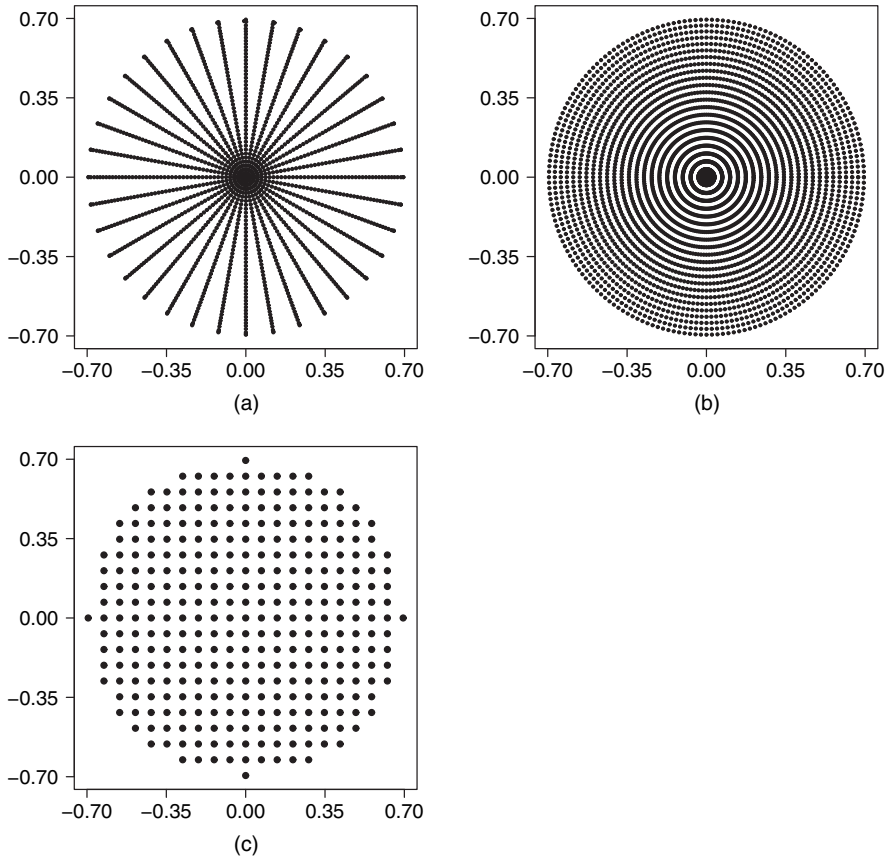

Fig. 1. (a) 46 × 36 spatial grid of observations, (b) 27 × 180 spatial grid of LFM-MIX model output and (c) knot points for the predictive process specifications: each grid is expressed from a polar projection where (0, 0) corresponds to the North Pole

2. Observational data and LFM-MIX model output

For this paper, LFM-MIX is calibrated using data from a geomagnetic storm which occurred on January 10th, 1997. For this storm, observations of ionospheric energy and energy×flux (which converts to megawatts per square metre) were drawn from remote sensing instruments and recorded approximately every 7 min between 2.00 p.m. and 4.00 p.m. Universal Time Coordinated (18 total time points) on the 46×36 spatial grid given in Fig. 1(a). For each time point, however, only a fraction of the grid is observed. The spatial grids in Fig. 1 are from the viewpoint of looking down on the Earth (for example the point $(0, 0)$ corresponds to the North Pole) with larger values of $\sqrt{(x^2 + y^2)}$ corresponding to lower latitudes. The spatial grid in Fig. 1(a) covers a latitude span from 45° to 90° (in steps of 1° latitude) and all longitudes (from 0° to 360° in steps of approximately 10°). Figs 2(a) and 2(d) display an example of the observed ionospheric energy and energy×flux at the second time point.

Given the solar wind data on January 10th, 1997, the discrete set of time points $\mathcal{T}$ corresponding to the observations and an input value $\mathbf{x} = (x_1, x_2, x_3)'$, the LFM-MIX model outputs ionospheric energy on the 27×180 spatial grid given in Fig. 1(b) for all $t \in \mathcal{T}$. Expressed in terms of latitudes and longitudes, the spatial grid in Fig. 1(b) covers a latitude span from 46° to 90° (with 46° latitude corresponding to the outer circle of grid points) along with all longitudes (steps of 2° from 0° to 358°). Owing to the misalignment of the observational and the LFM-MIX output grid (Figs 1(a) and 1(b)), this analysis follows Kleiber *et al.* (2013) by *a priori* using thin plate splines to krige the LFM-MIX output to the observational grid. Alternatively, this kriging

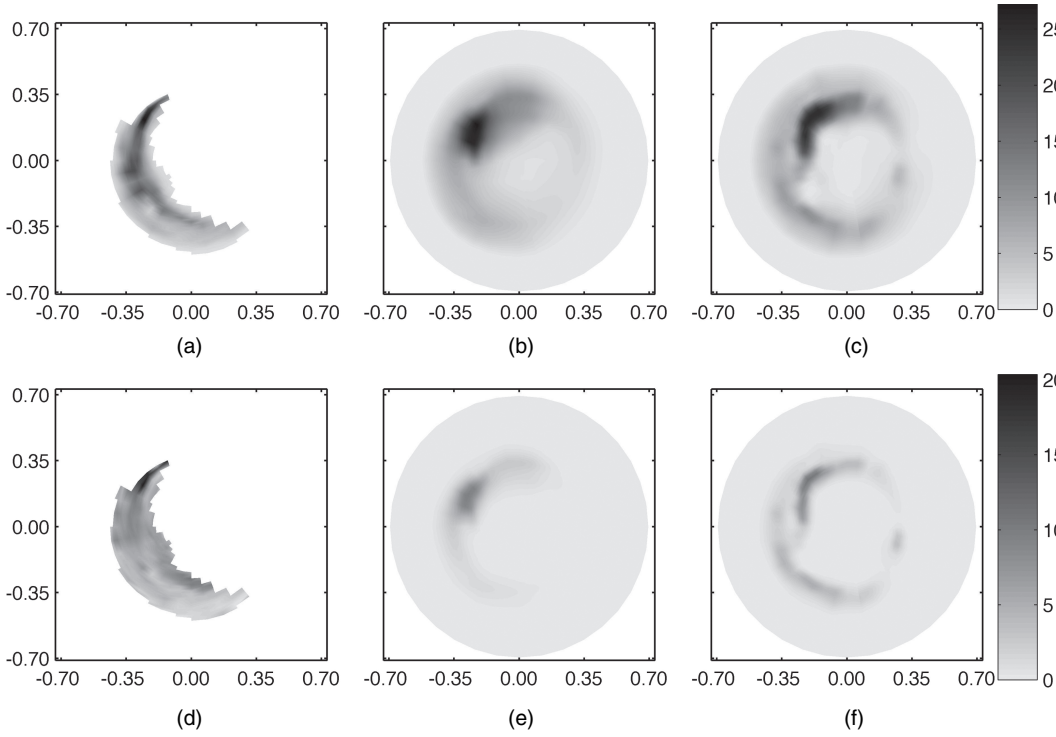

Fig. 2. (a)–(c) Ionospheric energy and (d)–(f) energy × flux at time $t = 2$ according to (a), (d) observations, (b), (e) single-fidelity LFM-MIX output and (c), (f) double-fidelity LFM-MIX output for $\mathbf{x} = (0.473, 1.591, 0.030)$: observed ionospheric energy and energy–flux are observed for only a fraction of the grid in Fig. 1(a) at each time point

could be done within the statistical modelling but given the density of the LFM-MIX grid in Fig. 1(a) *a priori* kriging to the observation grid introduces little error while greatly reducing the dimension of the spatial field.

According to the developers and users of LFM-MIX, the set of physically plausible input values is $\mathcal{X} = \{(x_1, x_2, x_3) : x_1 \in (0, 0.5), x_2 \in (0, 2.5), x_3 \in (0, 0.1)\}$. To explore $\mathcal{X}$ properly to find the most appropriate value $\chi \in \mathcal{X}$ which mimics ionospheric observations, we used the following strategy to choose $P_S = 20$ values of $\mathbf{x} \in \mathcal{X}$ to run the single-fidelity model. The first 16 points were chosen according to an orthogonal array Latin hypercube design (McKay *et al.*, 1979) following which an additional four values were chosen (totalling $P_S = 20$) according to a maximin space filling criterion (Johnson *et al.*, 1990). From these 20 values, an additional $P_D = 5$ were chosen according to a geometric space filling design of the original 20 (using `cover.design` within the `fields` package in R; Furrer *et al.* (2013)) to run the double-fidelity LFM-MIX model resulting in 25 distinct runs of the computer code. The design points are shown in Fig. 3 in Section 5. One run of the single-fidelity LFM-MIX model requires approximately 16 h to complete whereas the double-fidelity LFM-MIX model requires nearly 85 h by using an AMD Opteron 2.4 GHz Quad-Core processor with 125 Gbytes of random-access memory. Figs 2(b) and 2(e), and 2(c) and 2(f) display an example of single- and double-fidelity LFM-MIX model output respectively, at the same time point as in Figs 2(a) and 2(d), using an input value of $\mathbf{x} = (0.473, 1.591, 0.030)$. Contrasting Figs 2(b) and 2(e), and 2(c) and 2(f), note that the double-fidelity LFM-MIX can capture finer scale spatial features than the single-fidelity LFM-MIX model, albeit at the cost of additional computation.

3. Methodology

3.1. Single-fidelity model emulator

Owing to the physical constraint that energy and flux are positive values, let $S_e(\mathbf{l}, t; \mathbf{x})$ and $S_f(\mathbf{l}, t; \mathbf{x})$ denote the single-fidelity LFM-MIX output of log(energy) and log(flux) run at input setting $\mathbf{x} = (x_1, x_2, x_3)$, spatial location $\mathbf{l}$ and time period t. An exploratory analysis (see the scatter plots in the on-line supplement) gave evidence towards a linear relationship between $S_e(\mathbf{l}, t; \mathbf{x})$ and $S_f(\mathbf{l}, t; \mathbf{x})$ which is in harmony with the mathematical relationship given in equations (1) and (2). However, this linear relationship varied over space and time. We then choose to model

$$S_e(\mathbf{l}, t; \mathbf{x}) = \mu_{S_e}(\mathbf{l}, t) + w_{S_e}(\mathbf{l}, t; \mathbf{x}), \tag{4}$$

$$S_f(\mathbf{l}, t; \mathbf{x}) = \mu_{S_f}(\mathbf{l}, t) + \beta_{S_f}(\mathbf{l}, t)\, w_{S_e}(\mathbf{l}, t; \mathbf{x}) + w_{S_f}(\mathbf{l}, t; \mathbf{x}) \tag{5}$$

where $\mu_{S_e}(\mathbf{l}, t)$ and $\mu_{S_f}(\mathbf{l}, t)$ are spatiotemporal mean terms and determine the average energy and flux for each spatiotemporal location, $\beta_{S_f}(\mathbf{l}, t)$ is a spatiotemporal slope and $w_{S_e}(\mathbf{l}, t; \mathbf{x})$ and $w_{S_f}(\mathbf{l}, t; \mathbf{x})$ are Gaussian random effects where w_{S_e} and w_{S_f} are independent. Equations (4) and (5) specify the joint distribution of S_e and S_f via the marginal distribution for S_e and the conditional distribution for $S_f | S_e$ (Royle and Berliner, 1999). This specification was made for the following reasons. First, equations (1) and (2) suggest a natural ordering to the energy and flux variables. Specifically, equation (2) suggests that flux is linearly related to energy on the log-scale. Second, the conditional specification is a simple way to model a non-stationary relationship between S_e and S_f by allowing the slope and mean parameters to vary over space and time. And, third, the conditional specification is simpler computationally because computations can now deal with two smaller covariance matrices rather than a single large matrix. Note, however, that the joint process specification could also be done by using linear models of coregionalization (Gelfand

et al., 2004), multivariate covariance functions (see, for example, Gneiting *et al.* (2010) and Apanasovich *et al.* (2012)) or latent dimensions (Apanasovich and Genton, 2010).

The random effects w_{S_e} and w_{S_f} play the important role of correlating LFM-MIX output across spatial locations, time periods and input settings. This correlation allows for prediction to untried input settings. However, from Section 2, recall that available LFM-MIX output consists of 20 runs of 18 time periods of a 46×36 spatial grid resulting in 596160 correlated data points. Inverting and calculating the determinant of such a large covariance matrix is not computationally feasible so simplifying assumptions will need to be used. Following, for example, Bhat *et al.* (2012), using a separable covariance structure in input setting, time and space would reduce computation to inverting and calculating the determinant of a 20×20, 18×18 and a 1656×1656 matrix. Whereas the 20×20 and 18×18 matrix pose no computational challenges, repeatedly inverting a 1656×1656 matrix within a Markov chain Monte Carlo algorithm does and requires further dimension reduction. The strategy here, then, is to retain a Kronecker product structure for the temporal and input setting domain while reducing the dimension of the spatial covariance matrix. Details are as follows.

Let $\mathbf{w}^*_{S_e}(t;\mathbf{x}) = (w^*_{S_e}(\mathbf{l}^*_1, t;\mathbf{x}), \ldots, w^*_{S_e}(\mathbf{l}^*_K, t;\mathbf{x}))'$ be a vector of zero-mean Gaussian random effects at K spatial knot points $\mathbf{l}^*_1, \ldots, \mathbf{l}^*_K$. Furthermore, assume that the covariance for any two points $\mathbf{u}_i = (\mathbf{l}^*_i, t_i; \mathbf{x}_i)$ and $\mathbf{u}_j = (\mathbf{l}^*_j, t_j; \mathbf{x}_j)$ follows a separable form such that

$$\mathbb{C}\{w^*_{S_e}(\mathbf{u}_i), w^*_{S_e}(\mathbf{u}_j)\} = \sigma^2_{S_e} \mathcal{M}_3\{d_{\mathbf{l}}(i,j)|\phi_{S_e}\}\, \mathcal{M}_3\{d_t(i,j)|\psi_{S_e}\} \prod_{r=1}^{3} \mathcal{M}_3\{d_{x_r}(i,j)|\gamma_{S_e,r}\} \qquad (6)$$

where $\mathbb{C}(\cdot)$ denotes covariance, $d_{\mathbf{l}}(i,j) = \|\mathbf{l}^*_i - \mathbf{l}^*_j\|$, $d_t(i,j) = \|t_i - t_j\|$ and $d_{x_r}(i,j) = \|x_{ir} - x_{jr}\|$ denotes the Euclidean distance in the spatial, temporal and input setting domains respectively, and $\mathcal{M}_\nu(d|\phi)$ is the isotropic Matérn covariance function with smoothness parameter ν and decay parameter ϕ at a distance d. From equation (6), note the following assumptions. First, the covariance function for $w^*_{S_e}$ is separable, allowing a computationally feasible Kronecker product structure for the covariance matrix. The separability of the covariance function is key for producing a computationally feasible estimation method. Alternatively, non-separable spatiotemporal covariance functions like those proposed by Gneiting (2002) or Stein (2005) could be used to capture more complex spatiotemporal dependence structures. However, because the primary interest here is prediction to new input settings (as opposed to prediction to new spatiotemporal locations), we opt to use the simpler separable structure to account for space–time dependences. Separability in the input setting space is a standard choice (Santner *et al.*, 2003). Second, the smoothness parameters of each Matérn covariance function is fixed at 3. This value for the smoothness parameter was chosen to allow each of the spatial, temporal and input setting fields to be mean square differentiable but not infinitely so as is the case of the Gaussian covariance function (see Stein (1999)). Additionally, because estimating the smoothness is a notoriously difficult problem, fixing the smoothness is a common practice (Rasmussen and Williams, 2006; Gneiting *et al.*, 2012). And, third, Euclidean distance is used within each Matérn covariance function. Whereas Euclidean distance is appropriate for the temporal and input setting domains, because the spatial domain of interest is the northern hemisphere, it could be argued that such a distance metric is not appropriate for the spatial domain (see Banerjee (2005) and Jun and Stein (2007, 2008)). Note, however, that the spatial grid in Fig. 1(a) is not in typical latitude–longitude co-ordinates but is rather a polar projection. Euclidean distance on the projected polar grid is more appropriate than Euclidean distance of latitude–longitude co-ordinates because, for example, Euclidean distance on latitude–longitude co-ordinates would inappropriately quantify the distance between points at the North Pole. Furthermore, Gneiting (2013) showed that using great circle distance would not produce a positive definite covariance function with the degree

of smoothness that is seen in LFM-MIX output. For these reasons, we opted to use Euclidean distance on the polar project co-ordinates despite its deficiencies.

To reduce the dimension of the spatial domain from 1656 to $K \ll 1656$, let $w_{S_e}(\mathbf{l}, t; \mathbf{x})$ be given by the modified predictive process of Banerjee *et al.* (2004) and Finley *et al.* (2009) such that

$$w_{S_e}(\mathbf{l}, t; \mathbf{x}) = \mathbf{b}'_{w_e}(\mathbf{l})\, \mathbf{w}^*_{S_e}(t; \mathbf{x}) + \varepsilon_e(\mathbf{l}, t; \mathbf{x}) \tag{7}$$

where $\mathbf{b}_{w_e}(\mathbf{l}) = (b_{w_e,1}(\mathbf{l}), \ldots, b_{w_e,K}(\mathbf{l}))'$ is a K-vector of predictive process basis functions such that $b_{w_e,i}(\mathbf{l})$ is the ith element of $\boldsymbol{\sigma}'_{\mathbf{ll}^*}\boldsymbol{\Sigma}^{-1}_{\mathbf{l}^*}$ where $\boldsymbol{\Sigma}_{\mathbf{l}^*} = \mathrm{var}\{\mathbf{w}^*_{S_e}(t; \mathbf{x})\}$ and $\boldsymbol{\sigma}_{\mathbf{ll}^*} = \sigma^2_{S_e}(\mathcal{M}_3(\|\mathbf{l} - \mathbf{l}^*_1\|\,|\phi_{S_e}), \ldots, \mathcal{M}_3(\|\mathbf{l} - \mathbf{l}^*_K\|\,|\phi_{S_e}))'$. The $\varepsilon_e(\mathbf{l}, t; \mathbf{x})$ in equation (7) are the variance correction terms for the predictive process (see Finley *et al.* (2009)). We alter the framework of Finley *et al.* (2009) slightly by correlating the error terms across time and input setting by assuming that $\varepsilon_e(\mathbf{l}, t; \mathbf{x})$ follow a spatially independent zero-mean Gaussian process with covariance function

$$\mathbb{C}\{\varepsilon_e(\mathbf{l}_i, t_i; \mathbf{x}_i), \varepsilon_e(\mathbf{l}_j, t_j; \mathbf{x}_j)\} = \sigma^2_{\varepsilon_e}(\mathbf{l})\, \mathcal{M}_3\{d_t(i, j)|\psi_{S_e}\} \prod_{r=1}^{3} \mathcal{M}_3\{d_{x_r}(i, j)|\gamma_{S_e,d}\} \tag{8}$$

if $\mathbf{l}_i = \mathbf{l}_j$ and $\mathbb{C}\{\varepsilon_e(\mathbf{l}_i, t_i; \mathbf{x}_i), \varepsilon_e(\mathbf{l}_j, t_j; \mathbf{x}_j)\} = 0$ otherwise where $\sigma^2_{\varepsilon_e}(\mathbf{l}) = \sigma^2_{S_e}\{1 - \mathbf{b}'_{w_e}(\mathbf{l})\boldsymbol{\Sigma}_{\mathbf{l}^*}\mathbf{b}_{w_e}(\mathbf{l})\}$. From the predictive process specification in equation (7), note that $\mathrm{var}\{w_{S_e}(\mathbf{l}, t; \mathbf{x})\} = \sigma^2_{S_e}$. Furthermore, if $\mathbf{w}_{S_e}$ is the stacked vector of $\{w_{S_e}(\mathbf{l}, t; \mathbf{x})\}$ ordered by spatial locations, temporal locations and input settings locations, then $\mathrm{var}(\mathbf{w}_{S_e}) = \sigma^2_{S_e}(\mathbf{R}_{\mathbf{x}} \otimes \mathbf{R}_{\mathcal{T}} \otimes \mathbf{R}_{\mathcal{L}})$ where $\mathbf{R}_{\mathbf{x}} = \{\Pi^3_{r=1}\mathcal{M}_3\{d_{x_r}(i, j)\}\}_{i,j}$ is the 20×20 correlation matrix across input settings, $\mathbf{R}_{\mathcal{T}} = \{\mathcal{M}_3\{d_t(i, j)\}\}_{i,j}$ is the 18×18 correlation matrix across time periods and $\mathbf{R}_{\mathcal{L}} = \mathbf{B}_{w_e}\boldsymbol{\Sigma}_{\mathbf{l}^*}\mathbf{B}'_{w_e} + \mathbf{D}_{\varepsilon_e}$ controls the spatial correlation where $\mathbf{B}_{w_e}$ is a $1656 \times K$ matrix with ith row $\mathbf{b}'_{w_e}(\mathbf{l}_i)$ where $\mathbf{l}_i$ is the ith spatial location and $\mathbf{D}_{\varepsilon_e} = \sigma^{-2}_{S_e}\,\mathrm{diag}\{\sigma^2_{\varepsilon_e}(\mathbf{l}_1), \ldots, \sigma^2_{\varepsilon_e}(\mathbf{l}_{1656})\}$. Note that the rank of $\mathbf{R}_{\mathcal{L}}$ is K and $\mathbf{R}_{\mathcal{L}}$ can be inverted by using the Sherman–Morrison–Woodbury formula (see Harville (1997), section 4). Reducing the dimension of the spatial field resulted in, approximately, 80% savings in computation time.

To emphasize, by correlating the variance correction factors $\{\varepsilon_e(\mathbf{l}, t; \mathbf{x})\}$ across time and input setting, the predictive process specification in equation (7) reduces only the dimension of the spatial domain (not the temporal and input setting domains). Because we are data poor in the temporal and input setting domains, dimension reduction is neither necessary nor desirable. The predictive process was used here mainly for interpretability, i.e. the predictive process knot vector $\mathbf{w}^*_{S_e}(t; \mathbf{x})$ is interpreted as the original spatial surface at the knot points.

Having developed a computationally feasible specification for w_{S_e}, the random effects for the flux variable w_{S_f} in equation (5) were assumed to follow the same process structure as w_{S_e} in equation (7) but parameters were changed to be variable specific (i.e. all single-fidelity energy-specific parameters are replaced by single-fidelity flux-specific parameters).

3.2. Double-fidelity model emulator

As with single-fidelity LFM-MIX output, let $D_e(\mathbf{l}, t; \mathbf{x})$ and $D_f(\mathbf{l}, t; \mathbf{x})$ denote log(energy) and log(flux) for the double-fidelity model. Key pieces for an emulator of the double-fidelity model are

(a) leveraging the information in the single-fidelity model to predict the double fidelity and
(b) correlating the double-fidelity model across space, time and input setting.

Exploratory analysis suggests that the relationship between the single and double fidelity is linear but varied over the space–time domain. Hence, we model

$$D_e(\mathbf{l}, t; \mathbf{x}) = \mu_{D_e}(\mathbf{l}, t) + \beta_{D_e}(\mathbf{l}, t)\, w_{S_e}(\mathbf{l}, t; \mathbf{x}) + w_{D_e}(\mathbf{l}, t; \mathbf{x}), \tag{9}$$

$$D_{\mathrm{f}}(\mathbf{l},t;\mathbf{x}) = \mu_{D_{\mathrm{f}}}(\mathbf{l},t) + \beta_{D_{\mathrm{f}}}(\mathbf{l},t)\, w_{S_{\mathrm{f}}}(\mathbf{l},t;\mathbf{x}) + \theta_{D_{\mathrm{f}}}(\mathbf{l},t)\, w_{D_{\mathrm{e}}}(\mathbf{l},t;\mathbf{x}) + w_{D_{\mathrm{f}}}(\mathbf{l},t;\mathbf{x}) \tag{10}$$

where $\mu_{D_{\mathrm{e}}}(\mathbf{l},t)$ and $\mu_{D_{\mathrm{f}}}(\mathbf{l},t)$ are spatiotemporal mean fields that determine the mean energy and flux for the double-fidelity model across space and time, $\beta_{D_{\mathrm{e}}}(\mathbf{l},t)$, $\beta_{D_{\mathrm{f}}}(\mathbf{l},t)$ and $\theta_{D_{\mathrm{f}}}(\mathbf{l},t)$ are spatiotemporal slope fields, and $w_{D_{\mathrm{e}}}(\mathbf{l},t;\mathbf{x})$ and $w_{D_{\mathrm{f}}}(\mathbf{l},t;\mathbf{x})$ are random effects with $w_{D_{\mathrm{e}}} \perp\!\!\!\perp w_{D_{\mathrm{f}}}$ for all $(\mathbf{l},t;\mathbf{x})$. Note that equations (4), (5), (9) and (10) conditionally define a valid joint distribution $[S_{\mathrm{e}}, S_{\mathrm{f}}, D_{\mathrm{e}}, D_{\mathrm{f}}]$ by using our assumptions. Hence, these equations define a full statistical emulator for single- and double-fidelity LFM-MIX output and appropriately leverage single-fidelity information.

For the double-fidelity emulator in equations (9) and (10), note that we are data poor. Recall from Section 2 that the double-fidelity LFM-MIX was only run at a total of $P_{\mathrm{D}} = 5$ input settings compared with $P_{\mathrm{S}} = 20$ for the single-fidelity model. So few data points give only a few degrees of freedom to estimate all the parameters in equations (9) and (10). To increase the amount of information in the data to estimate the parameter surfaces, this analysis models the spatiotemporal mean and slope surfaces (i.e. $\mu_{D_{\mathrm{e}}}$, $\beta_{D_{\mathrm{e}}}$, $\mu_{D_{\mathrm{f}}}$, $\beta_{D_{\mathrm{f}}}$ and $\theta_{D_{\mathrm{f}}}$) by using predictive process basis functions. By using a basis function expansion, the dimension of these parameter surfaces is reduced and strength can be borrowed across spatiotemporal locations for estimation purposes. For example, if $\beta_{D_{\mathrm{e}}}(\mathbf{l},t) = \mathbf{b}'_{\beta_{\mathrm{e}}}(\mathbf{l},t)\boldsymbol{\beta}^*_{D_{\mathrm{e}}}$ where $\mathbf{b}'_{\beta_{\mathrm{e}}}(\mathbf{l},t)$ is a set of basis functions (in this case predictive process basis functions arising from the covariance function given in equation (6)) and $\boldsymbol{\beta}^*_{D_{\mathrm{e}}}$ is the corresponding reduced dimension vector of coefficients, then estimating an entire surface reduces to estimating the coefficients $\boldsymbol{\beta}^*_{D_{\mathrm{e}}}$. Under this predictive process specification, $\boldsymbol{\beta}^*_{D_{\mathrm{e}}}$ is then given a zero-mean Gaussian process prior with a covariance function that corresponds to the covariance function that is used to build the basis functions.

Our construction of a multiple-fidelity emulator follows most closely that of Qian and Wu (2008). Rather than the conditional specification in equations (9) and (10), however, Cumming and Goldstein (2009) proposed an alternative strategy for multiple-fidelity computer models, i.e. they used the same set of regressors for the single- and double-fidelity models (the $\mathbf{b}_{w_{\mathrm{e}}}(\mathbf{l})$ in equation (7)) but correlated the coefficients ($\mathbf{w}^*_{S_{\mathrm{e}}}(t;\mathbf{x})$). We note that this approach may reduce the dimension of the coefficient surfaces $\beta_{D_{\mathrm{e}}}(\mathbf{l},t)$, $\beta_{D_{\mathrm{f}}}(\mathbf{l},t)$ and $\theta_{D_{\mathrm{f}}}(\mathbf{l},t)$ and result in a more parsimonious model structure.

3.3. Observation model

Let $Y_{\mathrm{e}}(\mathbf{l},t;\boldsymbol{\chi})$ and $Y_{\mathrm{f}}(\mathbf{l},t;\boldsymbol{\chi})$ denote the observed log(energy) and log(energy × flux) at time t and spatial location $\mathbf{l}$. Recall that $\boldsymbol{\chi} = (\chi_1, \chi_2, \chi_3)'$ represents the unknown best input setting for the double-fidelity LFM-MIX model output to mimic field data. To emphasize, we wish to calibrate to the double-fidelity output as it represents a more accurate representation of the underlying physical processes in the observations. To estimate $\boldsymbol{\chi}$, relate the observations to double-fidelity LFM-MIX output at $\boldsymbol{\chi}$ via

$$Y_{\mathrm{e}}(\mathbf{l},t;\boldsymbol{\chi}) = D_{\mathrm{e}}\{\mathbf{R}(\lambda_t)\mathbf{l},t;\boldsymbol{\chi}\} + \delta_{\mathrm{e}}(\mathbf{l},t) + \zeta_{Y_{\mathrm{e}}}(\mathbf{l},t), \tag{11}$$

$$Y_{\mathrm{f}}(\mathbf{l},t;\boldsymbol{\chi}) = D_{\mathrm{e}}\{\mathbf{R}(\kappa_t)\mathbf{l},t;\boldsymbol{\chi}\} + D_{\mathrm{f}}\{\mathbf{R}(\kappa_t)\mathbf{l},t;\boldsymbol{\chi}\} + \delta_{\mathrm{f}}(\mathbf{l},t) + \zeta_{Y_{\mathrm{f}}}(\mathbf{l},t) \tag{12}$$

where $\delta_{\mathrm{e}}(\mathbf{l},t)$ and $\delta_{\mathrm{f}}(\mathbf{l},t)$ are independent spatiotemporal discrepancy functions, and $\zeta_{Y_{\mathrm{e}}}$ and $\zeta_{Y_{\mathrm{f}}}$ are Gaussian white noise random variables with variance $\sigma^2_{Y_{\mathrm{e}}}$ and $\sigma^2_{Y_{\mathrm{f}}}$ respectively. The $\mathbf{R}(\lambda_t)$ and $\mathbf{R}(\kappa_t)$ are clockwise rotation matrices given by

$$\mathbf{R}(\lambda_t) = \begin{pmatrix} \cos(\lambda_t) & \sin(\lambda_t) \\ -\sin(\lambda_t) & \cos(\lambda_t) \end{pmatrix}, \tag{13}$$

such that $\mathbf{R}(\lambda_t)\mathbf{l}$ rotates the location $\mathbf{l}$ clockwise by an angle λ_t.

According to equations (11) and (12), discrepancies between LFM-MIX and observations occur in two places. First, LFM-MIX differs from observations by a non-linear rotational bias of angle λ_t (for energy) and κ_t (for energy$\times$flux). This rotational discrepancy is apparent from Fig. 2 as the modes of LFM-MIX and the observations are misaligned. Furthermore, on the basis of personal discussions with the scientists at the CISWM, this rotational bias is highly informational in developing future versions of LFM-MIX. Second, after rotation, any remaining discrepancy between LFM-MIX and the observations is captured by the additive discrepancy terms δ_e and δ_f.

The additive discrepancy functions δ_e and δ_f are highly correlated with $\boldsymbol{\chi}$, i.e. a change of input setting (and, hence, a change of LFM-MIX output) necessarily changes the value of these discrepancy functions, leading to confounding of the input parameter $\boldsymbol{\chi}$ (see Bayarri *et al.* (2007)). Thus, to identify $\boldsymbol{\chi}$, let δ_e and δ_f be determined by a set of predictive process basis functions such that $\delta_e(\mathbf{l}, t) = \mathbf{b}'_{\delta_e}(\mathbf{l}, t)\boldsymbol{\delta}^*_e$ and $\delta_f(\mathbf{l}, t) = \mathbf{b}'_{\delta_f}(\mathbf{l}, t)\boldsymbol{\delta}^*_f$. Using basis functions not only gives dimension reduction for the spatiotemporal discrepancy fields but also restricts the flexibility of the discrepancy function. By so restricting the discrepancy surfaces, confounding of $\boldsymbol{\chi}$ is partially alleviated because the calibrated LFM-MIX output is forced to capture more of the spatiotemporal structure in the observations. As seen in Section 5, however, the discrepancy functions are still sufficiently flexible to capture the remaining structure in the observations.

Details on calculating the likelihood of observations and LFM-MIX output are provided in Appendix A.

3.4. Prior assumptions

The general strategy in selecting prior distributions was to be non-informative and to allow the data to inform the parameter value except in cases where prior information is necessary. We assumed $\chi_1 \sim \mathcal{U}(0.0, 0.5)$, $\chi_2 \sim \mathcal{U}(0, 2.5)$ and $\chi_3 \sim \mathcal{U}(0, 0.1)$ which accurately reflect the applied scientist's understanding of these parameters. Each variance term (e.g. $\sigma^2_{Y_e}$ and $\sigma^2_{S_e}$) was assumed to follow an inverse gamma distribution with shape parameter 2.01 and rate parameter 1.01 (such that the prior expectation is 1 and the prior variance is 100) with the exception of $\sigma^2_{Y_e}$ and $\sigma^2_{Y_f}$ which were fixed at known measurement error values obtained from CISWM scientists. Priors for the mean and slope surfaces of the single-fidelity emulator (e.g. $\mu_{S_e}(\mathbf{l}, t)$ in equation (4)) were assumed to be independent (over space and time) and uniform over a wide interval centred at zero. Alternatively, these surfaces could be given a prior that accounts for the spatiotemporal structure but smoothing these surfaces is not of interest in this study so we assume independence to facilitate computation.

For the double- and single-LFM-MIX output, the priors for the decay parameters within the covariance function in equation (6) were induced by assuming a discrete uniform distribution at ranges of 10%, 25%, 50%, 75% and 90% of the maximum corresponding distance. For example, consider $\gamma_{S_e,2}$ which corresponds to the correlation range in the dimension of $x_2 \in (0, 2.5)$. The prior for $\gamma_{S_e,2}$ was induced by assuming that the correlation range (the distance at which the correlation decays to 0.05) is distributed uniformly at distances of 10%, 25%, 50%, 75% and 90% of 2.5 (which is the maximum distance between two points on (0,2.5)). Such highly informative priors for the decay parameters are necessary because of equivalence of probability measures (see Zhang (2004)).

In contrast with the decay parameters in the covariance function of single- and double-LFM-MIX output, we chose to fix the decay parameters in specifying the predictive process basis functions for the mean and slope surfaces in equations (9) and (10) and for the discrepancy

functions (i.e. δ_e and δ_f) in equations (11) and (12). The values of the decay parameters for the mean and slope surfaces were chosen such that the corresponding ranges were 50% of the maximum distance between spatiotemporal grid points to increase the amount of information borrowed between neighbouring points. For the discrepancy functions, however, we chose the decay parameters such that the spatial and temporal ranges were 10% of the maximum distance between spatiotemporal grid points. We chose a smaller range for the discrepancy surfaces to ensure that the discrepancies accounted for less spatiotemporal structure in the observations than the LFM-MIX output (thus, as discussed above, partially alleviating confounding between the discrepancy surfaces and χ). For each of these mean, slope and discrepancy surfaces, however, we estimated a separate prior variance parameter which, according to Zhang (2004), provides sufficient flexibility to fit a wide variety of surfaces.

Because geomagnetic storms evolve smoothly over time, the rotation angles λ_t and κ_t are also expected to be a smooth function over time. To account for this, λ_t and κ_t were assumed to follow a natural cubic spline with 3 degrees of freedom (two knots). For example, $\lambda_t = \Sigma_{i=1}^{3} n_i(t)\nu_i$ where $n_i(t)$ are the natural cubic spline basis functions and ν_i is the corresponding coefficient. Vague Gaussian priors were used for the coefficients of the spline bases.

4. Validating the emulator

The performance of the emulator that was described in Sections 3.1 and 3.2 is vital to the overall calibration of LFM-MIX, i.e. because the emulator will be used to predict the double-fidelity LFM-MIX output at untried input settings, if the prediction of the emulator is wrong, LFM-MIX will not be well calibrated. Thus, before calibration of LFM-MIX, a series of cross-validation studies were performed to validate the prediction performance of the emulator.

For the cross-validation studies, each of the $P_D = 5$ double-fidelity runs were sequentially left out and the emulator was used to predict the held-out run. For the first cross-validation study (study 'A'), the emulator was fitted by using the remaining four double-fidelity runs along with all $P_S = 20$ single-fidelity runs. For the second cross-validation study (study 'B'), the emulator was fitted by using the remaining four double-fidelity runs and 19 of the single-fidelity runs excluding the single run at the same input setting as the held-out double-fidelity run. Study B is more realistic than study A because neither a single- nor a double-fidelity run will be available at the best input setting. For the final study (study 'C'), the emulator was fitted by using only the remaining four double-fidelity runs (i.e. ignoring all information contained in the single-fidelity runs). By comparing study C with the other two, the value of leveraging the single-fidelity runs to predict the double-fidelity runs is exposed.

In specifying each predictive process, $K = 317$ spatial knots were used with locations in Fig. 1(c). These locations were chosen by placing a 21^2 grid over the spatial domain of LFM-MIX and removing locations outside the circular spatial region in Fig. 1(b). An informal analysis (the results are not shown) in selecting the number of knots was performed before running the cross-validation study from which $K = 317$ was chosen on the basis of an adequate trade-off between computational expense and accuracy. A more formal analysis following that of Guhaniyogi *et al.* (2011) could have been used to select the number and location of knots but the computational expense of doing so was too great in this case. Recall that, even though the same spatial grid is used to define each predictive process, we estimate separate decay parameters in the covariance function for each variable.

2500 draws from the posterior were retained for analysis after an initial burn-in of 1000 draws using a Gibbs sampler with Metropolis–Hastings steps where required. Although these are very few iterations for a Markov chain Monte Carlo algorithm, trace plots, empirical auto-

Table 1. Empirical bias, root-mean-square error RMSE, 95% prediction interval coverage and 95% prediction interval width when predicting the hold-out sample in the cross-validation study for energy *E* and energy×flux *EF*†

Parameter	*Study*	*Results for the following leave-out samples:*									
		1		*2*		*3*		*4*		*5*	
		E	*EF*	*E*	*EF*	*E*	*EF*	*E*	*EF*	*E*	*EF*
bias	A	−0.01	−0.01	0.01	0.04	0.07	−0.03	−0.23	−0.06	0.02	0.02
	B	−0.50	−0.21	−0.54	−0.17	−0.07	−0.10	−0.32	−0.24	0.06	0.02
	C	−4.71	−4.13	0.86	0.23	0.07	−0.08	0.99	0.39	0.22	0.11
RMSE	A	0.11	0.02	1.96	0.76	1.20	0.27	1.76	0.59	0.93	0.34
	B	1.15	5.83	2.83	1.04	1.24	0.42	3.32	3.56	0.95	0.35
	C	8.63	12.72	3.31	1.06	1.54	0.48	3.34	3.88	1.67	0.58
coverage	A	0.92	0.95	0.96	0.95	0.96	0.96	0.97	0.96	0.97	0.97
	B	0.66	0.87	0.98	0.99	0.96	0.98	0.99	0.99	0.96	0.98
	C	0.17	0.02	0.30	0.29	0.54	0.61	0.21	0.37	0.43	0.45
width	A	0.18	0.02	2.94	0.71	1.92	0.40	3.08	0.81	1.75	0.40
	B	1.83	0.94	5.77	2.21	2.71	0.80	5.90	2.75	1.91	0.50
	C	1.12	1.58	1.40	0.30	0.73	0.18	1.27	0.47	0.68	0.12

†The studies correspond to the emulator predictive performance when A the single-fidelity LFM-MIX output at the cross-validated input setting is retained, B the single-fidelity LFM-MIX output at the cross-validated input setting is discarded and C *all* the single-fidelity LFM-MIX output is discarded (i.e. the emulator is fitted to only the double-fidelity LFM-MIX output). Using the single-fidelity LFM-MIX output increases the emulator prediction performance according to bias and RMSE. Practically useful errors are 5% of the variability of the observations which equates to biases of 0.77 and 0.27 for energy and energy × flux respectively.

correlation functions and Gelman–Rubin diagnostics (Gelman and Rubin, 1992) indicated that this was sufficient for convergence.

Table 1 compares the predictive performance of the emulator across studies in terms of bias,

$$\text{bias} = \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \{\hat{D}(\mathbf{l}_n, t; \mathbf{x}) - D(\mathbf{l}_n, t; \mathbf{x})\},$$

where $\hat{D}(\mathbf{l}, t; \mathbf{x})$ is the mean of the posterior predictive distribution for the double-fidelity output $D(\mathbf{l}, t; \mathbf{x})$ (either energy or energy×flux), root-mean-square error

$$\text{RMSE} = \sqrt{\left[\frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \{\hat{D}(\mathbf{l}_n, t; \mathbf{x}) - D(\mathbf{l}_n, t; \mathbf{x})\}^2 \right]},$$

95% central predictive interval coverage

$$\text{coverage} = \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \mathbb{I}_{\{\tilde{D}^{(0.025)}(\mathbf{l}_n, t; \mathbf{x}) \leqslant D(\mathbf{l}_n, t; \mathbf{x})\}} \mathbb{I}_{\{\tilde{D}^{(0.975)}(\mathbf{l}_n, t; \mathbf{x}) \geqslant D(\mathbf{l}_n, t; \mathbf{x})\}},$$

where $\mathbb{I}_{\mathcal{A}}$ is an indicator for the set $\mathcal{A}$ and $\tilde{D}^{(q)}(\mathbf{l}; \mathbf{x})$ is the qth quantile of the posterior predictive distribution for $D(\mathbf{l}, t; \mathbf{x})$, and average 95% central predictive interval width

$$\text{width} = \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \{\tilde{D}^{(0.975)}(\mathbf{l}_n, t; \mathbf{x}) - \tilde{D}^{(0.025)}(\mathbf{l}_n, t; \mathbf{x})\}.$$

According to Table 1 the bias and RMSE for studies A and B are smaller than those for study

C, suggesting that leveraging single-fidelity output to predict double-fidelity output is beneficial. Furthermore, note that, across the hold-out samples, study A had less or equal RMSE than study B, again, pointing to the importance of using the single-fidelity model output to predict untried input settings for the double-fidelity LFM-MIX. This result is also encouraging because exploration of input setting space can be accomplished by using the computationally cheap single-fidelity model (recall that the single-fidelity model requires approximately 0.5 days to run whereas the double-fidelity model requires approximately 3.5 days).

Note, also, that the coverage rates for study C are very poor whereas the corresponding rates for studies A and B are more acceptable. This result arises in that our experimental design includes only $P_{\mathrm{D}} = 5$ double-fidelity runs of the computer model. As demonstrated by Table 1, five points are far from adequate for exploring the input setting space. However, given the additional 20 runs of the single-fidelity model, the space is more adequately explored. The coverage rates for the first leave-out sample in study B seem to be lower than for the other hold-out samples. This is because the LFM-MIX output at this input setting is drastically different from that at any other observed output. Hence, the single-fidelity run is highly informative about the output from the double-fidelity run such that coverage rates for study A are superior to coverage rates under study B.

5. Calibration of LFM-MIX

For the calibration of LFM-MIX to observations, the grid given by Fig. 1(c) was again used to define each of the predictive processes. For analysis, 10000 draws from the posterior distribution of model parameters were obtained after an initial burn-in of 10000 draws and thinning to every 10th draw to decrease auto-correlation (i.e. the Markov chain Monte Carlo algorithm was run for a total of $10000 + 10000 \times 10 = 110000$ draws). Five separate Markov chain Monte Carlo chains were run with well-dispersed starting locations for $\boldsymbol{\chi}$ to detect possible multimodality in the posterior distribution. A final (sixth) Markov chain Monte Carlo chain was run by using a proposal distribution that was a mixture of Gaussian distributions constructed from the previous five chains and a random walk. Specifically, for the final chain, proposal values for $\boldsymbol{\chi}$ were generated from

$$0.95\,\mathcal{N}(\boldsymbol{\chi}^{(t-1)}, \mathbf{S}_{\chi}^{(t-1)}) + \frac{0.05}{5} \sum_{c=1}^{5} \mathcal{N}(\mathbf{m}_c, \mathbf{S}_c)$$

where $\boldsymbol{\chi}^{(t-1)}$ was the current value for $\boldsymbol{\chi}$ in the Markov chain Monte Carlo chain, $\mathbf{S}_{\chi}^{(t-1)}$ is the current estimate of $\mathrm{var}(\boldsymbol{\chi}|\mathbf{Y})$ adapted according to Andrieu and Thoms (2008) and $\mathbf{m}_c$ and $\mathbf{S}_c$ are the estimate of $\mathbb{E}(\boldsymbol{\chi}|\mathbf{Y})$ and $\mathrm{var}(\boldsymbol{\chi}|\mathbf{Y})$ according to the results from original chain $c = 1, \ldots, 5$. By using this proposal distribution, the final chain can adequately jump between possible modes discovered from the original five chains.

Fig. 3 displays the marginal and bivariate posterior kernel density estimates for $\boldsymbol{\chi}$. For χ_1 and χ_3 the posterior distributions suggest values that are near the edge of the respective parameter space with posterior means of $\hat{\chi}_1 = 0.47$ and $\hat{\chi}_3 = 0.02$ where we use the 'hat' notation to denote posterior mean. The value of χ_2, however, is centred closer to the centre of possible values with a posterior mean of $\hat{\chi}_2 = 1.59$. We re-emphasize that the tuning parameters x_1, x_2 and x_3 describe the state of precipitating electrons from the magnetosphere to the ionosphere. Rigorously characterizing this precipitation process, however, is outside the scope of LFM-MIX. As such, the posterior mean $\hat{\boldsymbol{\chi}}$ cannot be interpreted in terms of LFM-MIX but, rather, is interpreted as physically plausible conditions for this precipitation process such that LFM-MIX output resembles observational data as closely as possible.

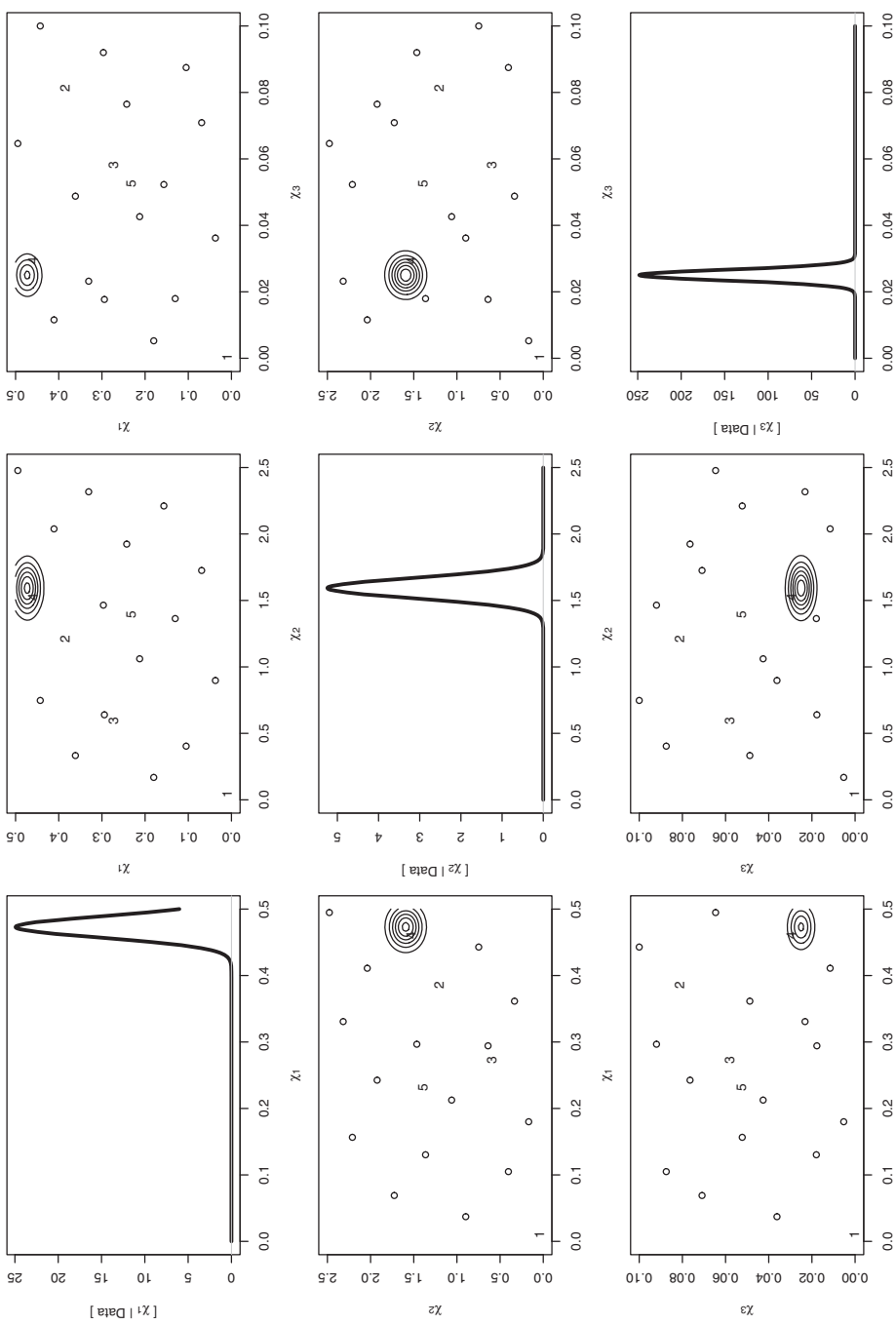

Fig. 3. Marginal (diagonal plots) and bivariate (off-diagonal plots) posterior kernel density estimates for χ: for the off-diagonal plots, numbers (corresponding to the leave-out sample in Table 1) denote design locations where LFM-MIX was run at double and single fidelity; ∘, design locations where LFM-MIX was run at single fidelity only

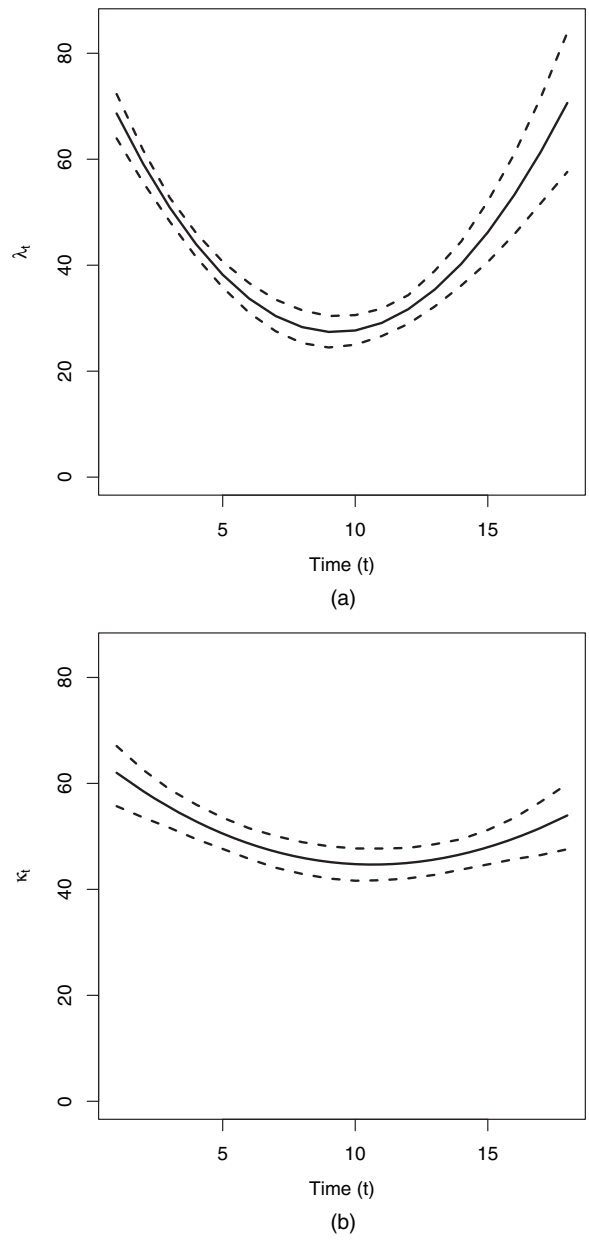

Fig. 4. Estimated counterclockwise rotations to align (a) the energy and (b) the energy × flux double-fidelity LFM-MIX output at χ with the observations: ———, posterior mean; – – –, 95% credible interval limits

We note that our estimate of the calibration value is near an observed run (run 4) of the double-fidelity LFM-MIX model. With only five runs of the double-fidelity model, having the good fortune of running the double-fidelity model near the 'correct' value is unlikely. Investigating this issue further, we determined that our estimate of $\boldsymbol{\chi}$ was near a design point for the following two reasons:

(a) short-range dependence across the input $\mathbf{x}$-space and

(b) insufficient detail in the single-fidelity model.

First, our estimates of the decay parameters (the γ-parameters in equation (6)) were large, suggesting a small correlation range across $\mathbf{x} \in \mathcal{X}$. These large estimates for the decay parameters suggest that LFM-MIX is highly variable as a function of $\mathbf{x}$. Hence, moving away from an observed LFM-MIX run results in predictions that resemble the mean field $\mu(\mathbf{s}, t)$ (which is inadequate). The best alternative is then to settle near an observed LFM-MIX run. Second, the single-fidelity LFM-MIX model lacked sufficient spatial detail to match the observations. This can be seen, partly, from Fig. 2 wherein the double-fidelity model has more spatial features than the corresponding single-fidelity output. Therefore, the double-fidelity LFM-MIX model is preferred to the single-fidelity model.

The relationship between χ_1, χ_2 and χ_3 is an important factor in selecting appropriate initial conditions for LFM-MIX. The estimate of the posterior correlation matrix for $\boldsymbol{\chi}$ is

$$\widehat{\text{corr}}(\boldsymbol{\chi}|\mathbf{Y}) = \begin{pmatrix} 1.00 & 0.00 & -0.10 \\ 0.00 & 1.00 & 0.10 \\ -0.10 & 0.10 & 1.00 \end{pmatrix}.$$

A posteriori, the correlations between the parameters are fairly small, suggesting that each input value may be set independently of one another and each controls a different aspect of LFM-MIX.

Fig. 4 displays the counterclockwise rotation that is necessary to align LFM-MIX and the observations (λ_t and κ_t). From Fig. 4, rotation of LFM-MIX is certainly required and, for some time periods, a counterclockwise rotation of about 60° is required. Whereas both the energy and the energy×flux fields are misaligned with the observations, the energy×flux field seems to be misaligned by an approximately time constant constant rotation. The rotation for the energy field, in contrast, requires a time varying rotation to align it with the observations. The requirement of a rotation of LFM-MIX to align with the observations is a previously unknown model discrepancy to CISWM scientists and current efforts are under way to explore this discrepancy in more detail.

Finally, Fig. 5 displays the calibrated LFM-MIX output to the observed ionospheric energy and energy×flux. Figs 5(b) and 5(e) display the rotated LFM-MIX output at the calibrated value $\boldsymbol{\chi}$ without the discrepancy term $\delta(\mathbf{s}, t)$. Comparing Figs 5(b) and 5(e) with the observations (Figs 5(a) and 5(d)), the rotational discrepancy is clearly aiding in aligning LFM-MIX with the observations. However, this comparison also clearly shows that the rotational discrepancy is insufficient given the large disparity between rotated LFM-MIX output and the observations. Figs 5(c) and 5(f) show the LFM-MIX output after rotations plus the discrepancy term $\delta(\mathbf{s}, t)$. By construction, the discrepancy term $\delta(\mathbf{s}, t)$ models the disparity between raw LFM-MIX output and the observations (hence, Figs 5(a) and 5(d), and 5(c) and 5(f) are nearly identical).

6. Conclusions

To use LFM-MIX to study space weather appropriately, proper initial conditions must be found. However, emulating and calibrating LFM-MIX to ionospheric observations poses several statistical challenges including high dimension, space–time correlations, multiple-fidelity output and non-additive discrepancies. To deal with these issues, this paper employed predictive processes as a low rank approach to emulate and leverage the multiple fidelities of LFM-MIX. Additionally, rotational and low rank discrepancies were introduced to account for differences between the LFM-MIX output and field observations.

Kleiber *et al.* (2013) also considered calibration of LFM-MIX but their results differed from

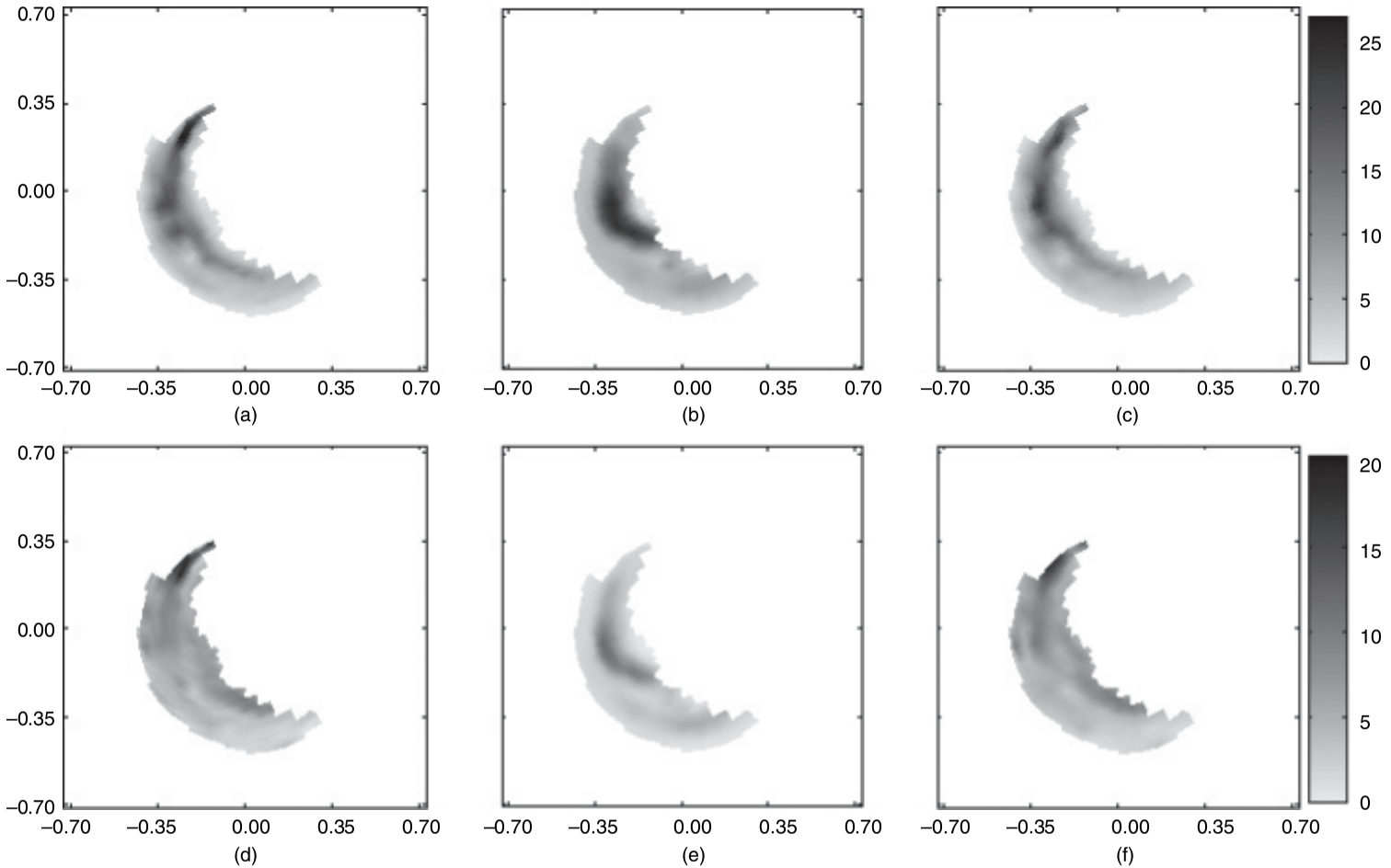

Fig. 5. (a)–(c) Ionospheric energy and (d)–(f) energy × flux at time $t = 2$ according to (a), (d) observations, (b), (e) posterior mean of calibrated double-fidelity LFM-MIX output after rotation and (c), (f) the posterior mean of the discrepancy-corrected double-fidelity LFM-MIX output: comparing raw LFM-MIX output in (b) and (e) with the observations in (a) and (d) reveals the value of including a discrepancy term to model any remaining variation in the observations that is not accounted for by LFM-MIX

those found here. Specifically, Kleiber *et al.* (2013) found that χ_2 was not well identified and $\chi_1 \approx 0.3$ and $\chi_3 \approx 0.4$ were appropriate calibration values. First, as suggested by equation (2), learning of χ_2 comes largely from the flux variable which was included in this analysis and not included in Kleiber *et al.* (2013). Second, the difference between the calibration results for χ_1 and χ_3 is mostly contingent on the choice of discrepancy function that is used. Here, rotations were used to align LFM-MIX output with observations more appropriately whereas Kleiber *et al.* (2013) chose not to use any discrepancy. Fundamentally, different discrepancies equate to different loss functions for evaluating the fit of LFM-MIX to the observations. Hence, it is not surprising to find different results based on different loss functions. This, inherently, raises the question about what discrepancy (loss function) is most appropriate to use here. This question is currently under investigation.

The fit of the statistical model identified appropriate settings for the three input values to LFM-MIX. However, the results herein are based only on the initial design described in Section 2. On-going work continues to run LFM-MIX at new design points to confirm these results and to explore new regions of the input parameter space. As computation is expensive in this application, future runs of LFM-MIX will require the use of more sophisticated sequential design techniques such as those proposed by Kleiber *et al.* (2013).

This paper used data from a single geomagnetic storm to calibrate LFM-MIX. However, a key question yet remains in how LFM-MIX reacts to different storms, i.e. would the same value for $\boldsymbol{\chi}$ be found by using data from multiple storms? In other words, are the results that are discussed in this paper specific to this storm or do they generalize to all storms? The answer to this question is beyond the scope of this paper as only the single storm was available to us. Yet, continued work in calibrating LFM-MIX will consider this question.

Acknowledgements

This work was supported by the CISWM which is funded by the Science and Technology Centers programme of the National Science Foundation under agreement ATM-0120950. This research is also supported in part by National Science Foundation grants AGS-0934488 and DMS-0707069. The National Center for Atmospheric Research is managed by the University Corporation for Atmospheric Research under the sponsorship of the National Science Foundation. This work was primarily done while M. J. Heaton was a postgraduate scientist at the National Center for Atmospheric Research. The authors thank Derek Bingham and C. Shane Reese for helpful insight.

Appendix A: Details on likelihood calculation

Let $\mathbf{S}_e$ and $\mathbf{S}_f$ denote the vector of single-fidelity LFM-MIX output of log(energy) and log(flux) stacked according to spatial locations, temporal locations and input settings, i.e. $\mathbf{S}_e$ is a $(1656 \times 18 \times 20) \times 1$ vector. Similarly, define $\mathbf{D}_e$ and $\mathbf{D}_f$ as the double-fidelity LFM-MIX output. Let $\mathbf{Y}_e$ and $\mathbf{Y}_f$ denote the vector of observations of log(energy) and log(energy $\times$ flux). Let $\mathbf{S}_e^*$, $\mathbf{S}_f^*$, $\mathbf{D}_e^*$ and $\mathbf{D}_f^*$ denote the LFM-MIX output of log(energy) and log(flux) run at the 'best' input setting $\boldsymbol{\chi}$. Finally, let $\boldsymbol{\Theta}$ denote all model parameters and $[\cdot]$ denote a general distribution function.

Given observations and the runs of LFM-MIX as described in Section 2, the likelihood for $\boldsymbol{\Theta}$ is given by

$$[\mathbf{Y}_e, \mathbf{Y}_f, \mathbf{D}_f^*, \mathbf{D}_e^*, \mathbf{S}_f^*, \mathbf{S}_e^*, \mathbf{D}_f, \mathbf{D}_e, \mathbf{S}_f, \mathbf{S}_e|\boldsymbol{\Theta}] = [\mathbf{Y}_e, \mathbf{Y}_f|\mathbf{O}^*, \mathbf{O}, \boldsymbol{\Theta}][\mathbf{O}^*|\mathbf{O}, \boldsymbol{\Theta}][\mathbf{O}|\boldsymbol{\Theta}]$$

where $\mathbf{O}^* = \{\mathbf{D}_f^*, \mathbf{D}_e^*, \mathbf{S}_f^*, \mathbf{S}_e^*\}$ is the LFM-MIX output at $\boldsymbol{\chi}$ and $\mathbf{O} = \{\mathbf{D}_f, \mathbf{D}_e, \mathbf{S}_f, \mathbf{S}_e\}$ is the set of observed LFM-MIX runs. According to equations (11) and (12), $\mathbf{Y}_e$ and $\mathbf{Y}_f$ are independent normally distributed random variables with means

$$\mathbb{E}(\mathbf{Y}_e|\mathbf{O}^*, \mathbf{O}, \boldsymbol{\Theta}) = R_\lambda(\mathbf{D}_e^*) + \boldsymbol{\delta}_e,$$
$$\mathbb{E}(\mathbf{Y}_f|\mathbf{O}^*, \mathbf{O}, \boldsymbol{\Theta}) = R_\kappa(\mathbf{D}_f^*) + R_\kappa(\mathbf{D}_e^*) + \boldsymbol{\delta}_f$$

and $\text{var}(\mathbf{Y}_e|\mathbf{O}^*, \mathbf{O}, \boldsymbol{\Theta}) = \sigma^2_{Y_e}\mathbf{I}$ and $\text{var}(\mathbf{Y}_f|\mathbf{O}^*, \mathbf{O}, \boldsymbol{\Theta}) = \sigma^2_{Y_f}\mathbf{I}$ where $\boldsymbol{\delta}_e$ and $\boldsymbol{\delta}_f$ are the vectors of discrepancy terms and $R_\lambda(\cdot)$ and $R_\kappa(\cdot)$ appropriately rotate $\mathbf{D}_e^*$ and $\mathbf{D}_f^*$ to align with the observations. Subsequently, by equations (4), (5), (9) and (10), the joint distribution $[\mathbf{O}^*|\mathbf{O}, \boldsymbol{\Theta}]$ is factored as

$$[\mathbf{D}_f^*|\mathbf{D}_e^*, \mathbf{S}_e^*, \mathbf{S}_f^*, \mathbf{O}, \boldsymbol{\Theta}][\mathbf{D}_e^*|\mathbf{S}_f^*, \mathbf{S}_e^*, \mathbf{O}, \boldsymbol{\Theta}][\mathbf{S}_f^*|\mathbf{S}_e^*, \mathbf{O}, \boldsymbol{\Theta}][\mathbf{S}_e^*|\mathbf{O}, \boldsymbol{\Theta}]. \tag{14}$$

Each conditional distribution in expression (14) is Gaussian with

$$\mathbf{D}_f^*|\mathbf{D}_e^*, \mathbf{S}_e^*, \mathbf{S}_f^*, \mathbf{O}, \boldsymbol{\Theta} \sim \mathcal{N}\{\boldsymbol{\mu}_{D_f} + \boldsymbol{\beta}_{D_f} \odot \tilde{\mathbf{S}}_f^* + \boldsymbol{\theta}_{D_f} \odot \tilde{\mathbf{D}}_e^* + (\mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1} \otimes \mathbf{I}_{18} \otimes \mathbf{I}_{1656})\tilde{\mathbf{D}}_f,$$
$$\sigma^2_{D_f}(1 - \mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1}\mathbf{r}'_{\boldsymbol{\chi}\mathbf{x}})(\mathbf{R}_\mathcal{T} \otimes \mathbf{R}_\mathcal{L})\},$$
$$\mathbf{D}_e^*|\mathbf{S}_f^*, \mathbf{S}_e^*, \mathbf{O}, \boldsymbol{\Theta} \sim \mathcal{N}\{\boldsymbol{\mu}_{D_e} + \boldsymbol{\beta}_{D_e} \odot \tilde{\mathbf{S}}_e^* + (\mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1} \otimes \mathbf{I}_{18} \otimes \mathbf{I}_{1656})\tilde{\mathbf{D}}_e, \sigma^2_{D_e}(1 - \mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1}\mathbf{r}'_{\boldsymbol{\chi}\mathbf{x}})(\mathbf{R}_\mathcal{T} \otimes \mathbf{R}_\mathcal{L})\},$$
$$\mathbf{S}_f^*|\mathbf{S}_e^*, \mathbf{O}, \boldsymbol{\Theta} \sim \mathcal{N}\{\boldsymbol{\mu}_{S_f} + \boldsymbol{\beta}_{S_f} \odot \tilde{\mathbf{S}}_e^* + (\mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1} \otimes \mathbf{I}_{18} \otimes \mathbf{I}_{1656})\tilde{\mathbf{S}}_f, \sigma^2_{S_f}(1 - \mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1}\mathbf{r}'_{\boldsymbol{\chi}\mathbf{x}})(\mathbf{R}_\mathcal{T} \otimes \mathbf{R}_\mathcal{L})\},$$
$$\mathbf{S}_e^*|\mathbf{O}, \boldsymbol{\Theta} \sim \mathcal{N}\{\boldsymbol{\mu}_{S_f} + (\mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1} \otimes \mathbf{I}_{18} \otimes \mathbf{I}_{1656})\tilde{\mathbf{S}}_e, \sigma^2_{S_e}(1 - \mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}\mathbf{R}_\mathbf{x}^{-1}\mathbf{r}'_{\boldsymbol{\chi}\mathbf{x}})(\mathbf{R}_\mathcal{T} \otimes \mathbf{R}_\mathcal{L})\},$$

where '$\odot$' denotes the Hadamard (pointwise product), $\mathbf{r}_{\boldsymbol{\chi}\mathbf{x}}$ is a vector denoting the correlations across input settings (i.e. the correlations between LFM-MIX run at $\boldsymbol{\chi}$ and those run at the observed input settings). Note that, in each conditional distribution above, the values for the covariance function were assumed to be fidelity specific (see Sections 3.1 and 3.2 for more details). The final piece of the likelihood is the joint distribution $[\mathbf{S}_e, \mathbf{S}_f, \mathbf{D}_e, \mathbf{D}_f|\boldsymbol{\Theta}]$. This distribution is factored and expressed in a similar manner to expression (14) above.

References

Andrieu, C. and Thoms, J. (2008) A tutorial on adaptive MCMC. *Statist. Comput.*, **18**, 343–373.

Apanasovich, T. V. and Genton, M. G. (2010) Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, **97**, 15–30.

Apanasovich, T. V., Genton, M. G. and Sun, Y. (2012) A valid Matérn class of cross-covariance functions for multivariate random fields with any number of dimensions. *J. Am. Statist. Ass.*, **107**, 180–193.

Banerjee, S. (2005) On geodetic distance computations in spatial modeling. *Biometrics*, **61**, 617–625.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman and Hall–CRC.

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008) Gaussian predictive process models for large spatial data sets. *J. R. Statist. Soc.* B, **70**, 825–848.

Bayarri, M. J., Berger, J. O., Kennedy, M. C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., Lin, C. H. and Tu, J. (2009) Predicting vehicle crashworthiness: validation of computer models for functional and hierarchical data. *J. Am. Statist. Ass.*, **104**, 929–943.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H. and Tu, J. (2007) A framework for validation of computer models. *Technometrics*, **49**, 138–154.

Bhat, K. S., Haran, M., Olsen, R. and Keller, K. (2012) Inferring likelihoods and climate system characteristics from climate models and multiple tracers. *Environmetrics*, **23**, 345–362.

Calder, C. A., Berrett, C., Shi, T., Xiao, N. and Munroe, D. K. (2011) Modeling space-time dynamics of aerosols using satellite data and atmospheric transport model output. *J. Agric. Biol. Environ. Sci.*, **16**, 495–512.

Cressie, N. and Johannesson, G. (2008) Fixed rank kriging for very large spatial data sets. *J. R. Statist. Soc.* B, **70**, 209–226.

Cumming, J. and Goldstein, M. (2009) Small sample Bayesian designs for complex high-dimensional models based on information gained using fast approximations. *Technometrics*, **51**, 377–388.

Finley, A. O., Sang, H., Banerjee, S. and Gelfand, A. E. (2009) Improving the performance of predictive process modeling for large datasets. *Computnl Statist. Data Anal.*, **53**, 2873–2884.

Fricker, T. E., Oakley, J. E. and Urban, N. M. (2013) Multivariate emulators with non-separable covariance matrices. *Technometrics*, **55**, 47–56.

Furrer, R., Nychka, D. and Sain, S. (2013) fields: tools for spatial data. *R Package Version 6.7.6*. (Available from `http://CRAN.R-project.org/package=fields`.)

Gelfand, A. E., Schmidt, A. M., Banerjee, S. and Sirmans, C. F. (2004) Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). *Test*, **12**, 1–50.

Gelman, A. and Rubin, D. (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.*, **7**, 457–511.

Gneiting, T. (2002) Nonseparable, stationary covariance functions for space-time data. *J. Am. Statist. Ass.*, **97**, 590–600.
Gneiting, T. (2013) Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, **19**, 1327–1349.
Gneiting, T., Kleiber, W. and Schlather, M. (2010) Matérn cross-covariance functions for multivariate random fields. *J. Am. Statist. Ass.*, **105**, 1167–1177.
Gneiting, T., Ševčíková, H. and Percival, D. B. (2012) Estimators of fractal dimension: assessing the roughness of time series and spatial data. *Statist. Sci.*, **27**, 247–277.
Gratiet, L. L. (2013) Bayesian analysis of hierarchical multi-fidelity codes. *J. Uncertn. Quantificn*, **1**, 244–269.
Guhaniyogi, R., Finley, A. O., Banerjee, S. and Gelfand, A. E. (2011) Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics*, **22**, 997–1007.
Harville, D. A. (1997) *Matrix Algebra from a Statisticians Perspective*. New York: Springer.
Higdon, D. (2002) Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues* (eds C. Anderson, V. Barnett, P. C. Chatwin and A. H. El-Shaarawi), pp. 37–56. New York: Springer.
Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008) Computer model calibration using high-dimensional output. *J. Am. Statist. Ass.*, **103**, 570–583.
Higdon, D., Swall, J. and Kern, J. (1999) Non-stationary spatial modeling. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 761–768. Oxford: Oxford University Press.
Johnson, M. E., Moore, L. M. and Ylvisaker, D. (1990) Minimax and maximin distance designs. *J. Statist. Planng Inf.*, **26**, 131–148.
Jun, M. and Stein, M. L. (2007) An approach for producing space-time covariance functions on spheres. *Technometrics*, **49**, 468–479.
Jun, M. and Stein, M. L. (2008) Nonstationary covariance models for global data. *Ann. Appl. Statist.*, **2**, 1271–1289.
Kang, E. L. and Cressie, N. (2011) Bayesian inference for the spatial random effects model. *J. Am. Statist. Ass.*, **106**, 972–983.
Kaufman, C. G., Bingham, D., Habib, S., Heitmann, K. and Frieman, J. A. (2011) Efficient emulators of computer experiments using compactly supported correlation functions, with application to cosmology. *Ann. Appl. Statist.*, **5**, 2470–2492.
Kennedy, M. C. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussion). *J. R. Statist. Soc.* B, **63**, 425–464.
Kleiber, W., Sain, S. R., Heaton, M. J., Wiltberger, M., Bingham, D. and Reese, C. R. (2013) Uncertainty quantification for a multi-fidelity dynamical model of the magnetosphere. *Ann. Appl. Statist.*, **7**, 1286–1310.
Liu, F., Bayarri, M. J. and Berger, J. O. (2009) Modularization in Bayesian analysis with emphasis on analysis of computer models. *Baysn Anal.*, **4**, 119–150.
McKay, M. D., Beckman, R. J. and Conover, W. J. (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–245.
National Research Council Committee on the Societal and Economic Impacts of Severe Space Weather Events (2008) *Severe Space Weather Events—Understanding Societal and Economic Impacts: a Workshop Report*. Washington, DC: National Academies Press.
Paulo, R., Garcia-Donato, G. and Palomo, J. (2012) Calibration of computer models with multivariate output. *Computnl Statist. Data Anal.*, **56**, 3959–3974.
Phillips, T. (2012) Solar storm dumps gigawatts into Earth's upper atmosphere. In *Science News*, Mar. 22nd. National Aeronautics and Space Administration. (Available from `http://science.nasa.gov/science-news/science-at-nasa/2012/22mar_saber/`.)
Qian, Z. G. and Wu, C. F. (2008) Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, **50**, 192–204.
Rasmussen, C. and Williams, C. K. I. (2006) *Gaussian Processes for Machine Learning*, vol. 1. Cambridge: MIT Press.
Rougier, J. C. (2008) Efficient emulators for multivariate deterministic functions. *J. Computnl Graph. Statist.*, **17**, 827–843.
Royle, J. A. and Berliner, L. M. (1999) A hierarchical approach to multivariate spatial modeling and prediction. *J. Agric. Biol. Environ. Statist.*, **4**, 29–56.
Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989) Design and analysis of computer experiments (with discussion). *Statist. Sci.*, **4**, 409–423.
Santner, T. J., Williams, B. J. and Notz, W. I. (2003) *The Design and Analysis of Computer Experiments*. New York: Springer.
Sparks, R. S. J., Aspinall, W. P., Chapman, N. A., Hill, B. E., Kerridge, D. J., Pooley, J. and Taylor, C. A. (2013) Technological facilities, infrastructure and hazardous materials. In *Risk and Uncertainty Assessment for Natural Hazards* (eds J. Rougier, S. Sparks and L. J. Hill), pp. 445–480. Cambridge: Cambridge University Press.
Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.
Stein, M. L. (2005) Space-time covariance functions. *J. Am. Statist. Ass.*, **100**, 310–321.
Sun, Y., Li, B. and Genton, M. (2011) Geostatistics for large datasets. In *Space–Time Processes and Challenges*

Related to Environmental Problems (eds E. Porcu, J. M. Montero and M. Schlather), pp. 55–77. New York: Springer.

Tarantola, A. (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: Society for Industrial and Applied Mathematics.

Wiltberger, M., Weigel, R. S., Lotko, W. and Fedder, J. A. (2009) Modeling seasonal variations of auroral particle precipitation in a global-scale magnetosphere-ionosphere simulation. *J. Geophys. Res.*, **114**, article A01204.

Zhang, H. (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Statist. Ass.*, **99**, 250–261.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Emulating and calibrating the multiple-fidelity Lyon–Fedder–Mobarry magnetosphere–ionosphere coupled computer model'.