Feature-based calibration of a global

magnetosphere-ionosphere model for geomagnetic storms

by

Victoria A. Slattum

B.S., University of Colorado at Colorado Springs, 2014

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Master of Science

Department of Applied Mathematics

2017

This thesis entitled:
Feature-based calibration of a global magnetosphere-ionosphere model for geomagnetic
storms
written by Victoria A. Slattum
has been approved for the Department of Applied Mathematics

            _____

Prof. William Kleiber

            _____

Prof. Jem Corcoran

            _____

Prof. Manuel Lladser

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both
the content and the form meet acceptable presentation standards of scholarly work in the
above mentioned discipline.

Slattum, Victoria A. (M.S., Applied Mathematics)

Feature-based calibration of a global magnetosphere-ionosphere model for geomagnetic storms

Thesis directed by Prof. William Kleiber

Geomagnetic storms play a significant role in space weather physics and have the potential to impact our daily lives. Widespread impacts of space weather physics can include power grid outages, air traffic rerouting, and disruption of GPS signals. The Lyon-Fedder-Mobarry global magnetosphereionosphere coupled model (LFM-MIX) is a computer model used at the Center for Integrated Space Weather Modeling (CISM) to study Sun-Earth interactions by simulating geomagnetic storms. LFM-MIX uses solar wind observations to perform a magnetohydrodynamic (MHD) simulation of the magnetosphere (LFM) and couples it with an electrostatic model of the ionosphere (MIX). Given a set of input parameters and solar wind data, LFM-MIX numerically solves the MHD equations and outputs a large bivariate spatiotemporal field of ionospheric energy and flux. These input parameters are unknown and we focus on quantifying them. The currently available methods are insufficient for our data set due to its high dimensionality, thus we develop our own method based on statistical calibration. Here, statistical calibration refers to the process of fitting a model to observed data by adjusting the input parameters. Our approach, which we call feature-based calibration, involves calculating some goodness of fit criterion between model output and observed data, then predicting its value over the entire feasible parameter space and locating the minimum of the predicted surface. We apply this approach to several goodness of fit criteria based on different defining features of the data.

Dedication

To Mom and Dad.

Acknowledgements

First of all I would like to thank my advisor, Will Kleiber for his patience throughout the writing process and for answering all of my questions, no matter how silly they were at the time. Thanks to my family for their love and support throughout my education. In particular, I would like to thank my dad and granddad for their financial contributions to my graduate school career. Most importantly, I want to thank my mom for encouraging me to continue pursuing my masters degree. None of this would have happened without her belief in my ability. Finally, I would like to thank Greg Oman for his emotional support and for helping me stay mostly sane throughout the writing process.

Contents

Tables

Table

Figures

Figure

Chapter 1

Introduction

Space weather is a branch of space physics concerned with conditions in the space surrounding Earth, including the magnetosphere, ionosphere, and thermosphere [18]. This differs from terrestrial weather, which occurs in Earth's atmosphere (the troposphere and stratosphere) [18]. Although space weather events occur outside of Earth's atmosphere, they still have the ability to impact our daily lives.

Strong space weather events that can be observed from Earth are caused by emissions from the Sun interacting with Earth's atmosphere [15]. The Sun is constantly emitting a solar wind composed of charged particles, which transmits the solar magnetic field through space. Sometimes when the solar wind is particularly strong it can affect Earth's magnetic field and ionosphere. It is these disturbances in Earth's magnetic field and ionosphere that trigger geomagnetic storms.

The most intense space weather storms are caused by enormous magnetic eruptions stemming from the Sun's sunspot regions. These eruptions are initiated by solar flares, or bursts of electromagnetic radiation composed of X-rays and ultraviolet light. If these X-rays come in contact with Earth's ionosphere they can cause the ionosphere to temporarily lose its ability to reflect long-range radio waves, which is known as a Radio Blackout. In addition to causing Radio Blackout, the X-rays and ultraviolet radiation heat the outer part of Earth's atmosphere, which causes it to expand and increases the drag on Earth-orbiting satellites, especially those in high geosynchronous orbit, thus reducing their lifetime in orbit [9]. These

changes in the atmosphere, along with the intense radio emissions from solar flares, can also decrease the precision of GPS measurements [9].

Many of these huge magnetic eruptions also render a huge cloud of plasma and concentrated solar magnetic field, which is launched through space at millions of miles per hour. This cloud is known as a coronal mass ejection, or CME. It is the interaction between a CME and Earth's magnetosphere that causes the largest magnitude geomagnetic storms. It typically takes a CME 3 to 5 days to reach the Earth, triggering a geomagnetic storm which can last somewhere between a few hours and a few days. The strongest geomagnetic storms have been known to persist for up to a week [15].

CMEs are typically associated with solar flares, but they can also occur independent of solar flares. Both CMEs and solar flares occur more frequently during the active phase of the solar cycle, which lasts approximately 11 years. The most recent maximum in solar activity occurred in April 2014 [9].

The most serious effects of space weather on human activity occur during major geomagnetic storms [9]. In addition to damaging Earth-orbiting satellites and causing Radio Blackout, geomagnetic storms have also been known to cause temporary loss of electricity over large regions. The high currents in the magnetosphere translate to high currents in the power lines, which can cause electric transformers and power stations to blow out. This typically happens at high latitudes, where the induced currents are the greatest, and in regions with long power lines, where the ground is poorly conducting [9]. The most well-known instance of this occurred in Quebec in 1989, when the entire province lost power for nine hours [11].

In addition to the numerous negative impacts geomagnetic storms have on our lives, they have one positive effect. From an aesthetic point of view, geomagnetic storms enhance the Earth's auroras, causing them to appear brighter and move closer to the equator. See Figure 1.1.

Figure 1.1: Aurora australis over Earth's atmosphere viewed from the International Space Station (ISS) [4].

Human interest in space weather began in 1724, when the English geophysicist George Graham noticed that the needle of his magnetic compass was repeatedly deflected from magnetic north throughout the day. In 1889 English physicist Arthur Schuster eventually proved, via analysis of magnetic observatory data, that this was due to electrical currents flowing through the ionosphere and magnetosphere. With the rise of technology over the last two centuries, the study of space weather has drastically changed since the 1800's. Today's state-of-the-art computer models can do more than just track the location of a geomagnetic storm, they can simulate and even predict future storms. In this thesis we focus on one model in particular, the LFM-MIX.

LFM-MIX is a global magnetospheric-ionospheric coupled simulation model used at

the Center for Integrated Space Weather Modeling (CISM) to study interactions between the Sun and Earth via simulation of geomagnetic storms. LFM-MIX takes solar wind and magnetic field data as input, along with a set of unknown input values, and solves the magnetohydrodynamic equations, producing a large spatiotemporal field of ionospheric energy and flux. Our goal is to determine the optimal set of unknown input values which results in the most accurate prediction of the field observations. In computer experiments literature, this process of tuning physical model parameters to observations is known as an inverse problem, or sometimes is referred to as a calibration problem [11].

The layout of this thesis is as follows. In Chapter 2, we describe the LFM-MIX model and our dataset in further detail. In Chapter 3 we discuss calibration in further detail, present methods employed in the literature and discuss challenges specific to this problem, such as the issues introduced by the dimensionality of the data and the uncertainty in the measurement of the observed geomagnetic storm. We end the chapter by presenting our method of feature-based calibration. Chapter 4 is where we introduce several statistics gathered from the data and apply feature-based calibration. Then, in Chapter 5 we validate the results from Chapter 4 by implementing leave one out cross validation. We hold out each model as a new set of observations, then calibrate the remaining models to the new observations and compare the estimated parameter values to the true values to determine the accuracy of our method. Finally, in Chapter 6 we conclude by analyzing our results and making suggestions for future work.

Chapter 2

Data and LFM-MIX Model

The LFM-MIX is a magnetospheric-ionospheric coupled model used for simulation of geomagnetic storms. It is made up of two standalone models, the Lyon Fedder Mobarry global MHD model (LFM) and the Magnetosphere Ionosphere Coupler/Solver (MIX).

2.1   LFM

The heart of the LFM model is a time-dependent calculation of the state of the magnetosphere via numerical solution of the ideal magnetohydrodynamic (MHD) equations. Typically, the LFM model is tightly coupled to a realistic model of the polar ionospheres (such as the MIX model), and requires only solar wind plasma and magnetic field data. The LFM model is structured with the following essential design considerations in mind: high resolving power transport, maintaining $\nabla B = 0$ and low β, and using an adapted grid [20].

High resolving power transport refers to the plasma transport used in LFM, which is designed to fulfill two goals: to avoid false extrema upon calculation and to provide the highest resolution possible. Maintaining $\nabla B = 0$ (up to roundoff error) is also necessary for the LFM model, as nonzero values of ∇B can cause the magnetic stress tensor to yield nonphysical magnetic forces. In addition to ensuring $\nabla B = 0$, the LFM model must also keep a low β value. β represents the ratio of plasma pressure to the magnetic pressure, and is generally small in the inner magnetosphere. The choice of grid used in the LFM code is the final crucial aspect of the model's design. It uses a distorted spherical grid where the

(r, θ) coordinate system is tailored to the problem at hand. The azimuthal coordinate ϕ, on the other hand, is not modified for the specific problem; instead, it is stretched slightly near the equator and compressed slightly near the poles to give a somewhat more uniform mapping of the magnetospheric grid in the ionosphere [14]. This process occurs in a separate program prior to initialization of the LFM code.

Although LFM is currently the only model that can yield a reasonably accurate simulation of the magnetosphere-ionosphere system, the model suffers from a few limitations. For instance, LFM assumes the ideal MHD equations for the magnetospheric plasma, which do not take into account various significant processes, such as inner magnetospheric drifts and microscopic reconnection physics [20]. This lack of drift physics causes an absence of trapped ring current, which is known to have a considerable effect on the electrodynamics of geomagnetic storms [13]. Furthermore, LFM treats the auroral acceleration region in a simplified manner and does not take into account ionospheric outflow [20].

In this thesis we focus on the LFM-MIX coupled model, but the LFM model can be coupled with other models to simulate different aspects of Earth's atmosphere. For instance, the Coupled Magnetosphere Ionosphere Thermosphere Model (CMIT), which is composed of LFM, MIX, and the Thermosphere Ionosphere Electrodynamic General Circulation Model (TIEGCM), is used to simulate Earth's magnetosphere-ionosphere-thermosphere system. Coupling LFM with the Rice Convection Model (RCM) instead of MIX also simulates Earth's magnetosphere-ionosphere system, but trades some strengths of LFM-MIX to improve upon weaknesses in the LFM model. For example, LFM-RCM improves upon LFM by taking into account the trapped ring current [24].

2.2   MIX

Although the main function of the Magnetosphere Ionosphere Coupler/Solver (MIX) code is to determine the inner boundary conditions for global MHD models, such as LFM, it has other useful applications. When in solver mode, MIX can use an existing empiri-

cal model of ionospheric conductance as input to solve for the electrostatic potential. Or, when in solver/coupler mode, it can take conductance inputs from an external ionosphere-thermosphere coupled model and calculate the potential [21]. Currently, work is being done to use MIX in communicating between other geospace models, such as RCM and CMIT.

2.3   Data Set

The data set under consideration in this thesis arises from the LFM-MIX model. Taking solar wind, initial magnetic field strength, and level of ultraviolet light from the Sun as boundary conditions, LFM solves the MHD equations and passes on information about the energy and number flux (henceforth referred to as flux) of the electrons precipitating into the ionosphere to the MIX model. In the MIX model, three physical equations define the energy and flux inputs:

$$\varepsilon_0 = \alpha c_s^2, \quad f_0 = \beta \rho \sqrt{\varepsilon_0}, \quad \varepsilon_\| = \frac{R J_\| \sqrt{\varepsilon_0}}{\rho} \tag{2.1}$$

where ε_0 is energy, f_0 is flux, c_s^2 is the speed of sound, ρ is the density of the innermost cells in the magnetospheric grid, $\varepsilon_\|$ represents the field aligned electric potential energy difference, and $J_\|$ represents the upward field aligned current. α, β, and R are unknown tuning parameters designed to account for physical processes outside the scope of LFM-MIX. α accounts for the effects of calculating electron temperature from the single fluid temperature, β adjusts for possible plasma anisotropy and controls a loss filling cone, and R allows for scaling of the parallel potential drop based on the sign of the current and accounts for the possibility of being outside the regime of the scaling [11]. Here we seek to identify the optimal setting of $\theta = (\alpha, \beta, R)$ such that some goodness of fit criterion is minimized.
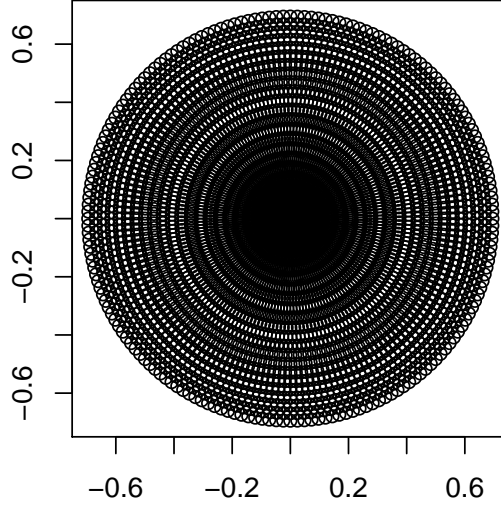
Figure 2.1: Grid of spatial locations where the LFM-MIX model was run.

Our observational data set is a bivariate spatiotemporal field of energy (in keV) and flux (in $\frac{1}{cm^2s}$) from a geomagnetic storm which occurred from October 21-22, 2001. Energy and flux were measured approximately every 3 minutes, between 1:43am (October 21) and 11:59pm UTC (October 22) (815 time points total), on a subset of the 26×180 grid of 4680 spatial locations in Figure 2.1. The subset of the spatial grid where the observations are defined changes for each time point, which further complicates our analysis. In terms of latitudes and longitudes, the spatial grid spans $46°$ to $90°$ latitude (with $46°$ latitude corresponding to the outer circle of grid points) and all longitudes (in steps of $2°$ from $0°$ to $358°$) [6]. Our model data comes from 20 runs of LFM-MIX at different input settings (α, β, R) in the hyperrectangle of physically feasible values: $[0, 0.5] \times [0, 2.5] \times [0, 0.1]$. To simplify our analysis, we normalize the (α, β, R) values, resulting in the hyperrectangle $[0, 1] \times [0, 1] \times [0, 1]$ as the new grid for calibration. Each LFM-MIX run contains energy and flux data defined on the same spatiotemporal grid as the observations. Figures 2.2 and 2.3 show examples of the observed energy and flux, respectively, for time $t = 114$ and three

different model outputs for the same time.

We notice that the flux models are several orders of magnitude larger than the observed flux, while observed energy and models are on a similar order of magnitude. In Chapter 4 we will introduce a multiplicative bias to the models in hopes of correcting this. We can see from Figure 2.3 that the flux models also need to be rotated. In Chapter 4 we will also introduce two different rotational biases, one constant and one time-varying.
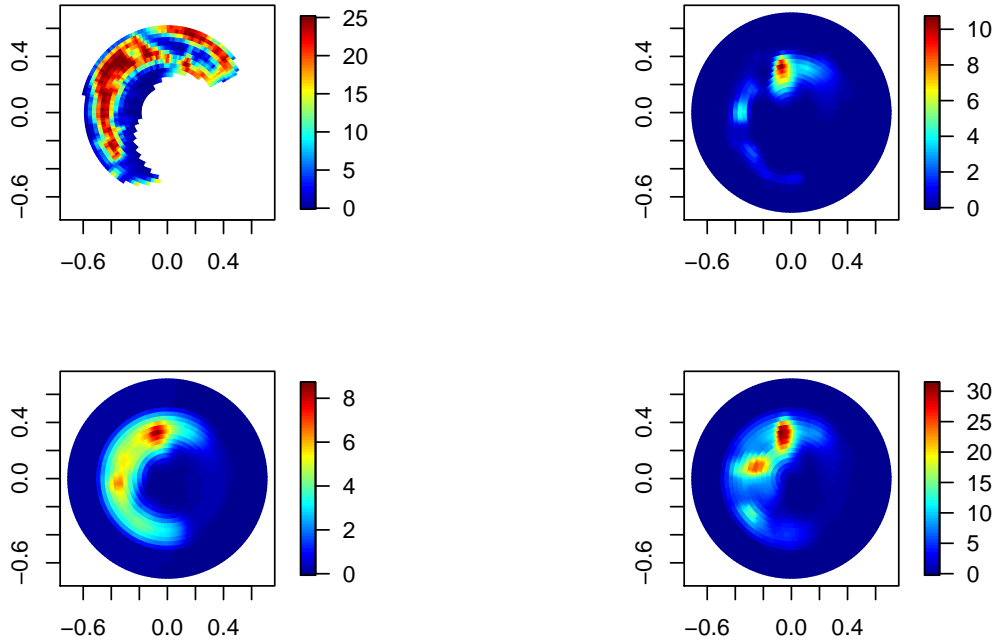
Figure 2.2: Top left: Observed energy at time $t = 114$. Top right: LFM-MIX energy output for $\theta = (0.39, 1.21, 0.08)$. Bottom left: LFM-MIX output for $\theta = (0.18, 0.17, 0.01)$. Bottom right: LFM-MIX output for $\theta = (0.16, 2.21, 0.05)$.
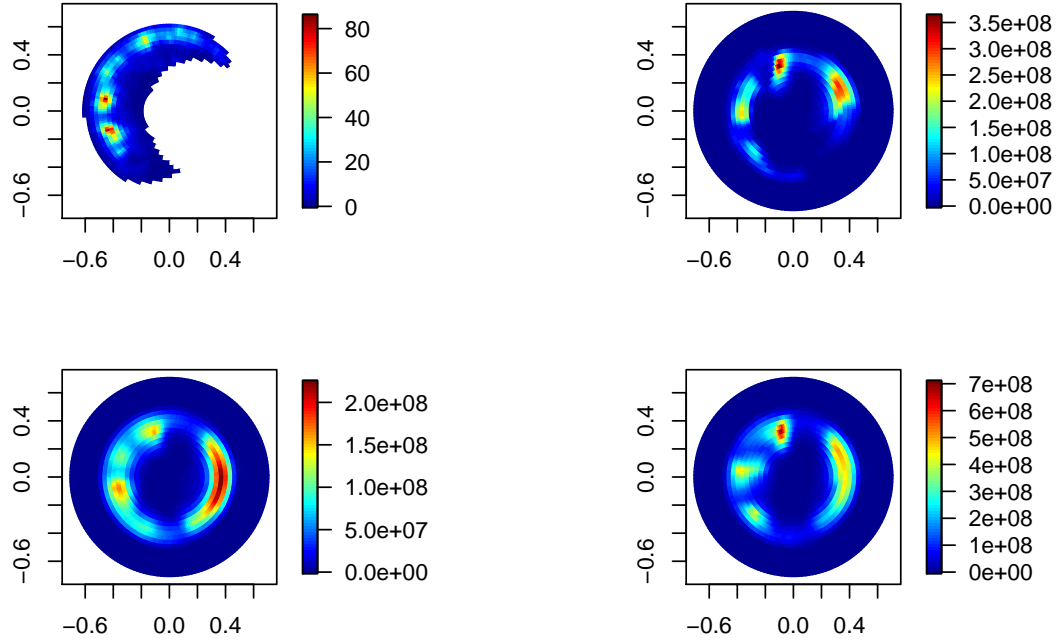
Figure 2.3: Top left: Observed flux at time $t = 114$. Top right: LFM-MIX flux output for $\theta = (0.39, 1.21, 0.08)$. Bottom left: LFM-MIX output for $\theta = (0.18, 0.17, 0.01)$. Bottom right: LFM-MIX output for $\theta = (0.16, 2.21, 0.05)$

Before moving on to the next chapter we need to introduce some notation. We will let S denote the 26×180 grid of spatial locations in Figure 2.1 and T represent the set of 815 time points where the storm was observed. The observed energy will be denoted by $E_0(s, t)$ and the observed flux by $F_0(s, t)$, for $s \in S$ and $t \in T$. We define θ to be the set of (α, β, R) values corresponding to each model run, so $\theta = \{(\alpha_1, \beta_1, R_1), \ldots, (\alpha_{20}, \beta_{20}, R_{20})\}$ and denote the parameters corresponding to model j by $\theta_j = (\alpha_j, \beta_j, R_j)$. Then the energy and flux for model j are denoted by $E_{\theta_j}(s, t)$ and $F_{\theta_j}(s, t)$, respectively. And finally, the optimal input setting we are trying to identify will be denoted by $\theta_0 = (\alpha_0, \beta_0, R_0)$.

Chapter 3

Method: Feature-based Calibration

In this chapter we introduce statistical calibration, which provides the basis of our approach, then present a brief overview of different approaches to calibration found in the literature. Finally, we discuss several challenges presented by our data set and end the chapter by developing our method of feature-based calibration.

3.1   Calibration

Calibration is a term with many different meanings, depending on the field of study. Scientific calibration, for example, is the process of determining the scale of a measuring instrument based on an informative calibration experiment. While similar to scientific calibration, a statistical calibration problem has a more complicated form. A statistical calibration problem is a type of inverse prediction, which some authors refer to as inverse regression, rather than calibration. To better explain a statistical calibration problem, we borrow a univariate example from [17]: Suppose there is a chemist who wants to establish a calibration curve for measuring the amount of a certain chemical A in samples sent to the laboratory. There are two stages in the calibration process: experiment and prediction. In the first stage, each of the samples with known amounts of chemical A is measured again, this time with the relatively quick, inexpensive *test method* we wish to calibrate. The samples with known amounts of chemical A have been measured by an extremely accurate *standard method*, which is slow and expensive to implement. The resulting data is used to estimate the calibration

curve f, ending the first stage of the calibration process. In the prediction stage, the test method is used to analyze samples with unknown amounts of chemical A, and the amount of chemical A is predicted for each sample.

Statistical calibration problems are separated into two classes: absolute and comparative calibration. It is important to distinguish between the two types because, although they are both called calibration, they are different concepts and lead to different issues in statistical modelling [17]. In absolute calibration, a quick or non-standard measurement technique is calibrated against the standard technique, which is known to be accurate. The chemist example above is an absolute calibration problem. Comparative calibration, on the other hand, is the process by which one instrument or measurement technique is calibrated against another, with neither being the standard technique. In this thesis, we will focus on absolute calibration.

3.2   Literature Review

Before we discuss the challenges presented by our data set, we outline a few popular approaches found in the literature. We begin with the general approach of Kennedy and O'Hagan in [10].

In their classic paper *Bayesian calibration of computer models*, Kennedy and O'Hagan introduce the method of Bayesian calibration. The basic idea behind their approach is to use the observed data to derive the posterior distribution of the unknown input parameters, θ. Then, a predictive distribution is obtained to quantify the 'residual uncertainty' about θ. Their approach treats the computer model as a 'black box', meaning that it does not make use of any underlying information about the mathematical model, except for the relationships between the inputs and outputs. They employ the following statistical model:

$$z_i = \zeta(\mathbf{x_i}) + e_i = \rho\eta(\mathbf{x_i}, \theta) + \delta(\mathbf{x_i}) + e_i$$

where z_i are the observations, $\zeta(\mathbf{x_i})$ is the true value of the real process when the known inputs take values $\mathbf{x_i}$, e_i is the observation error for the ith observation, ρ is an unknown

regression parameter, $\eta(\mathbf{x_i}, \theta)$ is the model output given $\mathbf{x_i}$ and θ as inputs, and $\delta(\cdot)$ is a model inadequacy function that is independent of $\eta(\cdot, \cdot)$ [10].

Next, we describe the approach of Kleiber et al. in [11], which was developed with the LFM-MIX model in mind. Although they also focus on the LFM-MIX model, their data set differs from ours in that they have model data at two different resolutions. The main idea behind the approach of Kleiber et al. is to utilize an empirical orthogonal function (EOF) decomposition of the low resolution spatial field and an EOF decomposition of the discrepancy function which links the high and low resolutions of the model. We discuss EOF analysis in further detail in the next section.

Following the approach of Kennedy and O'Hagan above, Kleiber et al. exploit the following statistical model:

$$Y(\mathbf{s}, t) = H(\mathbf{s}, t, \theta) + \varepsilon(\mathbf{s}, t)$$

where $Y(\mathbf{s}, t)$ denotes observed energy at spatial location $\mathbf{s}$ and time t; $H(\mathbf{s}, t, \theta)$ is the high resolution model output at location $\mathbf{s}$, time t, and optimal input parameters θ; and $\varepsilon(\mathbf{s}, t)$ is measurement error, which is normally distributed with mean zero and variance τ^2. Their approach differs from that of Kennedy and O'Hagan in that they do not employ a model discrepancy term. The reasons behind this decision will be discussed in section 3.4.3, where we develop our own statistical model. Finally, the additive discrepancy function $\delta(\mathbf{s}, t, \theta)$ is used to link the high and low resolution model data:

$$H(\mathbf{s}, t, \theta) = L(\mathbf{s}, t, \theta) + \delta(\mathbf{s}, t, \theta)$$

Finally, we end our literature review by briefly describing the approach of Heaton et al. in [6]. Their analysis builds on and extends that of Kleiber et al. by considering both energy and flux data and including discrepancy between the model output and observations. They note that exploratory analysis suggests a linear relationship between energy and flux data for the low resolution model, thus we arrive at the following relationships for the low resolution model output:

$$L_e(\mathbf{s}, t, \theta) = \mu_{L_e}(\mathbf{s}, t) + w_{L_e}(\mathbf{s}, t, \theta)$$

and

$$L_f(\mathbf{s}, t, \theta) = \mu_{L_f}(\mathbf{s}, t) + \beta_{L_f}(\mathbf{s}, t)w_{L_e}(\mathbf{s}, t, \theta) + w_{L_f}(\mathbf{s}, t, \theta)$$

where $\mu_{L_e}(\mathbf{s}, t)$ and $\mu_{L_f}(\mathbf{s}, t)$ represent the average energy and flux at spatial location $\mathbf{s}$ and time t, $\beta_{L_f}(\mathbf{s}, t)$ is a spatiotemporal slope, and $w_{L_e}(\mathbf{s}, t, \theta)$ and $w_{L_f}(\mathbf{s}, t, \theta)$ are independent Gaussian random effects.

Heaton et al. also note that exploratory analysis suggests a linear relationship between the low and high resolution models that varies over the spatiotemporal domain. Thus, the high and low resolution model outputs are related as follows:

$$H_e(\mathbf{s}, t, \theta) = \mu_{H_e}(\mathbf{s}, t) + \beta_{H_e}(\mathbf{s}, t)w_{L_e}(\mathbf{s}, t, \theta) + w_{H_e}(\mathbf{s}, t, \theta)$$

and

$$H_f(\mathbf{s}, t, \theta) = \mu_{H_f}(\mathbf{s}, t) + \beta_{H_f}(\mathbf{s}, t)w_{L_f}(\mathbf{s}, t, \theta) + \theta_{H_f}(\mathbf{s}, t)w_{H_e}(\mathbf{s}, t, \theta) + w_{H_f}(\mathbf{s}, t, \theta)$$

where $\theta_{H_f}(\mathbf{s}, t)$ is a spatiotemporal slope field.

3.3   Challenges

As previously mentioned, our data set presents several challenges which render currently available methods ineffective. First of all, the models and observations are very highly multivariate spatiotemporal fields; with only one storm and 20 LFM-MIX runs we have over 76,284,000 correlated points (4680 grid locations for the 20 model runs at 815 time points, plus the observed storm at a subset of the 4680 grid locations over 815 time points). The traditional approach of Kennedy and O'Hagan in [10] is very difficult to implement for large spatiotemporal data sets; in our case it would require inverting a covariance matrix of dimension $76,284,000 \times 76,284,000$, which is far beyond our computing capabilities.

In addition to the high dimensionality of our data, uncertainty in the measurement of the observations presents further challenges. Upon analysis of total energy (which represents the amount of energy in the storm at each time point) in chapter 4, we notice in Figure 4.5 that the observations and models have similar amounts of energy at the beginning of the storm, but later on the models all show two large spikes in energy, where the observations

do not. There are two possible reasons for this behavior: the LFM-MIX model is not very accurate or the observations were not measured very accurately. The LFM-MIX has been a very popular method of simulating the magnetosphere-ionosphere system for some time and is known to be very realistic, thus we suspect this behavior is due to uncertainty in measurement of the observations. Furthermore, we already know that the observations were measured via satellite which has a limited field of vision, (this is why the spatial domain where the observations are defined changes over time), thus it is not unreasonable to assume that even the measurements over the observed spatial domain are not entirely accurate. In other words, we suspect that Earth's curvature prevents the satellite from observing the entire storm. For example, in Figure 1.1 the part of the storm barely visible on the horizon is likely not recorded.

Following the approach of Kleiber et al. in [11], we attempt empirical orthogonal function (EOF) analysis to reduce dimensionality of our data. EOF analysis involves decomposing the data set in terms of orthogonal basis functions, which are computed from the data. The ith basis function is chosen to be orthogonal to the first $i - 1^{st}$ and to account for as much variance as possible. It is very similar to principal components analysis (PCA), except that EOF analysis finds both time series and spatial patterns, while PCA only finds spatial patterns [23]. First, the data is arranged into a matrix so that each row represents a particular time point and the columns are the spatial data points. Next, a singular value decomposition is performed, resulting in $M = U \cdot D \cdot V^t$, where M is the data matrix, the columns of U are the EOFs, the columns of V are the principal components, and the diagonal values of D represent the amplitudes of the EOFs. However, it is unclear how to proceed with this process in our case, due to the irregular grids on which the observed data are defined. To construct the data matrix M, we would need the same number of spatial points for each time point, but since our observations are defined on different grids for each time point, we have a different number of spatial points for each time point. Thus, we quickly abandon the EOF analysis approach.

3.4 Feature-based Calibration

After providing sufficient background on statistical calibration, we are ready to introduce the main focus of this thesis, our method of feature-based calibration. The following is a general outline of the process:

(1) Generate statistics based on energy and flux data

    (a) Calculate total energy and flux, introduce rotational and multiplicative biases, etc. See Chapter 4 for further details.

(2) Calculate SSE between models and observations for each statistic

(3) For each statistic, implement test method we wish to calibrate

    (a) Calculate empirical variogram based on the SSEs for each statistic

    (b) Fit Gaussian variogram model to the empirical variogram

    (c) Calculate the covariance function using the parameters from the fitted Gaussian variogram model

    (d) Krige statistics over 3D normalized parameter space

        (i) Construct a confidence interval for each statistic at the minimum of the kriged surface

(4) Validate the test method

    (a) Perform leave-one-out cross validation experiment for energy and flux

In this section, we describe the test method in full detail. The statistics for calibration are generated in Chapter 4, and the test method is applied. Finally, we conclude the calibration process in Chapter 5, where we discuss validation of the test method. Here, we note that our method differs slightly from the traditional method of absolute calibration in that

we hold out model data for validation, rather than validating on new, unseen data. This difference is simply due to new data being expensive and time-consuming to acquire.

Our test method is implemented as follows: For each statistic, we measure the sum of squared errors (SSE) between the observed values and each model. This results in 20 measures of error for each statistic, which are located on the three dimensional grid of normalized θ values for the model runs. This grid will henceforth be referred to as the *normalized parameter space*. Next, we predict the SSE values at new locations in the normalized parameter space (where we didn't run the LFM-MIX model) and find the minimum of the predicted surface. Prediction of the SSE values at new locations is performed with a very popular geostatistical method known as kriging.

3.4.1   Kriging

Kriging is an interpolation method developed by French mathematician Georges Matheron in 1960, based on the 1951 Master's thesis of Danie G. Krige, after whom the method is named. Kriging, also known as Gaussian process regression or Wiener-Kolmogorov prediction, yields an interpolation function based on a covariance or variogram model derived from the data, rather than a theoretical model of the interpolating function. Since kriging is based on prior covariances it is able to yield the best linear unbiased prediction (under suitable assumptions on the priors).

The basic idea of kriging is to predict the value of a random function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point. Kriging is similar to prediction in time series: given past values, typically measured at regular intervals, predict the signal at some future time. To analyze the signal, first the spectrum is modeled then a filter, or predictor, is constructed. The main difference between the kriging approach and the time series approach is that kriging takes place in a spatial setting, rather than a temporal setting, where there is no concept of past or future. There are two other differences between kriging and prediction in time series that we need

to address. In kriging, the data is typically irregularly spaced, which renders many of the commonly used methods in digital signal processing ineffective. Second, kriging requires a continuous, rather than discrete, location indexing space.

Before describing the mechanics of kriging, we must first introduce some notation. Let $\{Z(x) : x \in D \subset \mathbb{R}^n\}$ denote a random function used as a model for the variable of interest $\{z(x) : x \in D \subset \mathbb{R}^n\}$, which represents reality. In other words, $z(x)$ is a realization of $Z(x)$. We will denote the set of points where $Z(x)$ has been sampled by $S = \{x_\alpha : \alpha = 1, \ldots, N\}$. And the kriging estimator of $Z(x_0)$ will be denoted by $Z^*(x_0)$. Finally, to make our notation simpler, we will represent values of functions at sample points by the subscripts of the points. For example, $Z_\alpha = Z(x_\alpha)$, $m_\alpha = m(x_\alpha)$ will represent the mean value of Z_α, $\sigma_{\alpha\beta} = \sigma(x_\alpha, x_\beta)$ will represent the covariance between Z_α and Z_β.

Depending on the stochastic properties of the random field under investigation, we can derive different methods for calculating the weights, thus resulting in different types of kriging. Three of the most popular methods are ordinary, simple, and universal kriging. Let x_0 denote a spatial location where we wish to estimate the value of the random function, $Z(x_0)$. Ordinary kriging assumes an unknown constant mean over the neighborhood of x_0. Simple kriging assumes a known mean and covariance, and a constant and known first moment over the entire domain. And universal kriging assumes the mean follows a general polynomial trend. In this thesis, we focus of the method of simple kriging.

For simple kriging, the estimator Z^* is defined as follows:

$Z^*(x_0) = m_0 + \sum_\alpha \lambda_\alpha (Z_\alpha - m_\alpha)$

where λ_α is the weight placed on Z_α. And the kriging weights are given by the *simple kriging system* below:

$$\sum_\beta \lambda_\beta \sigma_{\alpha\beta} = \sigma_{\alpha 0}, \qquad \alpha = 1, 2, \ldots, N$$

In matrix notation, this is:

$$\boldsymbol{\Sigma}\lambda = \sigma_{\mathbf{0}}$$

where $\boldsymbol{\Sigma} = [\sigma_{\alpha\beta}]$ is the $N \times N$ matrix of covariances between the data, $\lambda = [\lambda_\alpha]$ is the

N-vector of kriging weights, and $\sigma_\mathbf{0} = [\sigma_{\alpha 0}]$ is the N-vector of covariances between the data and the target point, x_0. Finally, we define the *simple kriging variance* as follows:

$$V(x_0) = \sigma_{00} - \lambda^T \sigma_\mathbf{0}$$

where σ_{00} is the variance of Z_0. We will use the kriging variance to construct a confidence interval for the simple kriging estimator at the minimum, Z_0^* as follows:

$$[Z_0^* - 2\sqrt{V_0}, Z_0^* + 2\sqrt{V_0}]$$

3.4.2   Covariance Functions and Variograms

As previously mentioned, we need to know the covariance function in order to apply simple kriging. Simple kriging assumes that the covariance function and variogram are both stationary and isotropic. In this context, stationary means that the random process only depends on the separation of any two points, rather than their actual locations. And here, isotropic means that the random process only depends on the Euclidean distance (and not the direction) between the points. Since we are assuming that our random process is stationary and isotropic (together known as homogeneous), we can use one of the common covariance models, such as the constant, linear, exponential, or Gaussian. Here we choose the Gaussian covariance model, which is defined as follows:

$$C(h) = c \cdot \exp\left(-\frac{h^2}{a^2}\right), \qquad a, c > 0$$

where a and c are parameters derived from the variogram.

The theoretical variogram, denoted by $2\gamma(x, y)$, is a function describing the spatial dependence of a spatial random field or stochastic process. It is defined as the variance of the difference between field values at two locations across realizations of the field. Mathematically, this is:

$$2\gamma(x, y) = var(Z(x) - Z(y)) = E[((Z(x) - \mu(x)) - (Z(y) - \mu(y)))^2].$$

In the case of simple kriging we assume a constant mean μ, and this becomes:

$$2\gamma(x, y) = E[(Z(x) - Z(y))^2].$$

$\gamma(x, y)$ itself is known as the semivariogram, but is typically just called the variogram in

the literature. Following this convention, henceforth the term variogram will refer to the semivariogram $\gamma(x, y)$. To estimate the variogram, we calculate the sample or empirical variogram as follows:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} |z_i - z_j|^2$$

where z_i is the observed value at location x_i and $N(h)$ is the set of pairs of observations i, j such that $d(x_i, x_j) = h$ (within a given tolerance), where $d(x_i, x_j)$ is the Euclidean distance between spatial locations x_i and x_j.

In order to apply simple kriging, we need a variogram that is conditionally negative definite. In other words, for all weights $w_1, \ldots, w_n$ such that $\sum_{i=1}^{n} w_i = 0$, we require $\sum_{i=1}^{n} \sum_{j=1}^{n} w_i \gamma(x_i, x_j) w_j \leq 0$. Due to variability in the estimation of the empirical variogram, it may not fulfill this requirement. To fix this, we approximate the empirical variogram with a model function that ensures validity of the variogram. There are many different model functions to choose from, such as exponential, spherical, and Gaussian. In this thesis we make exclusive use of the Gaussian variogram model, defined as follows:

$$\gamma(h) = (c - n)\left[1 - \exp\left(-\frac{h^2}{a^2}\right)\right]$$

where c, n, and a are tuning parameters and h is the lag increment. c represents the *sill*, which is the limit of the variogram as h goes to infinity. n is the *nugget*, which is the value of the variogram at the origin. a is the *range*, which is just the distance at which the difference between the variogram and the sill becomes negligible.

Before moving on to the next subsection, we need to discuss selection of the lag increment h in the empirical variogram. The lag increment h controls the distances at which the empirical variogram is calculated. If h is too small the variogram will be too noisy, but if h is too large the variogram won't include enough points, which will make it too smooth. There is no rigorous method of selecting h in the literature, the ideal procedure depends on

the problem at hand. For our particular problem we assume that the nugget value is zero, thus an h that minimizes the nugget is desirable. We assume a zero nugget value because if we run LFM-MIX at the same input settings multiple times we will always get the same energy and flux output, resulting in the same SSEs; thus the variogram at zero lag should be zero. The nugget value is not the only deciding factor in our selection of h. We also require an empirical variogram that increases over a somewhat large range. Obviously the variogram needs to be increasing, since points that are closer together should be more similar than points that are farther apart. We also require a large range because a variogram that increases to its sill very quickly does not show as much spatial correlation as one that takes a long time to reach its sill.

3.4.3   Statistical Model

Finally, we are ready to introduce the statistical model for our problem. Actually, since we are examining several different statistics, we will have a different model for each statistic; however, they are all of the same general form. Following the approach of Kleiber et al., we use a method popularized by Kennedy and O'Hagan in [10]. We suppose there is an unknown setting, θ_0, for which the LFM-MIX model output is an adequate representation of reality. In particular, for the observations, $Y(s, t)$, at grid point s and time t, we have:

$$F(Y(s, t)) = F(M(s, t, \theta_0)) + \varepsilon(s, t)$$

where $M(s, t, \theta_0)$ is the LFM-MIX model output at the optimal input setting θ_0, $F(M(s, t, \theta_0))$ represents a particular statistic generated from the LFM-MIX model output, $F(Y(s, t))$ is the same statistic generated from the observed geomagnetic storm data, and $\varepsilon(s, t)$ denotes measurement error (which we assume to be normally distributed with mean zero and variance τ^2). For example, $F(\cdot)$ for rotated energy would be defined as follows:

$$F(M(s, t, \theta_0)) = R(\hat{r})M(s, t, \theta_0)$$

where $R(\hat{r})$ is a clockwise rotation matrix given by $R(\hat{r}) = \begin{bmatrix} \cos(\hat{r}) & \sin(\hat{r}) \\ -\sin(\hat{r}) & \cos(\hat{r}) \end{bmatrix}$

and $\hat{r}$ is the rotation that minimizes some measure of error between the observations and rotated models. Like Kleiber et al., we also do not include a model discrepancy term. There are two reasons for this: model discrepancy is not part of the traditional setup for large-scale inverse problems such as ours and since we have only one geomagnetic storm, and any model bias term would be confused with $\varepsilon(s, t)$ without severe simplifying assumptions [11].

Chapter 4

Application of Method

In this chapter we introduce several statistics gathered from the data, then apply our method of feature-based calibration to each one and present the results. Table 4.1 and Figures 4.15-4.17 in Section 4.13 summarize the results for each statistic. Then, in Section 4.14 we introduce a sigmoid weight function to the statistics in hopes of improving upon the results of Section 4.13.

4.1   Energy

The first, and simplest, statistic we will consider is the energy SSE, denoted by $SSE(E_\theta)$. We define the energy SSE for model j as follows:

$SSE(E_{\theta_j}) = \sum_{t \in T} \sum_{s \in S} (E_0(s,t) - E_{\theta_j}(s,t))^2$. Upon calculation of the energy SSE we notice that model 8, corresponding to $(\alpha, \beta, R) = (0.180, 0.169, 0.005)$ minimizes $SSE(E_\theta)$. Now we initiate the calibration process by computing the empirical variogram for $SSE(E_\theta)$. Using a bin width of 0.22, we find that the empirical variogram has an increasing trend and a small nugget value, as desired (see the black line in Figure 4.1). Next, we fit a Gaussian variogram model to the empirical variogram. To do this, we computed the Gaussian model at $c = a = \{0.1, 0.2, \ldots, 10.0\}$, and found that using $c = 0.1$ and $a = 0.2$ minimizes the SSE between the empirical and fitted variograms (see Figure 4.1).

Now that we have established an appropriate variogram model, we can finish calibration of $SSE(E_\theta)$ by kriging the SSE values over the grid of equally spaced locations in the three

dimensional normalized parameter space $[0,1] \times [0,1] \times [0,1]$. Figures 4.2-4.4 show 2D slices of the predicted field at the minimum of the normalized parameter space, which corresponds to the point $(0.39, 1.35, 0.06)$ in the original (unnormalized) parameter space.
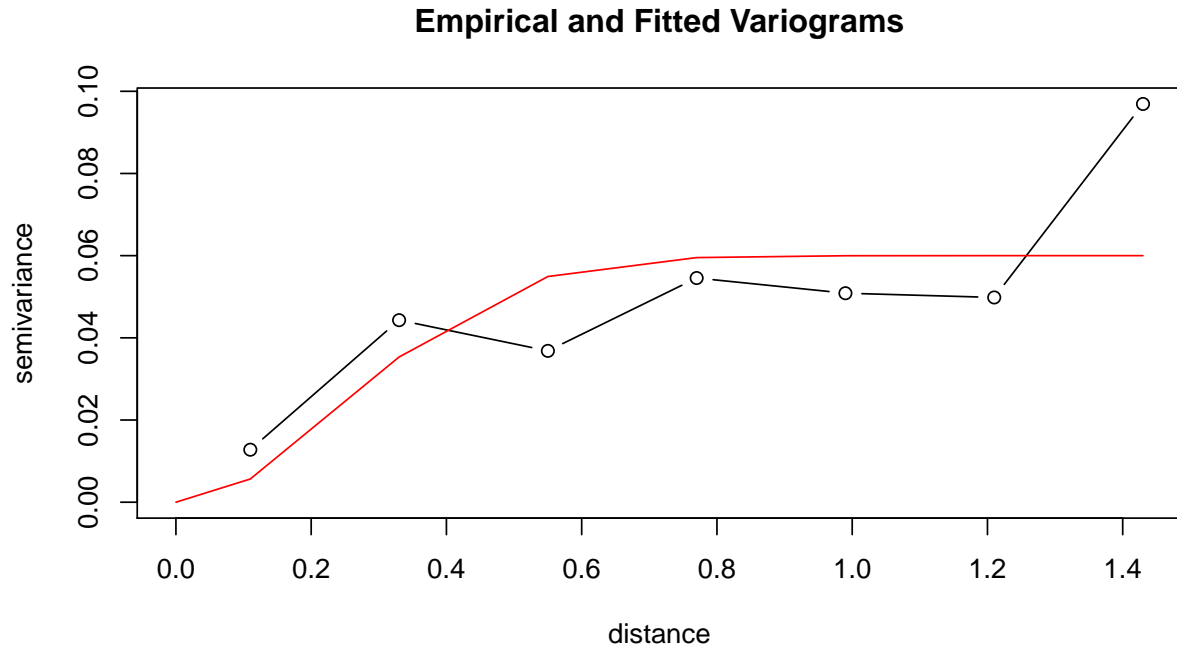
Figure 4.1: Empirical and fitted variograms for the energy SSE. Empirical variogram is black and fitted model is in red.
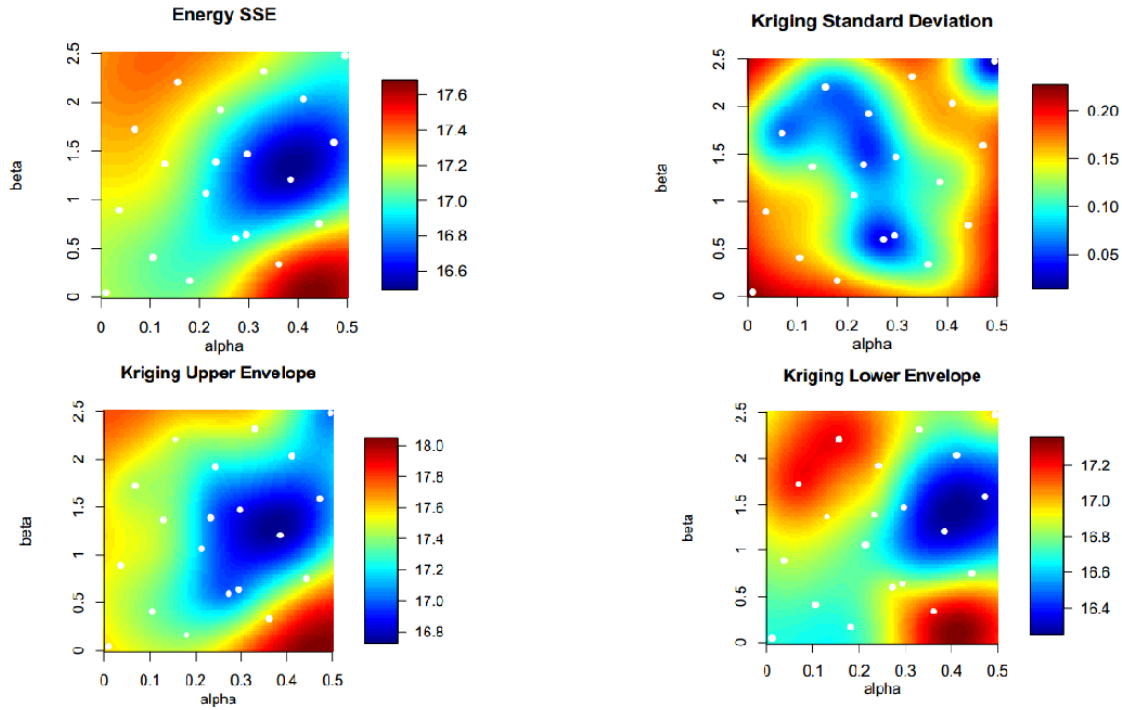
Figure 4.2: Top left: (α, β) slice of predicted surface at the minimum of $SSE(E_\theta)$. White dots are locations of LFM-MIX model runs. Top right: (α, β) slice of kriging prediction variance at the minimum of $SSE(E_\theta)$. Bottom left: (α, β) slice of upper bound of 95% confidence interval at its minimum. Bottom right: (α, β) slice of lower bound of 95% confidence interval at its minimum.
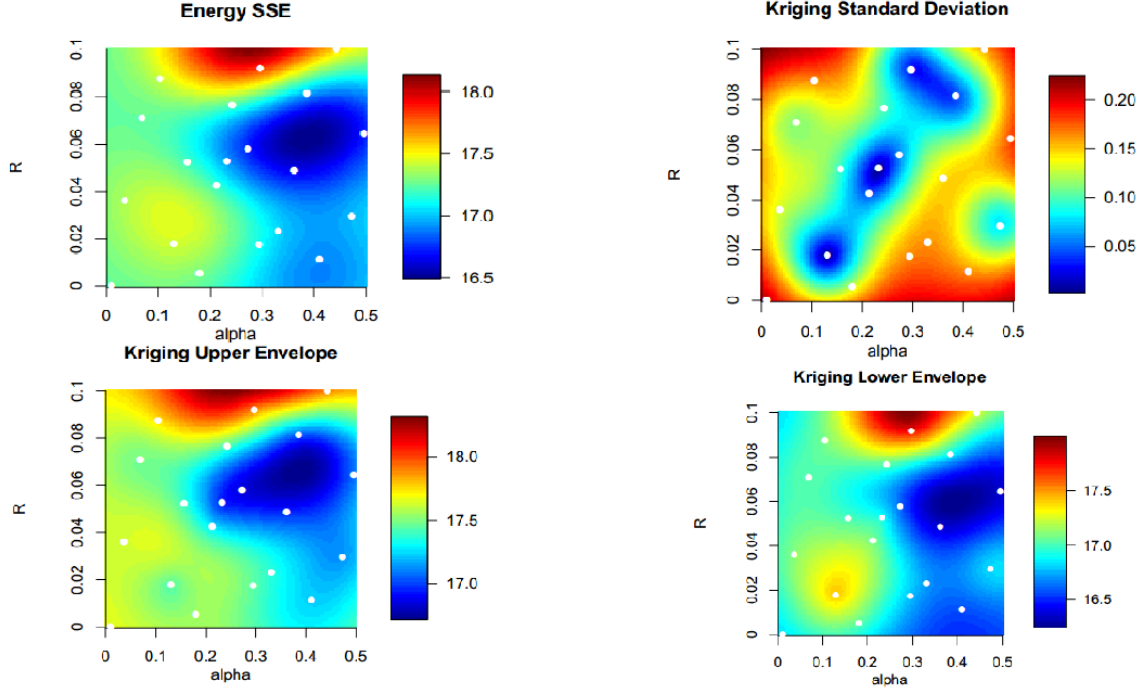
Figure 4.3: Top left: (α, R) slice of predicted surface at the minimum of $SSE(E_\theta)$. White dots are locations of LFM-MIX model runs. Top right: (α, R) slice of kriging prediction variance at the minimum of $SSE(E_\theta)$. Bottom left: (α, R) slice of upper bound of 95% confidence interval at its minimum. Bottom right: (α, R) slice of lower bound of 95% confidence interval at its minimum.
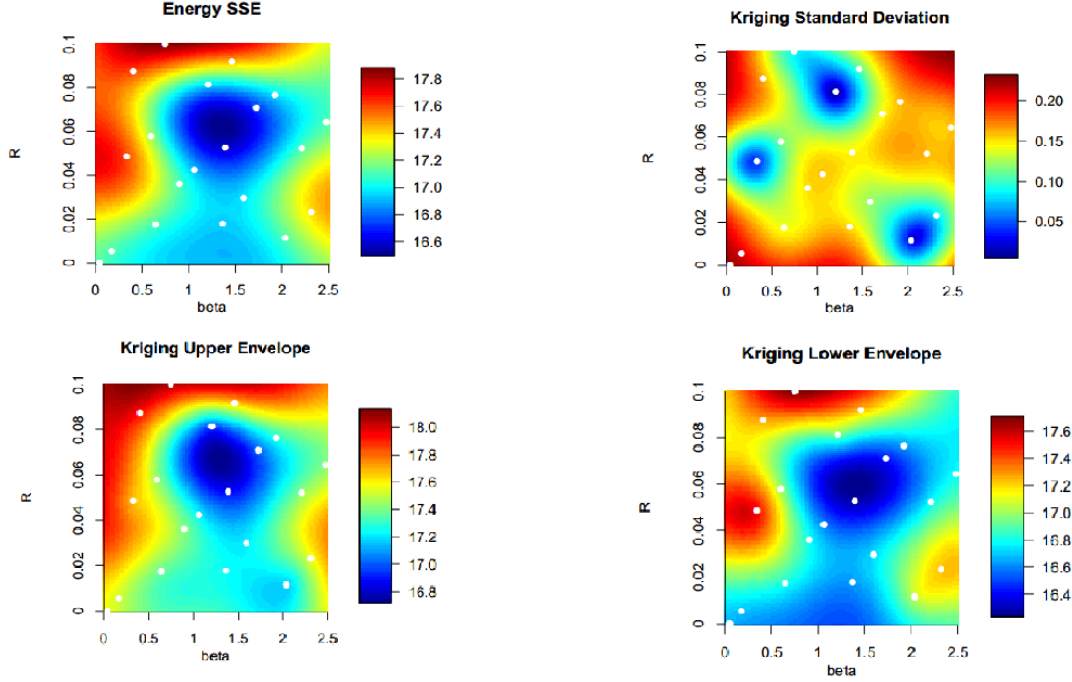
Figure 4.4: Top left: (β, R) slice of predicted surface at the minimum of $SSE(E_\theta)$. White dots are locations of LFM-MIX model runs. Top right: (β, R) slice of kriging prediction variance at the minimum of $SSE(E_\theta)$. Bottom left: (β, R) slice of upper bound of 95% confidence interval at its minimum. Bottom right: (β, R) slice of lower bound of 95% confidence interval at its minimum.

4.2   Rotated Energy

Both exploratory analysis and review of the literature suggest that a rotation may be required to better align LFM-MIX output with the observed storm. The need for a rotational bias is clear from Figure 2.3. Furthermore, Heaton et al. even claim (on the basis of personal discussion with scientists at CISM) that this rotational bias is highly informational in developing future versions of LFM-MIX [6]. Here we use a slightly different approach from that of Heaton et al., while they employ a nonlinear time-varying rotational bias, we begin with a constant rotational bias, then introduce our own time-varying rotational bias in the next section.

We begin by rotating each of the models clockwise around the origin (which corresponds to the North Pole) in increments of $2°$ (from $2°$ to $358°$). Model j rotated by i degrees is denoted by $R^e_{\theta_j,i}(s,t)$. Then we calculate the SSE between the observations and rotated models and, for each model, choose the rotation that yields the minimum SSE. Mathematically, the rotated energy SSE for model j is defined as follows: $SSE(R^e_{\theta_j}) = \min_{i\in\{2,4,\dots,358\}}\sum_{t\in T}\sum_{s\in S}(R^e_{\theta_j,i}(s,t) - E_0(s,t))^2$. Initially, we find that model 8 (with $(\alpha,\beta,R) = (0.18, 0.169, 0.005)$) rotated $104°$ minimizes the rotated energy SSE, $SSE(R^e_\theta)$.

Now we can commence the calibration process by computing the empirical variogram for $SSE(R^e_\theta)$. Using a bin width of 0.22, we find that the empirical variogram for $SSE(R^e_\theta)$ looks almost identical to the empirical variogram for $SSE(E_\theta)$ in Figure 4.1. Next, we fit a Gaussian variogram model to the empirical variogram. Calculating the Gaussian fit for inputs $c = \{0.01, 0.02, \dots, 1.00\}$ and $a = \{0.1, 0.2, \dots, 10.0\}$, we find that $c = 0.03$ and $a = 0.4$ give the best fitted Gaussian model, which looks very similar to the fitted model in Figure 4.1. Finally, we end the calibration process for $SSE(R^e_\theta)$ by kriging the values over the same normalized 3D parameter space as in section 4.1. Upon kriging, we find that the minimum is located at $(0, 0, 0.08)$ in the original parameter space. We notice that the minimum location differs significantly from that of the previous section. This is no surprise; we expect rotated minimum locations to differ from non-rotated minimum locations.

4.3   Energy with Varying Rotation

In this section, we expand upon the idea of section 4.2 by introducing a time-varying rotational bias à la Heaton et. al. We begin by measuring the SSE between observed energy and each rotated model for each time point and rotation. Then, for each model and time point, we choose the rotation that minimizes the SSE, and call this quantity $r_{\theta_j}(t)$.

Now we are ready to develop our time-varying rotational bias. Since geomagnetic storms evolve smoothly over time, the rotational bias should be a smooth function over time [6]. While Heaton et al. choose a rotational bias that follows a natural cubic spline

(with vague Gaussian priors used for the coefficients of the spline bases), we take a slightly different approach. Instead, we fit natural splines of varying degrees of freedom to $r_{\theta_j}(t)$ and minimize the SSE between observed energy and rotated energy to determine the optimal degree of spline fit for each model. We calculate the spline basis functions and coefficients using `ns` in the `splines` package and fit them to the data using `predict` in the `stats` package in R. The following algorithm illustrates this process step-by-step:

(1) For each model j, calculate SSE between observations and each rotation i, for each time point:

   (a) $SSE(R^e_{\theta_j,i})(t) = \sum_{s\in S}(E_0(s,t) - R^e_{\theta_j,i}(s,t))^2$

(2) Find the rotation that minimizes the SSE for each model and time point:

   (a) $r_{\theta_j}(t) = \underset{i\in\{2,4,\ldots,358\}}{\operatorname{argmin}}\ SSE(R_{\theta_j,i})(t)$

(3) Fit splines with degrees of freedom $d = \{1, 2, \ldots, 20\}$, to each $r_{\theta_j}(t)$:

   (a) $S_{\theta_j}(d,t) = \sum_{k=1}^{d} n_k(t)\nu_k$

   (b) $n_k(t)$ are the natural spline basis functions and ν_k is the corresponding coefficient.

(4) Find the degree of freedom that minimizes the SSE between observations and models rotated by $S_{\theta_j}(d,t)$:

   (a) $d_{\theta_j} = \underset{d\in\{1,2,\ldots,20\}}{\operatorname{argmin}} \sum_{s\in S}\sum_{t\in T}(E_0(s,t) - R^e_{\theta_j,S_j(d,t)}(s,t))^2$

(5) The time-varying rotational bias for each model is the best spline fit:

   (a) $\lambda_{\theta_j}(t) = \sum_{k=1}^{d_{\theta_j}} n_k(t)\nu_k$

Finally we can define our next statistic, the SSE for energy with time-varying rotation:

$SSE(R_{\theta_j}^{ev}) = \sum_{s \in S} \sum_{t \in T} (E_0(s,t) - R_{\theta_j}^{ev}(s,t))^2$, where $R_{\theta_j}^{ev}(s,t) = R_{\theta_j, \lambda_{\theta_j}(t)}^{e}(s,t)$. Initially, we find

that model 8 (corresponding to $(\alpha, \beta, R) = (0.18, 0.169, 0.005)$) with a spline of 17 degrees

of freedom minimizes $SSE(R_{\theta_j}^{ev})$.

We begin the calibration process by computing the empirical variogram for $SSE(R_{\theta}^{ev})$.

Using a bin width of 0.22, we find that the empirical variogram looks almost identical to the

empirical variograms for $SSE(E_\theta)$ and $SSE(R_{\theta}^{e})$ (see Figure 4.1). Next, we fit a Gaussian

model to the empirical variogram using inputs $c = a = \{0.01, 0.02, \ldots, 1.00\}$, and find that

$c = 0.03$ and $a = 0.49$ yield the best Gaussian fitted model. Naturally, the Gaussian model

fit looks very similar to that of Figure 4.1. Now we can end the calibration process by kriging

$SSE(R_{\theta}^{ev})$ over the 3D parameter space of section 4.1. We find that the minimum of the

kriged surface is located at $(0.41, 1.43, 0.06)$ in the original parameter space. We notice that

the minimum locations for energy and energy with varying rotation are very similar in α

and R, and the β values differ on a slightly larger scale. This behavior is no surprise, given

that energy is only indirectly related to β via the governing equations 2.1, thus resulting in β

being not as well-defined as α and R for the energy statistics. Analysis of the flux statistics

later in this chapter will shed more light on the optimal range of β.

4.4   Total Energy

Next, we will consider the total energy, which represents the amount of energy in

the geomagnetic storm at any time point. Mathematically, the total energy for model j

is defined as follows: $T_{\theta_j}^{e}(t) = \sum_{s \in S} E_{\theta_j}(s,T)$, where $E_{\theta_j}(s,T)$ is the energy for model j at

spatial location s over all time points, T. Now we define the total energy SSE: $SSE(T_{\theta_j}^{e}) =$

$\sum_{t \in T} (T_{\theta_j}^{e}(t) - T_0^{e}(t))^2$, where $T_0^{e}(t)$ is the total energy of the observations.

Upon plotting the total energy time series for each model, we notice that the models

all have the same structure over time, only with different amounts of energy. See Figure 4.5

below. Again, we notice that model 8 ($(\alpha, \beta, R) = (0.18, 0.169, 0.005)$) minimizes the total energy SSE.
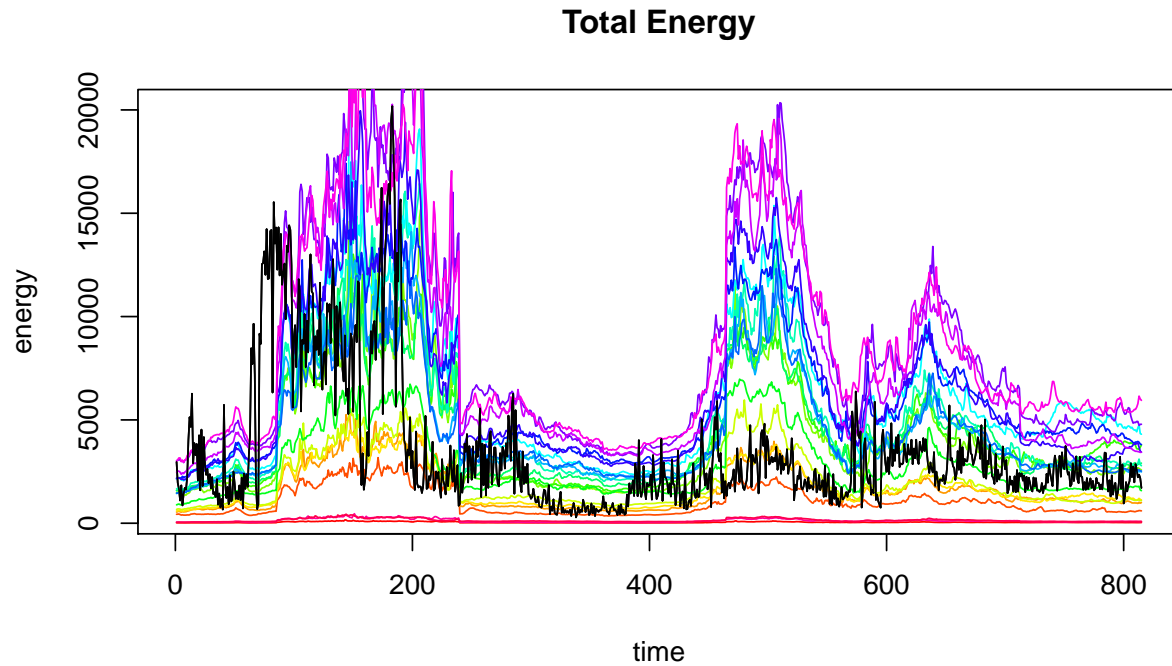
Figure 4.5: Total energy time series. Observations are in black and the models are in color.

Now we can initialize the calibration process, as before, by calculating the empirical variogram. Again, we use a bin width of 0.22 and find that the empirical variogram has the desired qualities: increasing trend and small nugget value. However, as we can see from Figure 4.6, the empirical variogram has a slightly different shape than that of the previous energy statistics.

Empirical and Fitted Variograms
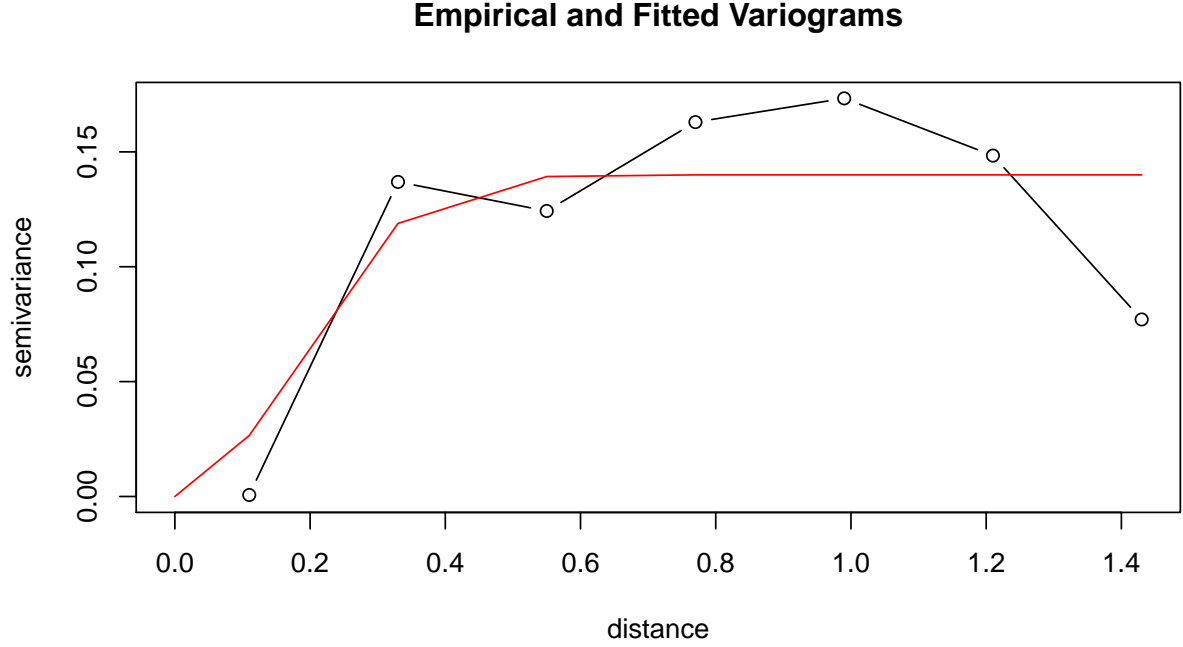
Figure 4.6: Empirical and fitted variograms for the total energy SSE. Empirical variogram is in black and fitted model is in red.

To continue the calibration process, we fit a Gaussian model to the empirical variogram. Using input values $c = \{0.1, 0.2, \ldots, 10.0\}$ and $a = \{0.1, 0.2, \ldots, 10.0\}$, we find that $c = a = 0.3$ yields the best fit. See Figure 4.6 above. Next, we krige the SSE values over the normalized parameter space $[0, 1] \times [0, 1] \times [0, 1]$, resulting in a minimum located at $(0.18, 0.15, 0.01)$ in the original parameter space. Here, we notice that the minimum locations for total energy and the previous energy statistics differ significantly. Once again, this is no surprise; we expect total energy minimums to differ from energy and rotated energy minimums. We also notice that β differs on a much larger scale than α and R. Again, this is because β is known to be not as well-defined as α and R for the energy statistics, since energy only depends on β indirectly via the governing equations 2.1.

4.5   Smooth Total Energy

In Figure 4.5, we notice that the total energy time series are very noisy, which suggests that smoothing may help us better isolate the true signal. In this section, we introduce smoothing for the total energy time series, and in the next section we expand upon this idea by introducing a multiplicative bias.

We smooth the total energy time series (both models and observations) using a simple moving average, also known as a Daniell window/kernel. The smoothing parameters, M_0 for observations and M_{θ_j} for models, represent the number of time points on either side of t that we average to calculate the smoothed value for time t. We choose the optimal smoothing parameters, M_0 and M_{θ_j}, via 10-fold cross validation. The following algorithm illustrates this process for M_0:

(1) For $k \in \{1, 2, \ldots, 10\}$, do steps 2-5:

(2) Randomly hold out 10% of total energy data, call these values $H(t)$

(3) Compute moving average for each $H(t)$ value and each $m \in \{2, 3, \ldots, 40\}$, denote it by $S(m, t)$:

    (a) $S(m, t) = \dfrac{1}{2m + 1} \displaystyle\sum_{i=t-m}^{t+m} T_0^e(i)$

(4) Measure error between $H(t)$ and $S(m, t)$ for each m:

    (a) $Err(m) = \displaystyle\sum_{t=1}^{|H(t)|} |S(m, t) - H(t)|$

(5) Find the m value that minimizes $Err(m)$ for kth fold of cross validation:

    (a) $M(k) = \underset{m \in \{2,3,\ldots,40\}}{\operatorname{argmin}} \; Err(m)$

(6) Average results for each fold of cross validation to get our optimal m value:

    (a) $M_0 = \frac{1}{10} \sum_{k=1}^{10} M(k)$

Now that we have calculated the smoothing parameters, we can define the observed smooth total energy as follows: $\tau_0^e(t) = \dfrac{1}{2M_0 + 1} \displaystyle\sum_{i=t-M_0}^{t+M_0} T_0^e(i)$. The smooth total energy for the models is defined analogously.

Finally, the smooth total energy SSE for model j is defined as follows: $SSE(\tau_{\theta_j}^e) = \displaystyle\sum_{t \in T}(\tau_0^e(t) - \tau_{\theta_j}^e(t))^2$.

And now we are ready to begin the calibration process by computing the empirical variogram. Using a bin width of 0.22, we discover that the empirical variogram looks nearly identical to that of the total energy SSE (see Figure 4.6). Next we fit a Gaussian variogram model, which again looks very similar to that of the total energy SSE. Using input values $c = a = \{0.1, 0.2, \ldots, 10.0\}$, we find that $c = 0.5$ and $a = 0.3$ yields the best fitted variogram. Finally, we end the calibration process for smooth total energy by kriging the SSE values over the normalized parameter space of the previous sections. This results in a minimum at $(0.20, 1.03, 0.05)$ in the original parameter space. This time, we notice that the α and R coordinates of the minimum are similar to those of total energy, while β differs on a larger scale. As previously mentioned, this is to be expected from the governing equations 2.1.

4.6   Scaled Smooth Total Energy

We end our analysis of the energy statistics by expanding on the idea of section 4.5. First, we must jump ahead for a brief moment. Upon initial investigation of the total flux data, we notice that the observations are several orders of magnitude smaller than the models, suggesting the need for a multiplicative bias. We decide to introduce a multiplicative bias, not only for the smooth total flux, but also for the smooth total energy.

After smoothing the total energy, we choose the optimal multiplicative bias, $C^\star$ for each model in the following way:

(1) Choose C values for model j as follows:

(a) Let $m_j = \dfrac{\max_{t \in T} \tau_0^e(t)}{\max_{t \in T} \tau_{\theta_j}^e(t)}$

(b) $C_j = \{0.8 \cdot m_j, 0.85 \cdot m_j, \ldots, 1.2 \cdot m_j\}$

(2) Choose C that minimizes SSE between each model and the observations, call it $C^\star$:

(a) $C^\star(j) = \underset{i \in C_j}{\operatorname{argmin}} \sum_{t \in T} (\tau_0^e(t) - i \cdot \tau_{\theta_j}^e(t))^2$

Finally, we define the SSE for scaled smooth total energy: $SSE(\hat{\tau}_\theta^e) = \sum_{t \in T} (\tau_0^e(t) - \hat{\tau}_{\theta_j}^e(t))^2$, where $\hat{\tau}_{\theta_j}^e(t) = C^\star(j) \cdot \tau_{\theta_j}^e(t)$. And now we can begin the calibration process by calculating the empirical variogram for $SSE(\hat{\tau}_\theta^e)$. Using a bin width of 0.26, we find that the empirical variogram has an increasing trend and small nugget value, as desired (see Figure 4.7). Next, we fit a Gaussian model to the empirical variogram, using inputs $c = a = \{0.01, 0.02, \ldots, 10.00\}$. We find that $c = 0.03$ and $a = 1.07$ yield the best fitted model. Finally, we end the calibration process by kriging the SSE values over the normalized parameter space, resulting in a minimum at $(0, 2.5, 0.1)$ in the original parameter space. We notice that these results vary drastically from those of the other energy statistics. We suspect this is due to a poorly fitting Gaussian variogram model. Perhaps analysis of the flux statistics in the next several sections will aid in explaining this behavior.

Empirical and Fitted Variograms
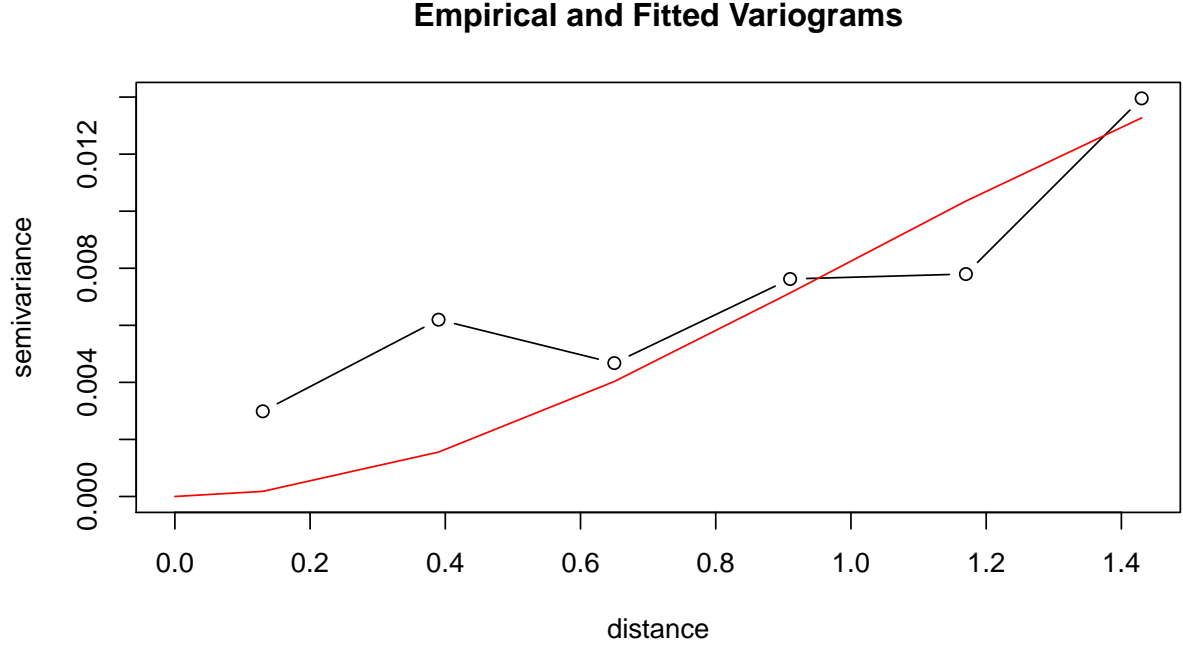
Figure 4.7: Empirical and fitted variograms for the scaled total energy SSE. Empirical variogram is in black and fitted model is in red.

4.7 Flux

In this section we begin our analysis of the flux statistics by examining the flux SSE, denoted by $SSE(F_\theta)$. We define the flux SSE for model j as follows:

$SSE(F_{\theta_j}) = \sum_{t \in T} \sum_{s \in S} (F_0(s,t) - F_{\theta_j}(s,t))^2$. Upon calculation of the flux SSE we notice that model 20, corresponding to $(\alpha, \beta, R) = (0.294, 0.64, 0.018)$ minimizes $SSE(F_\theta)$. Now we initiate the calibration process by computing the empirical variogram for $SSE(F_\theta)$. Using a bin width of 0.23, we find that the empirical variogram has an increasing trend and a small nugget value, as desired (see the black line in Figure 4.8). Next, we fit a Gaussian variogram model to the empirical variogram. To do this, we computed the Gaussian model at $c = a = \{0.1, 0.2, \ldots, 10.0\}$, and found that using $c = 3.7$ and $a = 0.2$ minimizes the SSE between the empirical and fitted variograms (see Figure 4.8 again).

Empirical and Fitted Variograms
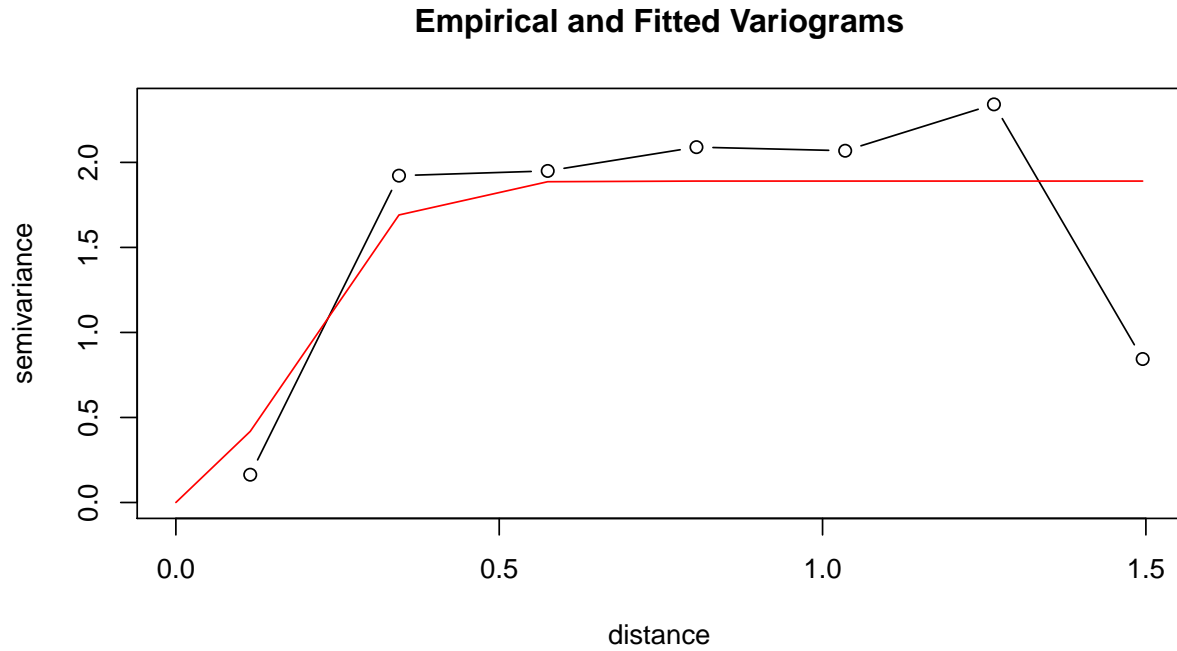
Figure 4.8: Empirical and fitted variograms for the flux SSE. Empirical variogram is in black and fitted model is in red.

Now that we have established an appropriate variogram model, we can finish calibration of $SSE(F_\theta)$ by kriging the SSE values over the normalized parameter space used in the previous sections. Figures 4.9-4.11 show 2D slices of the predicted field at the minimum of the normalized parameter space, which corresponds to the point $(0.29, 0.65, 0.02)$ in the original parameter space.
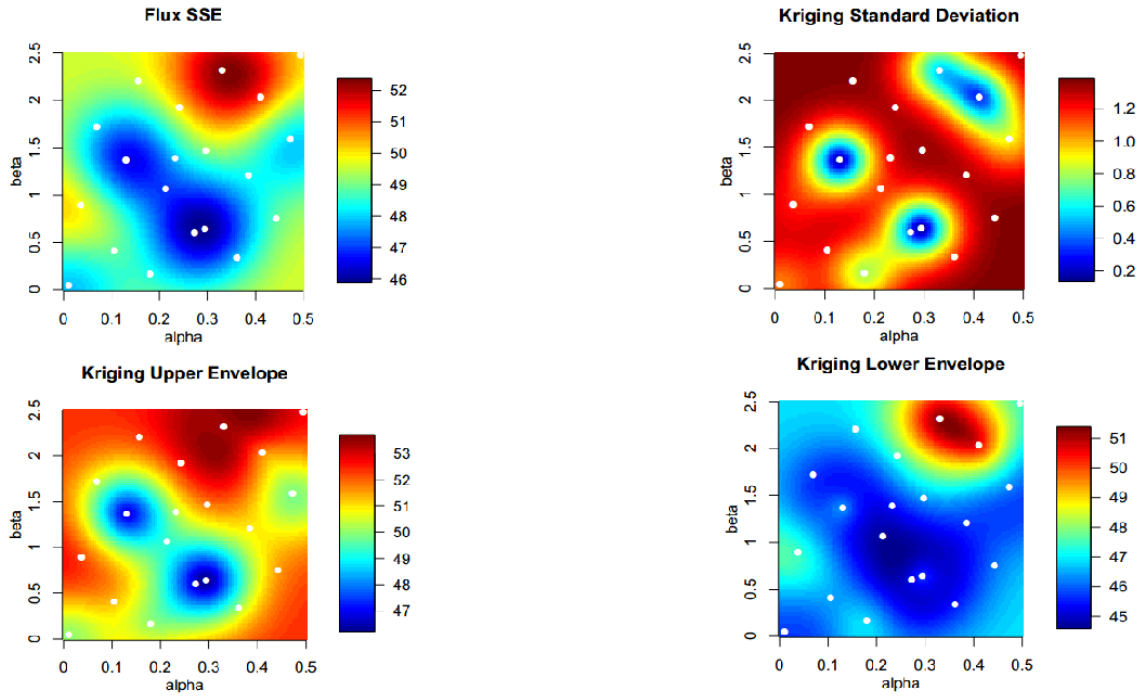
Figure 4.9: Top left: (α, β) slice of predicted surface at the minimum of $SSE(F_\theta)$. White dots are locations of LFM-MIX model runs. Top right: (α, β) slice of kriging prediction variance at the minimum of $SSE(F_\theta)$. Bottom left: (α, β) slice of upper bound of 95% confidence interval at its minimum. Bottom right: (α, β) slice of lower bound of 95% confidence interval at its minimum.
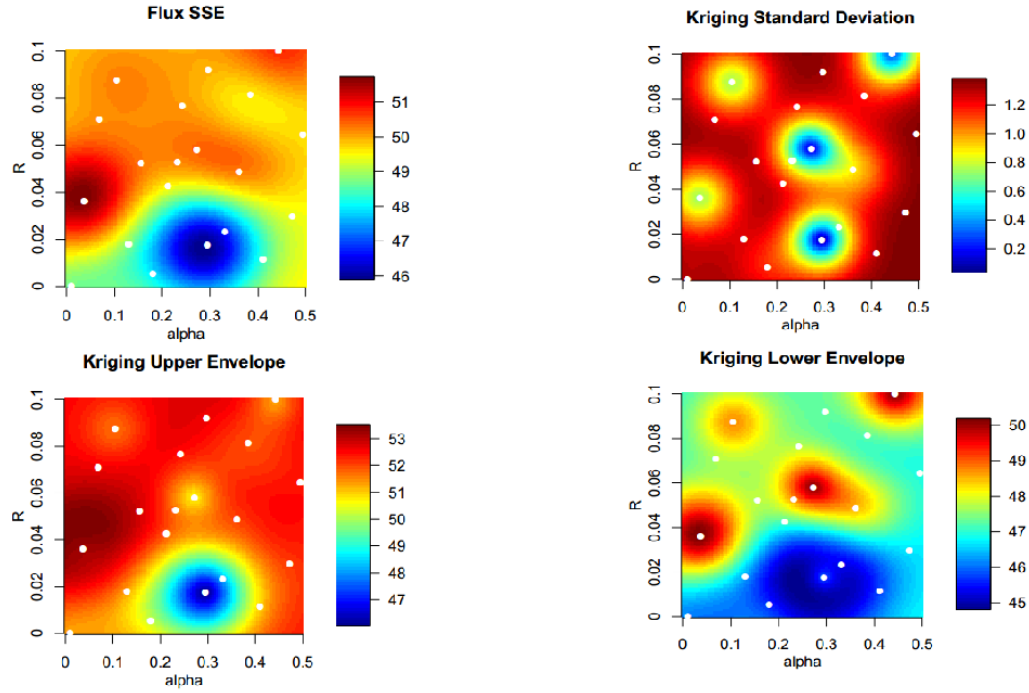
Figure 4.10: Top left: (α, R) slice of predicted surface at the minimum of $SSE(F_\theta)$. White dots are locations of LFM-MIX model runs. Top right: (α, R) slice of kriging prediction variance at the minimum of $SSE(F_\theta)$. Bottom left: (α, R) slice of upper bound of 95% confidence interval at its minimum. Bottom right: (α, R) slice of lower bound of 95% confidence interval at its minimum.
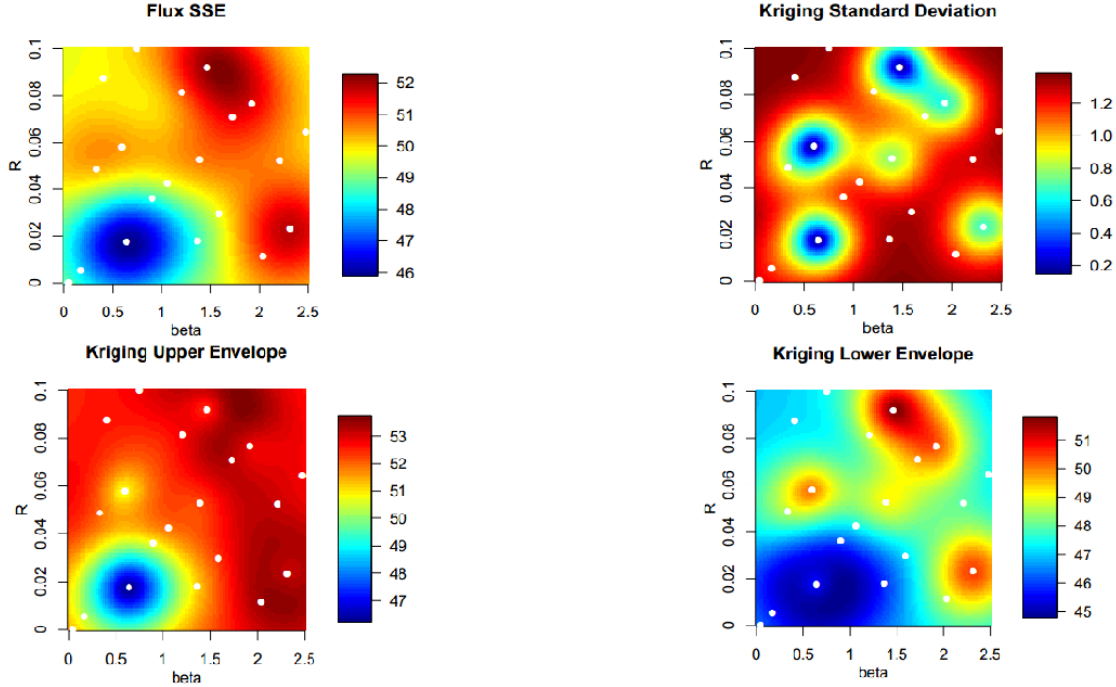
Figure 4.11: Top left: (β, R) slice of predicted surface at the minimum of $SSE(F_\theta)$. White dots are locations of LFM-MIX model runs. Top right: (β, R) slice of kriging prediction variance at the minimum of $SSE(F_\theta)$. Bottom left: (β, R) slice of upper bound of 95% confidence interval at its minimum. Bottom right: (β, R) slice of lower bound of 95% confidence interval at its minimum.

## 4.8   Rotated Flux

We proceed in a similar manner to that of section 4.2, by calculating the SSE between the observed flux and rotated models, and choosing the rotation that minimizes the SSE.

Mathematically, the rotated flux SSE for model j is defined as follows: $SSE(R^f_{\theta_j}) = \min_{i \in \{2,4,...,358\}} \sum_{t \in T} \sum_{s \in S} (R^f_{\theta_j,i}(s,t) - F_0(s,t))^2$. Initially, we find that model 20 (with $(\alpha, \beta, R) = (0.294, 0.64, 0.018)$) rotated 6° minimizes the rotated flux SSE, $SSE(R^f_\theta)$.

Now we can initiate the calibration process by computing the empirical variogram for $SSE(R^f_\theta)$. Using a bin width of 0.23, we find that the empirical variogram for $SSE(R^f_\theta)$

looks nearly identical to the empirical variogram for $SSE(F_\theta)$ in Figure 4.8. Next, we fit a Gaussian variogram model to the empirical variogram. Calculating the Gaussian fit for inputs $c = a = \{0.1, 0.2, \ldots, 10.0\}$, we find that $c = 3.7$ and $a = 0.2$ give the best fitted Gaussian model, which naturally looks very similar to the fitted model in Figure 4.8. Finally, we end the calibration process for $SSE(R_\theta^f)$ by kriging the values over the same normalized 3D parameter space as in previous sections. Upon kriging, we find that the minimum is located at $(0.29, 0.65, 0.02)$ in the original parameter space. We notice that the minimum locations for flux and rotated flux are identical.

4.9   Flux with Varying Rotation

In this section, we proceed in the same manner as section 4.3, by introducing a time-varying rotational bias for flux. Again, we will fit natural splines of varying degrees of freedom to $r_{\theta_j}(t)$ (the rotation that minimizes SSE for each time point), and minimize the SSE between observed flux and rotated flux to determine the optimal degree of spline fit for each model.

We define the SSE for flux with a time-varying rotational bias as follows: $SSE(R_{\theta_j}^{fv}) = \sum_{s \in S} \sum_{t \in T} (F_0(s, t) - R_{\theta_j}^{fv}(s, t))^2$, where $R_{\theta_j}^{fv}(s, t) = R_{\theta_j, \lambda_{\theta_j}(t)}^f(s, t)$. Initially, we find that model 20 (corresponding to $(\alpha, \beta, R) = (0.294, 0.64, 0.018)$) with a spline of 19 degrees of freedom minimizes $SSE(R_{\theta_j}^{fv})$.

We begin the calibration process by computing the empirical variogram for $SSE(R_\theta^{fv})$. Using a bin width of 0.23, we find that the empirical variogram looks almost identical to the empirical variograms for $SSE(F_\theta)$ and $SSE(R_\theta^f)$ (see Figure 4.8). Next, we fit a Gaussian model to the empirical variogram using inputs $c = a = \{0.1, 0.2, \ldots, 10.0\}$, and find that $c = 3.6$ and $a = 0.2$ yield the best Gaussian fitted model. Naturally, the Gaussian model fit looks very similar to that of Figure 4.8. Now we end the calibration process by kriging $SSE(R_\theta^{fv})$ over the 3D parameter space of the previous sections. We find that the minimum of the kriged surface is located at $(0.29, 0.68, 0.02)$ in the original parameter space.

4.10   Total Flux

Next we will consider the total flux, which represents the amount of flux in the geomagnetic storm at any time point. Mathematically, we define the total flux for model j as follows: $T_{\theta_j}^f(t) = \sum_{s \in S} F_{\theta_j}(s, T)$, where $F_{\theta_j}(s, T)$ is the flux for model j at spatial location s over all time points, T. Now we define the total flux SSE: $SSE(T_{\theta_j}^f) = \sum_{t \in T}(T_{\theta_j}^f(t) - T_0^f(t))^2$, where $T_0^f(t)$ is the total flux of the observations. Upon initial inspection of the total flux time series, we find once again that model 20 $((\alpha, \beta, R) = (0.294, 0.64, 0.018))$ minimizes $SSE(T_\theta^f)$.

Now we can initialize the calibration process, as before, by calculating the empirical variogram. Again, we use a bin width of 0.23 and find that the empirical variogram looks nearly identical to that of the flux, rotated flux, and flux with varying rotation SSEs. To continue the calibration process, we fit a Gaussian model to the empirical variogram. Using input values $c = a = \{0.1, 0.2, \ldots, 10.0\}$, we find that $c = 4.0$ and $a = 0.2$ yield the best fit. Next, we krige the SSE values over the normalized parameter space $[0, 1] \times [0, 1] \times [0, 1]$, resulting in a minimum located at $(0.29, 0.68, 0.02)$ in the original parameter space). We note that calibration of flux with varying rotation and total flux yield the same minimum.

4.11   Smooth Total Flux

Here we follow the same approach as in section 4.5, smoothing the total flux with a simple moving average. Again, we find the optimal smoothing parameters, M_0 and M_{θ_j}, by implementing 10-fold cross validation. After identifying M_0 and M_{θ_j}, we define the observed smooth total flux as follows: $\tau_0^f(t) = \frac{1}{2M_0 + 1} \sum_{i=t-M_0}^{t+M_0} T_0^f(i)$. And finally, the smooth total energy SSE for model j is defined as follows: $SSE(\tau_{\theta_j}^f) = \sum_{t \in T}(\tau_0^f(t) - \tau_{\theta_j}^f(t))^2$.

We begin the calibration process, as usual, by computing the empirical variogram. Using a bin width of 0.23, we discover that the empirical variogram has the desired qualities:

increasing trend and small nugget value (see Figure 4.12). Now we fit a Gaussian variogram model, using inputs $c = a = \{0.1, 0.2, \ldots, 10.0\}$, and find that $c = 8.6$ and $a = 9.9$ yield the best fit. Finally, we krige the SSE values over the normalized parameter space, resulting in a minimum at $(0.5, 0, 0)$ in the original parameter space.

Empirical and Fitted Variograms
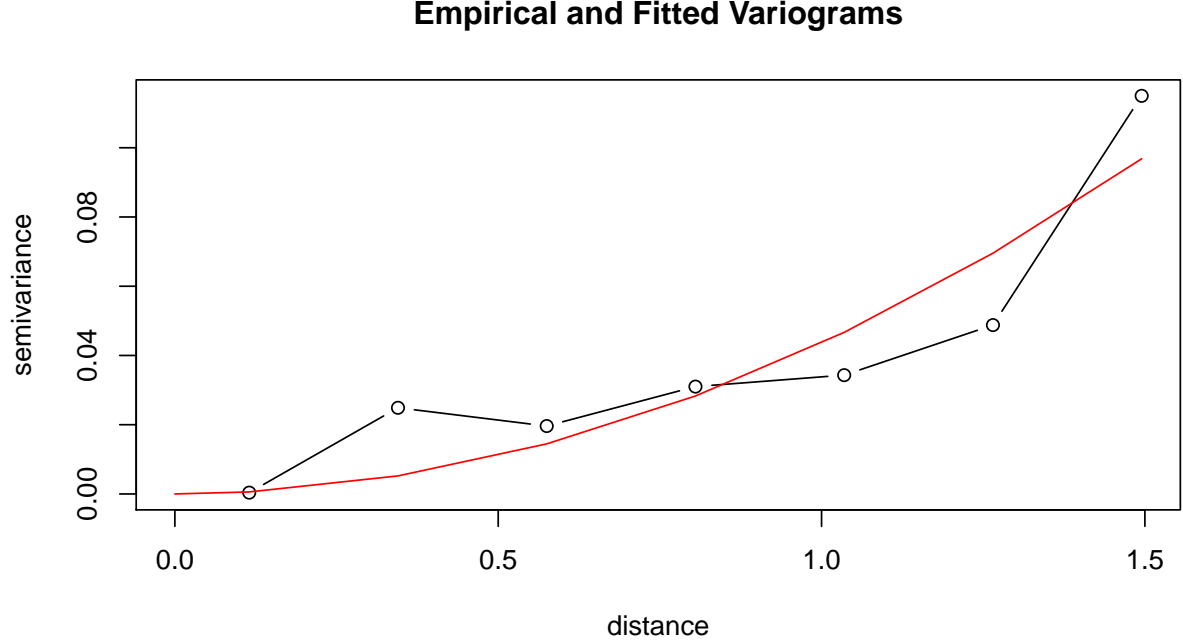
Figure 4.12: Empirical and fitted variograms for the smooth total flux SSE. Empirical variogram is in black and fitted model is in red.

4.12   Scaled Smooth Total Flux

After smoothing the total flux, we introduce a multiplicative bias in a manner similar to that of section 4.6. We define the SSE for scaled smooth total flux as follows: $SSE(\hat{\tau}_\theta^f) = \sum_{t \in T}(\tau_0^f(t) - \hat{\tau}_{\theta_j}^f(t))^2$, where $\hat{\tau}_{\theta_j}^f(t) = C^\star(j) \cdot \tau_{\theta_j}^f(t)$ and $C^\star(j)$ is the optimal multiplicative bias for model j. Now we can begin the calibration process by calculating the empirical variogram for $SSE(\hat{\tau}_\theta^f)$. We compute the empirical variogram for bin widths $h = \{0.10, 0.11, \ldots, 0.40\}$, but find that none of these variograms shows any indication of spatial correlation. This

means that we cannot continue calibration of this statistic, and must eliminate it from further analysis.

4.13   Discussion of feature-based calibration

We begin our discussion of the results of feature-based calibration by recalling equations 2.1, which govern the input parameter values. We recall that energy is directly related to α and R, while flux is directly related to β. These relationships suggest that α and R are well identified by the energy statistics, but β may not be. Analogously, these relationships suggest that β is well identified by the flux statistics, but α and R may not be.

In Table 4.1 and Figures 4.15-4.17, we notice that the minimums for the flux statistics center around $(0.29, 0.67, 0.02)$ in the original parameter space, while the energy minimums do not center on any one particular point. We also note that the significant difference between the rotated energy and energy minimums is to be expected; optimal parameter values derived from rotated models should be different from parameter values derived without rotations. Analogously, it is no surprise that the minimums for total energy and smooth total energy differ so much from the minimum for energy. However, the large differences between minimums for total energy and smooth total energy and total flux and smooth total flux are unexpected. These significant differences in minimums are likely due to oversmoothing of the observed total energy and total flux. To choose the degree of smoothing for each time series, we performed 10-fold cross validation with the goal of minimizing absolute error between smooth and unsmoothed time series. Then we averaged the results of the 10 folds to yield our smoothing parameter. In Figures 4.13-4.14 we can see that the smoothed time series do not capture the large spikes in observed total energy and flux around time 200, thus causing calibration of the smoothed statistics to yield very different minimums than that of the unsmoothed ones. Perhaps a better method of choosing the smoothing parameters would be to perform leave-one-out cross validation similar to Chapter 5. This would involve holding out one of the models as the new observations, then smoothing the time series and

choosing the smoothing parameter that minimizes the error between the true (α, β, R) value and the kriged estimate.
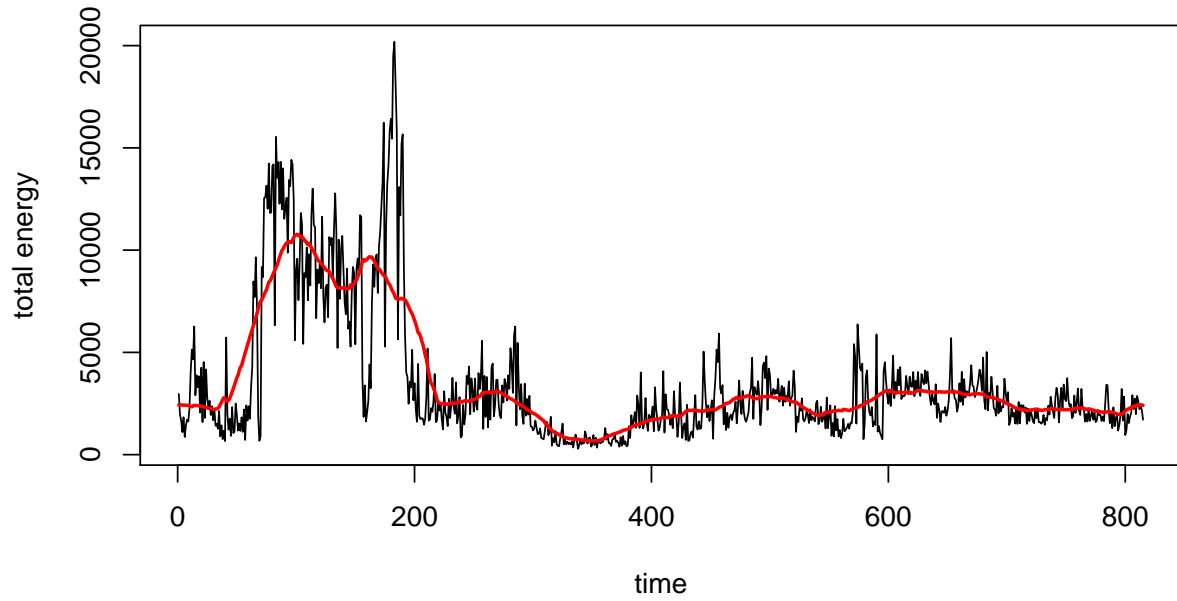
Figure 4.13: Total energy time series. Observed total energy is in black and smooth observed total energy is in red.
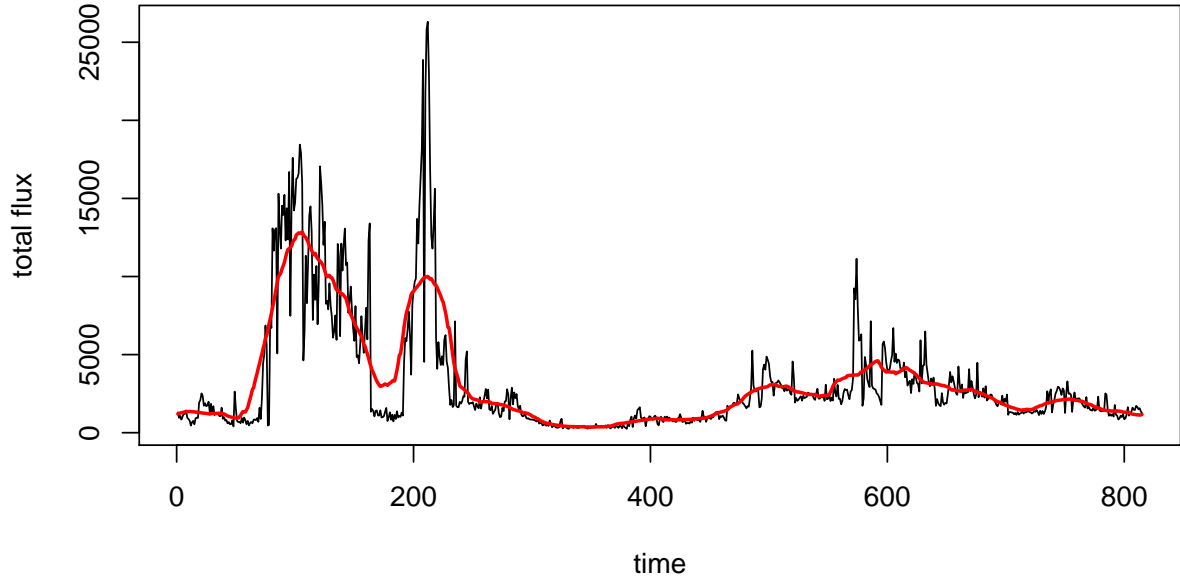
Figure 4.14: Total flux time series. Observed total flux is in black and smooth observed total flux is in red.

Some statistics show more spatial correlation than others, so to determine one optimal input setting θ_0 we should give more consideration to those statistics that show better spatial correlation. If we revisit the variogram plots in Figures 4.1, 4.6, 4.7, 4.8, and 4.12, we see, for example, that the energy variogram reaches its sill slower than the total energy variogram does, meaning that energy exhibits more spatial correlation than total energy. Following this ideology, we conclude that energy, rotated energy, energy with varying rotation, scaled smooth total energy, and smooth total flux exhibit the greatest spatial correlation. Here we must note that we should eliminate rotated energy, scaled smooth total energy, and smooth total flux from further consideration, as the extreme values of their minimums are likely to yield nonphysical results when used as input settings for the LFM-MIX.

We conclude our discussion of feature-based calibration results by comparing our results to those found in the literature. In [6], Heaton et al. center on $\theta_0 = (0.47, 1.59, 0.02)$ as

their optimal input setting. We find this encouraging, as our minimum for energy with varying rotation is $\theta = (0.41, 1.43, 0.06)$. Kleiber et al. do not center on one particular point, instead they show two dimensional slices of samples from the posterior distribution of the input parameters (see Figure 3 in [11]). They identify $\alpha \in (0.1, 0.2)$ and $\beta \in (0, 0.5)$ as a likely region for the input parameters, which agrees with our total energy minimum of $(0.18, 0.15, 0.01)$. They also identify $\alpha \in (0.1, 0.2)$ and $R \in (0.03, 0.05)$ as another likely region for the input parameters, which agrees with our smooth total energy minimum of $(0.20, 1.03, 0.05)$. Finally, a different set of authors also known as Kleiber et al. find that their posterior densities center on $\theta_0 = (0.22, 0.75, 0.029)$ [12]. Again, we find this encouraging, as our minimums for the flux statistics center on approximately $\theta_0 = (0.29, 0.67, 0.02)$.

Table 4.1: Results of feature-based calibration

Statistics	Min Locations	Min Values
Energy	(0.39, 1.35, 0.06)	16.504 ± 0.237
Rotated Energy	(0, 0, 0.08)	16.613 ± 0.088
Energy w/ Varying Rotation	(0.41, 1.43, 0.06)	16.676 ± 0.083
Total Energy	(0.18, 0.15, 0.01)	22.794 ± 0.034
Smooth Total Energy	(0.20, 1.03, 0.05)	22.099 ± 0.284
Scaled Smooth Total Energy	(0, 2.5, 0.1)	19.538 ± 0.027
Flux	(0.29, 0.65, 0.02)	45.946 ± 0.412
Rotated Flux	(0.29, 0.65, 0.02)	45.947 ± 0.412
Flux w/ Varying Rotation	(0.29, 0.68, 0.02)	46.337 ± 0.293
Total Flux	(0.29, 0.68, 0.02)	53.741 ± 0.308
Smooth Total Flux	(0.5, 0, 0)	20.211 ± 0.002
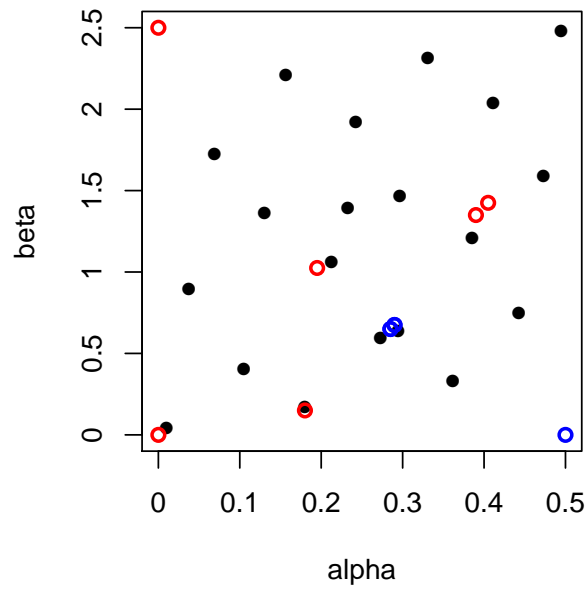
Figure 4.15: (α, β) slice of model locations and minimum locations. Model runs are the solid black points, energy minimums are the red open circles, and flux minimums are the blue open circles.
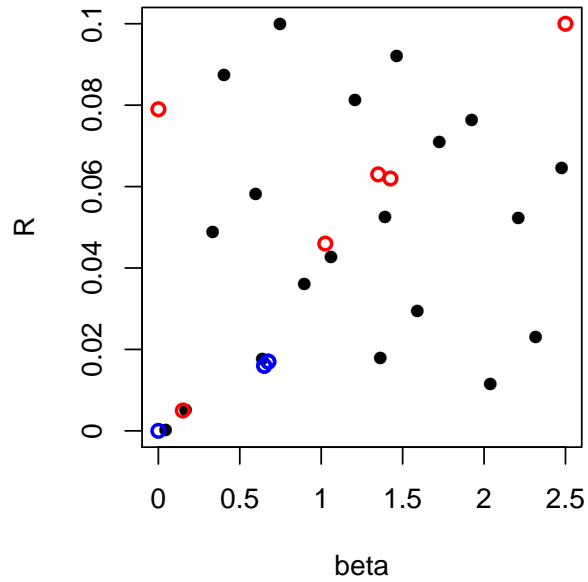
Figure 4.16: (β, R) slice of model locations and minimum locations. Model runs are the solid black points, energy minimums are the red open circles, and flux minimums are the blue open circles.
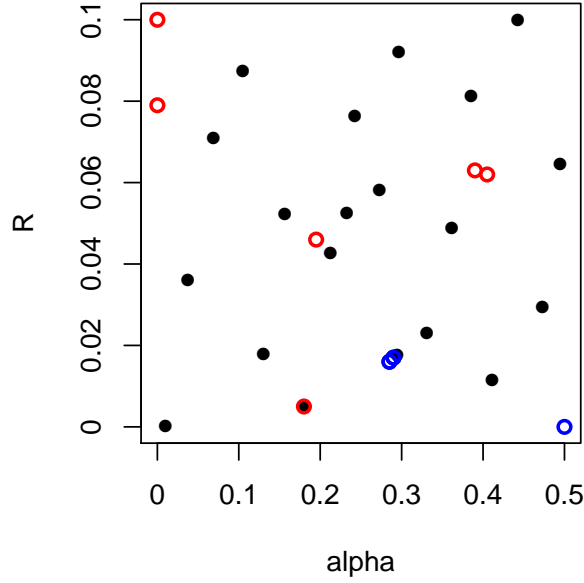
Figure 4.17: (α, R) slice of model locations and minimum locations. Model runs are the solid black points, energy minimums are the red open circles, and flux minimums are the blue open circles.

4.14   Weighted feature-based calibration

In an attempt to improve upon the results of our feature-based calibration, we introduce a sigmoidal weight function to our data. As discussed in Section 3.3, we know that the geomagnetic storm was measured by satellite, which did not observe the entire spatial domain. And we have reason to believe that even the measurements the satellite was able to make do not capture all of the geomagnetic activity that occurred in the observed spatial domain. To correct this in our measures of error between models and observations, we want to put less weight on the locations where the observed values are small, and more weight on the larger observations. Our weight function is defined as follows:

$$w(x) = \frac{1}{1 + e^{-(x-m)}}$$

where m denotes the median of the observed data. We choose a shifted standard logistic function in particular because it is a smooth monotonically increasing function with domain $\mathbb{R}$ and range $[0, 1]$. The reason behind the shift is to align the sigmoid's midpoint with the median of the observations.

The process of weighted feature-based calibration is nearly identical to that of the unweighted process, the only difference being in the calculation of the SSEs for kriging. Since our weight function needs to depend on both time and space, the weighted SSEs for the statistics based on total energy and total flux require a slightly different approach than that of the other statistics. For example, in calculating the weighted SSE for total energy we take the average of the sigmoid weight function evaluated at the observed energy, rather than using the sigmoid weight function evaluated at the observed total energy. Table 4.2 on the next page gives the formulas for the weighted SSEs for each statistic.

Table 4.2: Weighted SSEs

Statistics	Weighted SSEs for model j
Energy	$\sum\limits_{t\in T}\sum\limits_{s\in S} w(E_0(s,t))\cdot(E_0(s,t)-E_{\theta_j}(s,t))^2$
Rotated Energy	$\sum\limits_{t\in T}\sum\limits_{s\in S} w(E_0(s,t))\cdot(E_0(s,t)-R^e_{\theta_j}(s,t))^2$
Energy w/ Varying Rotation	$\sum\limits_{t\in T}\sum\limits_{s\in S} w(E_0(s,t))\cdot(R^{ev}_0(s,t)-R^{ev}_{\theta_j}(s,t))^2$
Total Energy	$\sum\limits_{t\in T}\left(\frac{1}{|S|}\sum\limits_{s\in S} w(E_0(s,t))\right)\cdot(T^e_0(t)-T^e_{\theta_j}(t))^2$
Smooth Total Energy	$\sum\limits_{t\in T}\left(\frac{1}{|S|}\sum\limits_{s\in S} w(E_0(s,t))\right)\cdot(\tau^e_0(t)-\tau^e_{\theta_j}(t))^2$
Scaled Smooth Total Energy	$\sum\limits_{t\in T}\left(\frac{1}{|S|}\sum\limits_{s\in S} w(E_0(s,t))\right)\cdot(\hat{\tau}^e_0(t)-\hat{\tau}^e_{\theta_j}(t))^2$
Flux	$\sum\limits_{t\in T}\sum\limits_{s\in S} w(F_0(s,t))\cdot(F_0(s,t)-F_{\theta_j}(s,t))^2$
Rotated Flux	$\sum\limits_{t\in T}\sum\limits_{s\in S} w(F_0(s,t))\cdot(F_0(s,t)-R^f_{\theta_j}(s,t))^2$
Flux w/ Varying Rotation	$\sum\limits_{t\in T}\sum\limits_{s\in S} w(F_0(s,t))\cdot(R^{fv}_0(s,t)-R^{fv}_{\theta_j}(s,t))^2$
Total Flux	$\sum\limits_{t\in T}\left(\frac{1}{|S|}\sum\limits_{s\in S} w(F_0(s,t))\right)\cdot(T^f_0(t)-T^f_{\theta_j}(t))^2$
Smooth Total Flux	$\sum\limits_{t\in T}\left(\frac{1}{|S|}\sum\limits_{s\in S} w(F_0(s,t))\right)\cdot(\tau^f_0(t)-\tau^f_{\theta_j}(t))^2$

We begin our discussion of the results of weighted feature-based calibration by comparing the weighted and unweighted calibration results. In Table 4.4 we see that the weighted and unweighted calibration results are similar for most statistics and note that only rotated energy, total energy, and smooth total flux differ significantly. We also notice that after weighted calibration the smooth total flux minimum agrees with the other flux statistics, validating our decision to apply a weight function to the data. Again, we remove scaled smooth total energy from further consideration due to the extreme value of its minimum.

Finally, we compare the results of weighted calibration to results found in the literature. Heaton et al. center on $\theta_0 = (0.47, 1.59, 0.02)$ in [6]. Again this is encouraging, as

Table 4.3: Results of weighted feature-based calibration

Statistics	Min Locations	Min Values
Energy	(0.42, 1.45, 0.06)	16.636 ± 0.073
Rotated Energy	(0.42, 1.38, 0.06)	16.796 ± 0.059
Energy w/ Varying Rotation	(0.45, 2.0, 0)	16.797 ± 0.050
Total Energy	(0.45, 1.93, 0.01)	21.464 ± 0.280
Smooth Total Energy	(0.21, 1.03, 0.05)	20.591 ± 9.075
Scaled Smooth Total Energy	(0, 2.5, 0.1)	19.750 ± 0.106
Flux	(0.29, 0.65, 0.02)	45.490 ± 0.412
Rotated Flux	(0.29, 0.68, 0.02)	45.493 ± 0.402
Flux w/ Varying Rotation	(0.29, 0.68, 0.02)	45.947 ± 0.398
Total Flux	(0.29, 0.68, 0.02)	52.325 ± 0.468
Smooth Total Flux	(0.29, 0.68, 0.02)	52.130 ± 0.307

the weighted energy minimum is $(0.42, 1.45, 0.06)$. In fact, the weighted energy minimum is closer to Heaton's result than the unweighted energy with varying rotation. As previously mentioned, in [11] Kleiber et al. show two dimensional slices of samples from the posterior distribution of the input parameters. Their findings indicate $\alpha \in (0.1, 0.2)$ and $R \in (0.03, 0.05)$ as a likely region for the optimal input parameters, which agrees with our weighted smooth total energy minimum of $(0.21, 1.03, 0.05)$. Finally, Kleiber et al. in [12] find their posterior densities center on $\theta_0 = (0.22, 0.75, 0.029)$, which is also encouraging since our weighted flux minimums center around $(0.29, 0.68, 0.02)$.

Table 4.4: Comparison of weighted and unweighted calibration results

Statistics	Min Locations (Unweighted)	Min Locations (Weighted)
Energy	(0.39, 1.35, 0.06)	(0.42, 1.45, 0.06)
Rotated Energy	(0, 0, 0.08)	(0.42, 1.38, 0.06)
Energy w/ Varying Rotation	(0.41, 1.43, 0.06)	(0.45, 2.0, 0)
Total Energy	(0.18, 0.15, 0.01)	(0.45, 1.93, 0.01)
Smooth Total Energy	(0.20, 1.03, 0.05)	(0.21, 1.03, 0.05)
Scaled Smooth Total Energy	(0, 2.5, 0.1)	(0, 2.5, 0.1)
Flux	(0.29 0.65 0.02)	(0.29, 0.65, 0.02)
Rotated Flux	(0.29, 0.65, 0.02)	(0.29, 0.68, 0.02)
Flux w/ Varying Rotation	(0.29, 0.68, 0.02)	(0.29, 0.68, 0.02)
Total Flux	(0.29, 0.68, 0.02)	(0.29, 0.68, 0.02)
Smooth Total Flux	(0.5, 0, 0)	(0.29, 0.68, 0.02)

Chapter 5

Cross Validation Experiment

5.1   Method

To validate the results of Chapter 4, we perform leave-one-out cross validation. In the interest of time, we will only do cross validation for energy, flux, weighted energy, and weighted flux. The following illustrates this process:

(1) For $k \in \{1, \ldots, 20\}$, do steps 2-9:

(2) Choose model k to be the new observations

(3) Measure SSE between new observations and the remaining models

(4) Calculate empirical variogram based on the SSEs

(5) Fit Gaussian model to empirical variogram

(6) Calculate the covariance function using the parameters from the fitted variogram model

(7) Predict SSE values by kriging over the normalized parameter space

(8) Locate minimum of predicted surface

(9) Measure Euclidean distance between predicted minimum and true parameter values

(10) Average Euclidean distances to yield a single measure of error

5.2   Results

Table 5.1 and Figures 5.1-5.4 summarize the results of our leave-one-out cross validation experiment. The rows of Table 5.1 show the cross validation errors when each model is held out as the new observations, with the last row yielding the average of the previous rows. As previously mentioned, each cross validation error is the average of the Euclidean distances between the kriged minimums and the true parameter values. The distances between kriged minimums and true parameter values are calculated in the normalized parameter space, rather than the original space $[0, 0.5] \times [0, 2.5] \times [0, 0.1]$. As a baseline for comparison, Table 5.1 also includes the cross validation error for randomly choosing a model run. This random cross validation error is calculated as follows:

(1) For $k \in \{1, \ldots, 20\}$ do steps 2-5:

(2) Hold out model k as the new observations

(3) Choose a random model run

(4) Calculate the Euclidean distance between the randomly chosen parameter values and the held out parameter values

(5) Save the average of these distances, call it e_r

(6) Perform steps 1-5 100 times

(7) Average the e_r's to yield a single measure of error

Our results indicate that energy and weighted energy calibration actually perform worse than if we just chose a random model run. Flux and weighted flux calibration, on the other hand, perform slightly better than choosing a random model run. From Figures 5.1-5.4 it seems that parameters in $[0.3, 0.5] \times [0.5, 1.3] \times [0.06, 0.1]$ were well estimated by energy calibration, parameters in $[0.25, 0.3] \times [1.5, 2.0] \times [0.07, 0.09]$ were well estimated by flux

calibration, parameters in $[0.3, 0.5] \times [0.5, 0.75] \times [0.06, 0.1]$ were well estimated by weighted energy, and parameters in $[0.25, 0.3] \times [1.5, 2.0] \times [0.08, 0.09]$ were well estimated by weighted flux calibration.
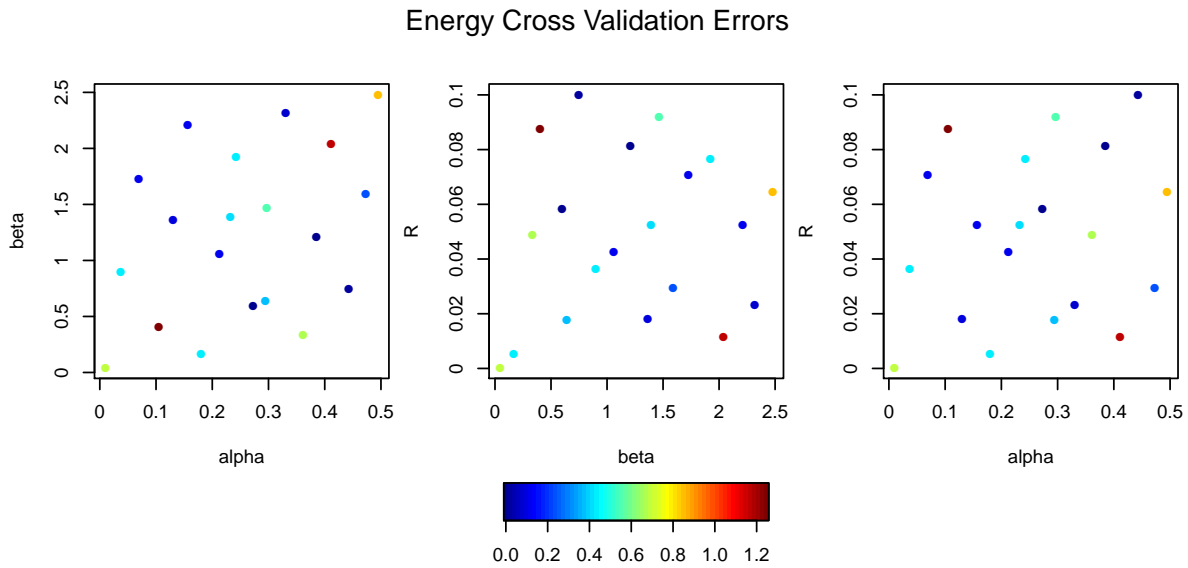

Figure 5.1: 2D slice plots of cross validation errors for energy calibration. The color of each point indicates the value of the cross validation error and the location of the point signifies which model was held out as the observations.

Table 5.1: Cross Validation Errors

| Held Out Model	Energy	Flux	Weighted Energy	Weighted Flux	Random
(0.01, 0.045, 0.0004)	0.8701	1.1522	0.8800	1.1522	
(0.385, 1.208, 0.081)	0.4024	0.6903	0.5655	0.6903	
(0.273, 0.597, 0.058)	0.4011	0.5103	0.4011	0.5457	
(0.473, 1.591, 0.03)	0.5655	0.6735	0.5655	0.6735	
(0.233, 1.392, 0.053)	0.6889	0.7376	0.6889	0.4983	
(0.037, 0.898, 0.036)	0.7014	0.7344	0.7073	0.7344	
(0.069, 1.726, 0.071)	0.4859	0.6941	0.4799	0.6941	
(0.18, 0.169, 0.005)	0.6947	0.8520	0.6947	0.8520	
(0.213, 1.061, 0.043)	0.4821	X	0.6423	X	
(0.243, 1.924, 0.076)	0.7039	0.2551	0.7039	0.2551	
(0.361, 0.333, 0.049)	0.8580	0.5641	0.7493	0.5641	
(0.411, 2.038, 0.012)	1.1935	0.6054	1.1935	0.6054	
(0.105, 0.404, 0.087)	1.2481	0.8088	1.2549	0.8088	
(0.13, 1.364, 0.018)	0.4799	0.7163	0.4799	0.7163	
(0.157, 2.211, 0.052)	0.4887	0.8098	0.4887	0.8098	
(0.297, 1.465, 0.092)	0.7816	0.2650	0.7676	0.2650	
(0.331, 2.317, 0.023)	0.4557	0.7381	0.4690	0.7381	
(0.443, 0.748, 0.1)	0.4225	0.5321	0.4364	0.5593	
(0.495, 2.478, 0.065)	0.9709	0.5727	0.9782	0.5727	
(0.294, 0.64, 0.018)	0.6657	0.4404	0.6657	0.4404	
Cross validation error	0.6780	0.6501	0.6906	0.6408	0.6753
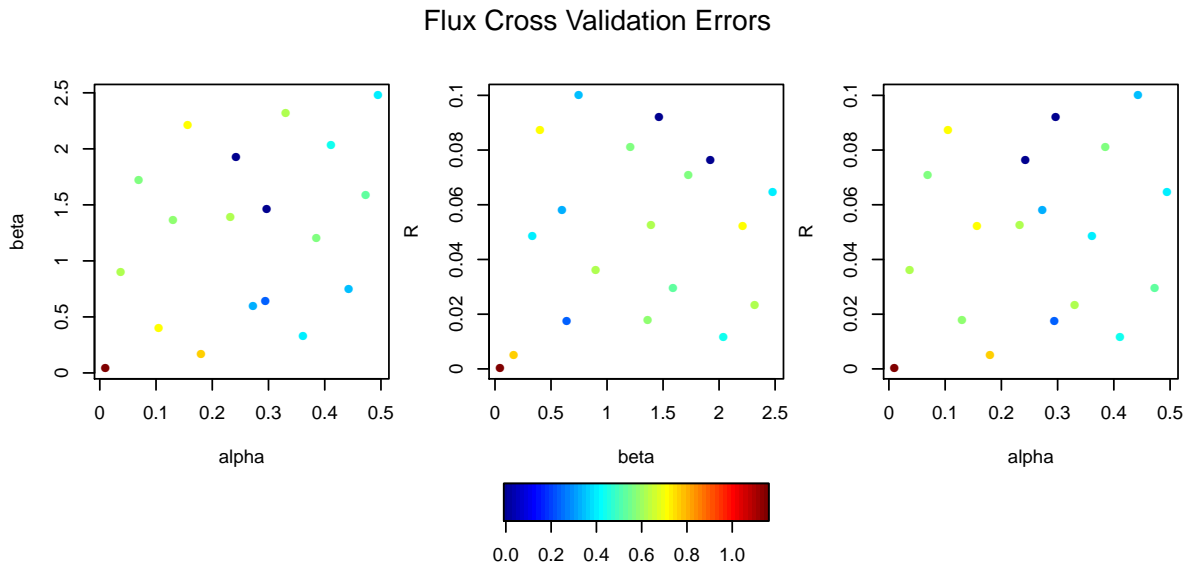
Figure 5.2: 2D slice plots of cross validation errors for flux calibration. The color of each point indicates the value of the cross validation error and the location of the point signifies which model was held out as the observations.

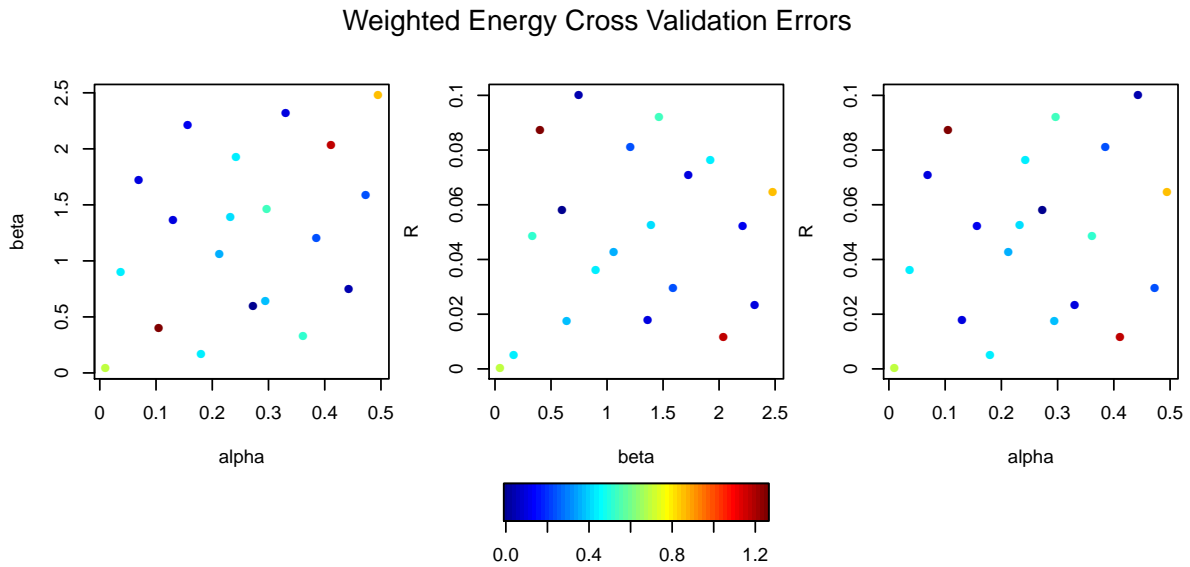Weighted Energy Cross Validation Errors

Figure 5.3: 2D slice plots of cross validation errors for weighted energy calibration. The color of each point indicates the value of the cross validation error and the location of the point signifies which model was held out as the observations.
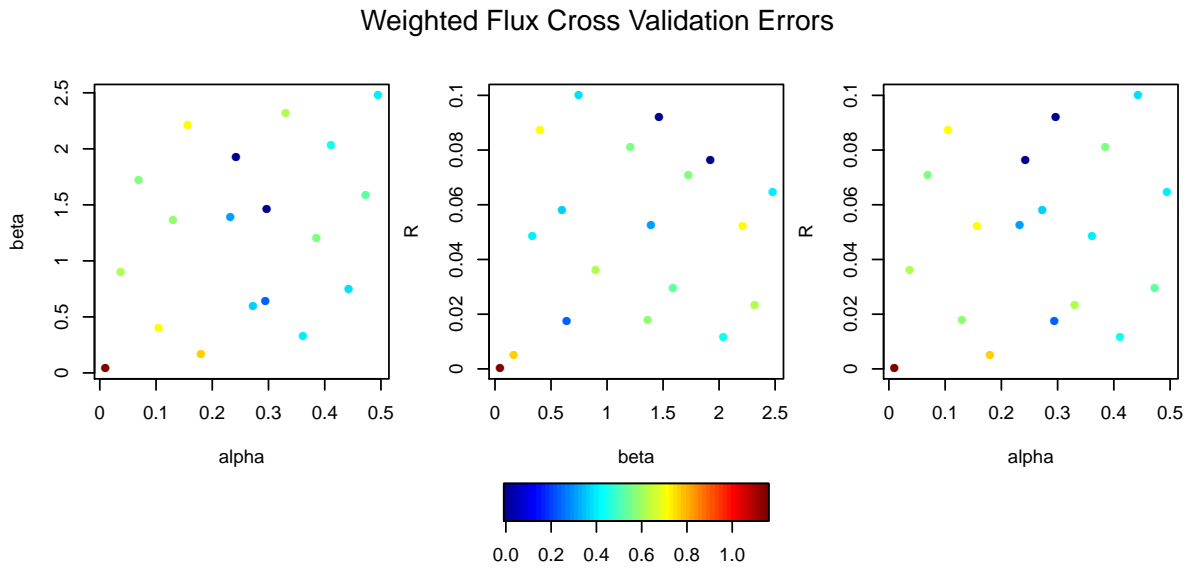
Figure 5.4: 2D slice plots of cross validation errors for weighted flux calibration. The color of each point indicates the value of the cross validation error and the location of the point signifies which model was held out as the observations.

Chapter 6

Conclusion

In this thesis we developed an approach based on statistical calibration to determine the optimal input parameters θ_0 for the LFM-MIX model for geomagnetic storms. High dimensionality of the data, uncertainty in the measurement of the observations, and observations defined on different subsets of the spatial grid presented additional challenges and rendered many of the available methods ineffective. Instead, we took a feature-based approach, meaning that we identified different optimal input settings θ_0 for different features of the data and then compared the results. Following the approach of Heaton et al. in [6], we introduced a time-varying rotational bias to the energy and flux data (in addition to a constant rotational bias). We also considered statistics based on total energy and flux, which are just time series representing the amount of energy (or flux) observed over the entire spatial grid. We attempted smoothing the total energy and flux to better identify the true signal, then introduced a multiplicative bias to better align the LFM-MIX models with the observed total energy and flux.

For each statistic, we applied our method of feature-based calibration, which is based on the popular geostatistical method of kriging. The main idea behind kriging is to predict the value of a random function by computing a weighted average of the known values of the function, which is based on a covariance model derived from the data. We calculated the SSE between the LFM-MIX models and the observations, then used kriging to predict these values at unsampled locations in the three dimensional (α, β, R) space of parameters. Finally,

we took θ_0 to be the minimum of this kriged surface. Then, in an attempt to correct the measurement error, we introduced a sigmoidal weight function to the data and performed calibration on each statistic once again. Finally, to validate our method of feature-based calibration we implemented leave-one-out cross validation. In the interest of time we were only able to perform cross validation for energy, flux, and weighted energy and flux.

We found that the minimums for the flux statistics centered around $(0.29, 0.67, 0.02)$ for unweighted calibration and $(0.29, 0.68, 0.02)$ for weighted calibration, while the energy minimums did not center on any one point. The results of the cross validation experiment suggested that some statistics identified particular regions of the parameter space better than others, but overall weighted flux calibration was the most accurate. Given these results and our analysis of the variograms for each statistic, we conclude that the optimal input parameter setting for the LFM-MIX model is $(0.29, 0.68, 0.02)$.

The clearest route for future research is to apply feature-based calibration to LFM-MIX output from multiple geomagnetic storms. A key question posed by authors before us is: how does LFM-MIX react to different storms? Would our analysis center on the same θ_0 values if we used data from multiple storms? It is beyond the scope of this thesis to answer this question with confidence, as we only had data from one storm available to us. However, we can begin to address this question by comparing our results to those of the authors in [6, 11, 12], who all used data from the same storm which occurred on January 10, 1997. Given that our results support the conclusions of the authors before us, we suspect that our results and those in the literature are not specific to any one storm, but instead generalize to all geomagnetic storms. To answer this question with more confidence, we propose obtaining the data from the 1997 storm, applying feature-based calibration, and comparing the results to those found herein.

Another suggestion for future research is to include other features in our feature-based calibration. For example, we could include a feature describing spatial extent of the geomagnetic storm. This would involve calculating the ratio of points above a certain threshold

to the number of points where the storm was observed. We could also implement a process similar to that of Kleiber et al. in [12] using "landmarks", or spatial features of the data we wish to focus on. For example, we could place four landmarks outlining the approximate boundary of the high energy spatial feature and a fifth at the spatial mode. The next step would be to measure the distance between model landmarks and observation landmarks and apply kriging to these distances.

To improve upon the accuracy of our method, we suggest future researchers obtain more LFM-MIX runs. Rules of thumb in the kriging literature range from a minimum of 20 to 100 points, with the consensus being around 30 [1]. We believe that an additional 10 LFM-MIX runs would significantly increase the reliability of our results. On a similar note, we also suggest obtaining data from a different geomagnetic storm to use in cross validation of the method. Finally, we propose applying both simple kriging and gradient enhanced kriging to the data and comparing their results. Gradient enhanced kriging, as the name suggests, is a special type of kriging that incorporates gradient information in its prediction. We suspect gradient enhanced kriging would increase the accuracy of our results, since assuming the random process is isotropic may be assuming too much.

Bibliography

[1] Standard guide for analysis of spatial variation in geostatistical site investigations. American Society for Testing and Materials (2010). 10.1520/d5922-96r10.

[2] CHILES, J.-P., AND DELFINER, P. Geostatistics: modeling spatial uncertainty, 2nd ed. Wiley, 2012.

[3] CRESSIE, N. Statistics for spatial data. John Wiley Sons, Inc., 1993.

[4] EXPEDITION 29 CREW. Photo of the aurora australis taken from the ISS. https://earthobservatory.nasa.gov/IOTD/view.php?id=52287, Sep 2011.

[5] GIORGINO, T. Computing and visualizing dynamic time warping alignments in R: The dtw package. Journal of Statistical Software 31, 7 (2009), 1–24.

[6] HEATON, M. J., KLEIBER, W., SAIN, S. R., AND WILTBERGER, M. Emulating and calibrating the multiple-fidelity lyon–fedder–mobarry magnetosphere–ionosphere coupled computer model. Journal of the Royal Statistical Society: Series C (Applied Statistics) 64, 1 (2015), 93–113.

[7] HIGDON, D., GATTIKER, J., WILLIAMS, B., AND RIGHTLEY, M. Computer model calibration using high-dimensional output. Journal of the American Statistical Association 103, 482 (2008), 570583.

[8] HIGDON, D., KENNEDY, M., CAVENDISH, J. C., CAFEO, J. A., AND RYNE, R. D. Combining field data and computer simulations for calibration and prediction. SIAM Journal on Scientific Computing 26, 2 (2004), 448–466.

[9] HOLMAN, G. Space weather: What impact do solar flares have on human activities? https://hesperia.gsfc.nasa.gov/sftheory/spaceweather.htm.

[10] KENNEDY, M. C., AND O'HAGAN, A. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, 3 (2001), 425464.

[11] KLEIBER, W., SAIN, S. R., HEATON, M. J., WILTBERGER, M., REESE, C. S., BINGHAM, D., ET AL. Parameter tuning for a multi-fidelity dynamical model of the magnetosphere. The Annals of Applied Statistics 7, 3 (2013), 1286–1310.

[12] Kleiber, W., Sain, S. R., and Wiltberger, M. J. Model calibration via deformation. SIAM/ASA Journal on Uncertainty Quantification 2, 1 (2014), 545–563.

[13] Kozyra, J. U., and Liemohn, M. W. Ring current energy input and decay. Space Science Reviews 109, 1-4 (Oct 2003), 105131.

[14] Lyon, J. G. Numerical methods used in the Lyon-Fedder-Mobarry Global code to model the magnetosphere. TERRAPUB, 2007, p. 101109.

[15] NOAA's National Weather Service. NWS space weather: Storm types. `http://www.nws.noaa.gov/om/space/storm_types.shtml`, Jan 2001.

[16] Nychka, D., Furrer, R., Paige, J., and Sain, S. fields: Tools for spatial data. `www.image.ucar.edu/fields`, 2015. R package version 8.15.

[17] Osborne, C. Statistical calibration: A review. International Statistical Review / Revue Internationale de Statistique 59, 3 (Dec 1991), 309336.

[18] Poppe, B. B., and Jorden, K. P. Sentinels of the sun: forecasting space weather. Johnson Books, 2006.

[19] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[20] Schmitt, P. Lyon Fedder Mobarry global MHD model. https://wiki.ucar.edu/display/LTR/LFM, May 2012.

[21] Schmitt, P. Mix: Magnetosphere ionosphere coupler/solver, Nov 2012.

[22] Soetaert, K. plot3D: Plotting Multi-Dimensional Data, Jan 2016. R package version 1.1.

[23] Stephenson, D. Empirical orthogonal function analysis, Sep 2000.

[24] Toffoletto, F., Sazykin, S., Spiro, R., Wolf, R., and Lyon, J. Rcm meets lfm: initial results of one-way coupling. Journal of Atmospheric and Solar-Terrestrial Physics 66, 15-16 (Sep 2004), 13611370.