Space Weather Forecasting

HARISH J. PALANTHANDALAM-MADAPUSI, DENNIS S. BERNSTEIN, and AARON J. RIDLEY

IDENTIFYING PERIODICALLY SWITCHING BLOCK-STRUCTURED MODELS TO PREDICT MAGNETIC-FIELD FLUCTUATIONS

The magnetosphere is the region of space dominated by the magnetic field of the Earth. The size of the magnetosphere and much of its dynamics are affected by the Sun's atmosphere, which flows supersonically away from the Sun past the Earth and the other planets in the form of solar wind. The Sun's magnetic field, which is commonly referred to as the interplanetary magnetic field (IMF), is embedded in the solar wind [1].

In the absence of solar wind, the Earth's magnetic field resembles that of a dipole magnet with symmetric magnetic field lines as shown by the dashed lines in Figure 1. However, as the solar wind flows past the Earth, the magnetic field lines of the Earth become severely distorted as shown in Figure 1. The magnetospheric region extends out to approximately 10 R_e (Earth radii) on the daytime side of the Earth and more than 40 R_e on the nighttime side of the Earth.

Although the solar wind typically flows at about 400 km/s, large expulsions of plasma from the sun, called coronal mass ejections (CMEs) or solar storms, can attain speeds of over 900 km/s. A typical cross-sectional solar wind distribution is shown in Figure 2, while a CME is shown in Figure 3.

High-energy CMEs, whose intensity varies with the 11-year cycle, interact with the Earth's magnetic field causing geomagnetic storms, which entail ionospheric currents and aurora in both the northern and
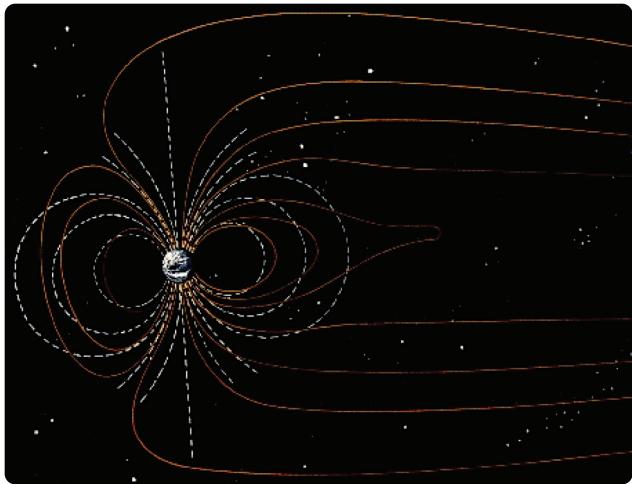
FIGURE 1 The Earth's magnetic field [36]. Due to the solar wind, the Earth's dipole-like magnetic field is distorted and stretched toward the nighttime side. The dashed lines indicate the dipole-like magnetic field lines in the absence of the solar wind, while the solid lines indicate the distorted magnetic field lines due to the solar wind. Image used with permission.

southern polar regions. Most of the time, the aurora and the ionospheric currents are minimal and have a negligible effect on technology. When a CME encounters the magnetosphere, however, the aurora and ionospheric currents increase, heating and expanding the upper atmosphere, which causes increased drag on satellites. In addition, large ionospheric current fluctuations can induce currents in power lines, which can overwhelm and destroy trans-

formers and electrical networks [2]. Additionally, as discussed in [3], solar storms can adversely affect animals, humans, and aircraft. It is therefore essential to be able to predict when and where large ionospheric current fluctuations are likely to occur.

Ultimately, the Sun controls the ionospheric currents and the aurora. When the Sun's atmosphere is calm, there is little aurora and small currents. However, large ejections of magnetic and plasma energy cause large disturbances in the ionosphere. To predict these large ejections of magnetic and plasma energy, satellites monitor the solar surface and image these events. However, these image measurements are not sufficient to accurately predict the subsequent interplanetary conditions. Therefore, researchers must wait until the ejections reach most of the way to the Earth, where additional measurements of the solar wind and IMF conditions are made. These additional measurements are made by the Advanced Composition Explorer (ACE) satellite, which orbits the Lagrangian gravitational null point between the Sun and the Earth. Further details about ACE and its orbit are discussed in "Advanced Composite Explorer." The solar wind measurements are made by ACE 30–90 min before the solar wind encounters the magnetopause, which is the boundary of the magnetosphere. Note that the data are available on Earth almost immediately due to the speed of radio propagation. However, because of the transport delay, which depends on the solar wind velocity component V_x, the solar wind measured by ACE at time t reach the magnetopause at time $t + D_1/V_x$, where $D_1 \approx 1.5e6$ km is the distance between ACE and the magnetopause, as illustrated
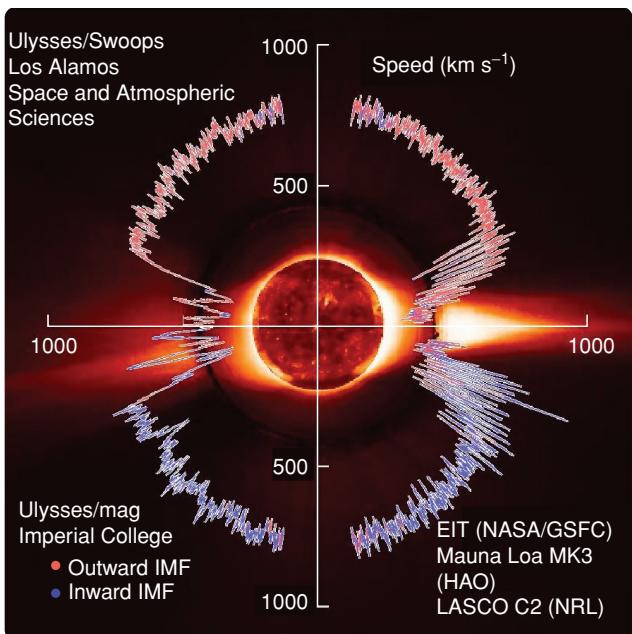
FIGURE 2 Solar wind [1]. In this cross-sectional view, the solar-wind distribution around the sun is shown along with the orientation of the IMF. The solar wind in some regions is seen to exceed 900 km/s. Image used with permission.
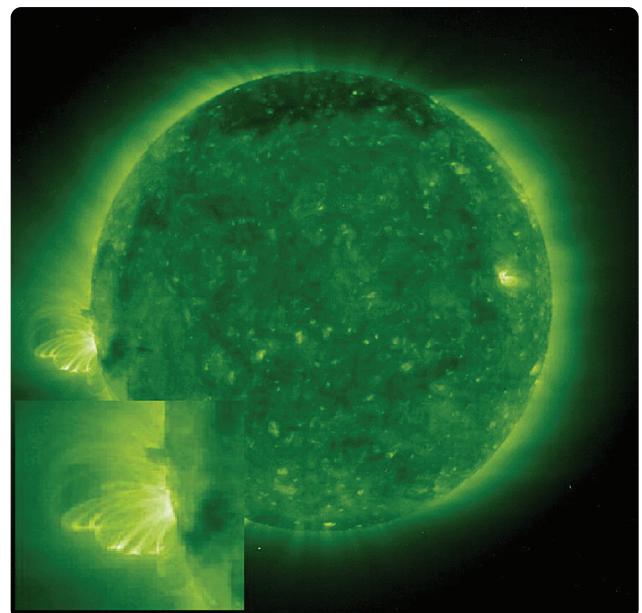
FIGURE 3 Coronal mass ejection [1]. This photograph shows the Sun's surface with a CME on the left side of the photograph. The CME is magnified in the insert. Large quantities of charged plasma along with magnetic energy associated with the plasma are released from the Sun's surface during a CME. Image used with permission.

in Figure 4. Consequently, this delay allows us to construct empirical models that have an input-driven predictive capability of 30–90 min before the solar wind reaches the magnetosphere. Note that the prediction horizon is determined by the solar wind velocity and this is outside of our control.

The magnetic perturbations caused by the interaction of the solar wind and the magnetosphere are measured on the Earth by ground-based magnetometer stations. The data used in this article are measurements made by ground-based magnetometers located at Thule (THL) in Greenland, Sondre Stromfjord (STF) in Greenland, and Kotelny (KTN) in Russia. The locations of these ground-based magnetometer stations are given in Table 1. Thule is a high-latitude station, Kotelny is a mid-latitude station, and Sondre Stromfjord is a low-latitude station. However, since the magnetic axis of the Earth differs from the geographic axis by about $11°$, in terms of the magnetic latitudes measured from the magnetic pole Thule is a high-latitude station, Kotelny is a low-latitude (auroral-zone) station, and Sondre Stromfjord is in between Thule and Kotelny. For reasons explained below, the dynamics of the magnetosphere are more complex at auroral latitude stations.

The size of the daytime-side portion of the magnetosphere is directly proportional to the solar wind speed, whereas the size of the nighttime-side portion, called the magnetosphere tail, fluctuates severely due to the storage and release of
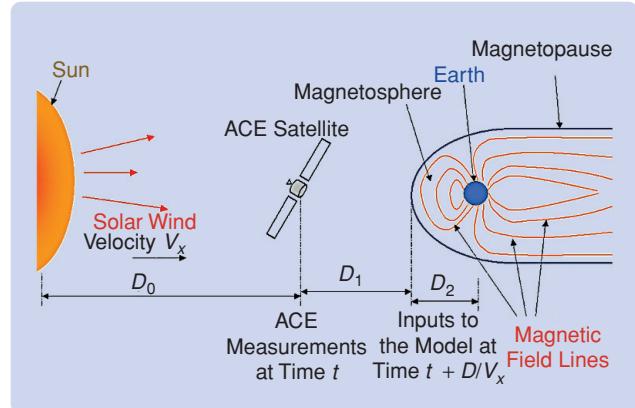
FIGURE 4 The Advanced Composition Explorer (ACE) satellite, magnetosphere, and time delay. The position of the ACE satellite is shown along with the time delay between the ACE measurements and the inputs to the model. The magnetopause (dark blue line) encloses the Earth's magnetosphere. The magnetic field lines (orange lines) are distorted due to the solar wind as explained in Figure 1. The ACE satellite orbits the first Lagrangian point, which is located at a distance D_1 (not drawn to scale) from the magnetopause. The solar wind flowing past the ACE satellite, whose velocity, density, magnetic field, and temperature are measured by ACE at time t, reaches the magnetosphere at time $t + D_1/V_x$, where D_1/V_x ranges from 30 to 90 min. Since the ACE measurements are available almost instantaneously, the identified model can predict conditions on the Earth between 30 and 90 min into the future.

Advanced Composition Explorer

The Advanced Composition Explorer (ACE) spacecraft [37] is an Explorer Mission managed by the Office of Space Science Mission and Payload Development Division of the National Aeronautics and Space Administration (NASA). The ACE spacecraft, which was launched on August 25, 1997, orbits the first Lagrangian null point (L1), which is a point of Earth-Sun gravitational equilibrium. L1 is approximately 1.5 million km from the Earth and approximately 148 million km from the Sun. By orbiting L1, which is outside the Earth's magnetosphere, ACE stays in a fairly constant position with respect to the Earth as the Earth revolves around the Sun.

The elliptical orbit of ACE has a major axis of roughly 6×10^5 km, which subtends an angle of $0.115°$ from the center of the Sun. At the ACE orbit's extremities, the satellite is roughly 3×10^5 km away from the Sun-Earth axis. However, since the

scale of the solar events is much larger than this distance [38], it is reasonable to assume that the solar wind conditions measured by ACE are the same as those encountered by the Earth.

The ACE satellite is 1.6-m long and 1-m high, not including the four solar arrays and magnetometer booms attached to two of the solar panels. The satellite spins at 5 rpm, with the spin axis pointed along the Earth-Sun line. Most of the scientific instruments are mounted on the top (sunward) deck.

The ACE satellite carries six high-resolution sensors and three monitoring instruments. From a vantage point of approximately 1% of the distance from the Earth to the Sun, ACE performs measurements over a wide range of energy and nuclear mass under all solar wind flow conditions. By continually monitoring the Sun, ACE provides real-time solar wind data and thus can provide a warning of solar storms about 30 to 90 min before they reach the Earth.

TABLE S1 Location of the ground-based magnetometers. Real-time Web-based prediction for these ground-based magnetometer stations are provided on the Web site [40]. Predictions for an additional 28 magnetometer stations are expected to be added. Magnetometers located at these stations measure magnetic perturbations caused by the solar wind and solar storms.

Magnetometer station	Symbol	Latitude	Longitude	Country
Island Lake	ISL	53.86 N	265.34 E	Canada
Tromsø	TRO	69.66 N	18.94 E	Norway
Longyearbyen	LYR	78.20 N	15.82 E	Norway
Dawson	DAW	64.05 N	220.89 E	Canada

magnetic energy. When magnetic energy is released in the magnetospheric tail, constituting a substorm [4], currents and high-energy particles are driven along field lines toward the Earth. The aurora results from these high-energy particles bombarding the atmosphere, while the currents cause large magnetic perturbations. Thus, magnetic field lines associated with the auroral zone map to locations in the magnetosphere where the storage and release cycles occur. The three magnetometer stations are thus governed by different dynamics. In particular, because of the storage-and-release cycle in the magnetosphere tail, the dynamics governing auroral-zone stations are more complex than at higher latitudes and thus more difficult to model. Moreover, the auroral-zone stations experience alternating periods of high-latitude dynamics and auroral-zone dynamics, depending on the size and shape of the magnetosphere. This variation significantly complicates the identification of dynamics governing auroral-zone stations.

The goal of this article is to use system identification methods to construct models that can be used to predict magnetic-field fluctuations. The inputs to the model are the solar wind conditions measured by ACE, while the outputs of the model are measurements made by ground-based magnetometer stations. By predicting ground-based magnetometer response using ACE data, our objective is to obtain advance warning of future disturbances. These warnings can be used to take steps to minimize damage to sensitive infrastructure [2].

EMPIRICAL MODELING OF THE MAGNETOSPHERE

The nonlinear partial differential equations governing the magnetospheric system involve exogenous drivers and feedback mechanisms [5], [6]. Since these first-principles models are computationally expensive, real-time implementation restricts the spatial and temporal resolution. Although empirical models are simplistic compared to first principles models, empirical models are useful for predicting specific quantities at specific locations. Empirical models developed for understanding and predicting magnetic-field fluctuations include neural network models

[7], time-series models [8], and statistical models [9].

In this article we identify Hammerstein-Wiener models of the magnetosphere by using ACE measurements as inputs and ground-based magnetometer data as outputs. The Hammerstein-Wiener model structure, which is shown in Figure 5, consists of linear dynamics, static input nonlinearities, and static output nonlinearities [10]–[20].

We make no claim that the Hammerstein-Wiener model structure can model all of the features of the magnetospheric system. However, knowledge of physics suggests that nonlinear functions of inputs drive the magnetospheric system. Moreover, our tests indicate that Hammerstein-Wiener models are useful for predicting magnetic-field fluctuations.

Although the present article focuses on Hammerstein-Wiener models, alternative identification methods of potential interest for the magnetospheric system include techniques for identifying linear parameter-varying (LPV) models and bilinear models [21]–[24]. Preliminary testing with these methods suggests that they are also useful for magnetospheric prediction. Comparisons with these methods are left for future work.

Since measurements of magnetospheric conditions are made by ground-based magnetometers, which rotate with a one-day periodicity with respect to the magnetosphere, the system has one-day periodicity. Subspace identification for discrete-time periodic systems is developed in [25], [26]. In particular, [25] uses lifting to recast the periodically time-varying identification problem as a set of linear time-invariant identification problems. In [26] an ensemble of data sets is used to identify a set of state-space matrices for each time step. For a time-varying periodic system with a period of ρ sample intervals, these methods require the calculation and storage of ρ state-space matrices. In the magnetospheric system, which has one-day periodicity with measurements sampled at every minute, 1440 sets of state-space matrices must be calculated and stored, while using decimated data every 15 min entails calculation and storage of 96 sets of state-space matrices. Since this approach is not computationally tractable, we develop an alternative approach wherein we identify periodically switching models that switch among a small number (six or less) of state-space matrices. We thus identify periodically switching Hammerstein-Wiener models to capture the time-varying nature of the system.

In this article, we consider subspace-based identification methods for periodically switching Hammerstein-Wiener models of the magnetospheric system. For this application, we develop and apply a Hammerstein-Wiener identification
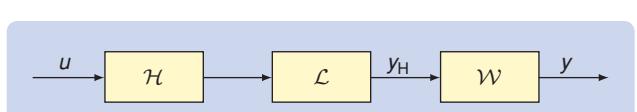
FIGURE 5 Hammerstein-Wiener model. The Hammerstein-Wiener model structure consists of a Hammerstein input nonlinearity $\mathcal{H}$ connected to a linear dynamical system $\mathcal{L}$ followed by a Wiener output nonlinearity $\mathcal{W}$. $\mathcal{H}$ and $\mathcal{W}$ are static multi-input, multi-output maps.

method, which is an extension of [27]–[29], involving optimizing basis functions iteratively. By comparing the prediction from these identified models with data measured subsequently, we validate the identified models based on their prediction performance.

An alternative approach to identifying periodically switching Hammerstein-Wiener models is to use the Hammerstein identification algorithm based on least-squares support vector machines (LS-SVM) [30]. Although these methods use basis-function expansions to represent the nonlinear maps, the actual basis functions need not be specified, instead the kernel functions, which are inner products of the basis functions, are chosen by the user.

A challenging aspect of real data, such as data from the magnetospheric and ionospheric system, is the occurrence of missing data points. Data may be missing due to instrument malfunction, environmental circumstances, excessive noise, or defective records. To address this problem, we describe a modified subspace algorithm that accommodates missing data points. Although this modification is straightforward, it is of great practical importance and is apparently not discussed in the literature.

It turns out that the technique for dealing with missing data is the key to identifying periodically switching models. Specifically, we partition each day into subintervals, and then we separately consider each repeating subinterval. In particular, we view all of the data except within the specified subinterval as missing, and we use the non-missing data to identify a time-invariant Hammerstein-Wiener model for the corresponding portion of the day. By partitioning the day into a small number of subintervals, we obtain a periodically switching model that requires the calculation and storage of fewer sets of state-space matrices than in [25] and [26]. We use this periodically switching identification method with the Hammerstein-Wiener identification algorithm to identify periodically switching Hammerstein-Wiener models for the magnetospheric system. A schematic of the various components of the identification scheme is shown in Figure 6.

HAMMERSTEIN-WIENER IDENTIFICATION
Consider the system

$$x_{k+1} = Ax_k + \mathcal{H}(u_k) + w_k, \tag{1}$$
$$y_k = \overline{\mathcal{W}}(x_k) + v_k, \tag{2}$$
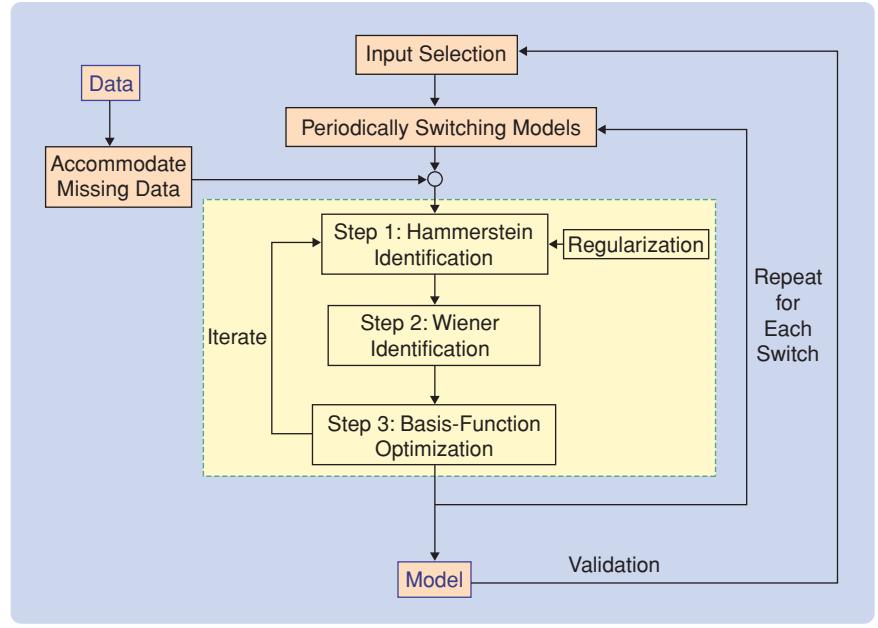
FIGURE 6 A flowchart showing the various steps in the identification scheme. Steps 1, 2, and 3 constitute the basic identification algorithm. Step 1 uses regularization, while steps 1–3 are iterated for optimized sets of basis functions. The procedure inside the shaded region is repeated for each set of state-space matrices in the periodically switching Hammerstein-Wiener model. This entire procedure is repeated with different input combinations, and the prediction performance is evaluated to determine the most relevant inputs.

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^l$, $A \in \mathbb{R}^{n \times n}$, $\mathcal{H} : \mathbb{R}^m \to \mathbb{R}^n$ is the Hammerstein nonlinear map, and $\overline{\mathcal{W}} : \mathbb{R}^n \to \mathbb{R}^l$ is the Wiener nonlinear map. We assume that measurements of u_k and y_k are available, A, $\mathcal{H}$, and $\overline{\mathcal{W}}$ are unknown, and $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^l$ are unknown zero-mean Gaussian white noise sequences. To estimate the state-space matrices and nonlinear maps, we present an identification method based on subspace algorithms and basis-function optimization. Based on this algorithm, periodically switching Hammerstein-Wiener models are constructed below for the magnetospheric system.

For convenience, (2) is decomposed as

$$y_{Hk} = Cx_k, \tag{3}$$
$$y_k = \mathcal{W}(y_{H,k}) + v_k, \tag{4}$$

where $y_{H,k} \in \mathbb{R}^{l_H}$, $C \in \mathbb{R}^{l_H \times n}$, $\mathcal{W} : \mathbb{R}^{l_H} \to \mathbb{R}^l$, and $l_H \geq l$. Equations (1) and (3), which represent the first two blocks in Figure 5, comprise the Hammerstein subsystem. Equation (4) represents the last block in Figure 5, which is the Wiener nonlinearity. The signal $y_{H,k}$ is the output of the Hammerstein subsystem. The identification method described in this section is a multistep method in which the Hammerstein subsystem is identified in the first step, the Wiener nonlinearity is identified in the second step, and the basis functions used to represent the nonlinearity are optimized in the third step.

Sequential Hammerstein-Wiener Identification with Basis-Function Optimization

Step One: Hammerstein Subsystem Identification

Consider the Hammerstein subsystem (1), (3). The Hammerstein nonlinearity $\mathcal{H}$ can be written in terms of its scalar-valued components as $\mathcal{H} = [\mathcal{H}_1 \cdots \mathcal{H}_n]^{\mathrm{T}}$, where, for $i = 1, \ldots, n, \mathcal{H}_i \colon \mathbb{R}^m \to \mathbb{R}$. Next, as in [27], we assume that each component $\mathcal{H}_i$ can be expanded in terms of basis functions $f_1, \ldots, f_r$ as

$$\mathcal{H} = \begin{bmatrix} \sum_{j=1}^{r} b_{1j} f_j \\ \vdots \\ \sum_{j=1}^{r} b_{nj} f_j \end{bmatrix}, \tag{5}$$

where the b_{ij}'s are coefficients of the basis-function expansions. Next, defining $f \colon \mathbb{R}^m \to \mathbb{R}^r$ as $f \triangleq [f_1 \cdots f_r]^{\mathrm{T}}$, it follows from (5) that

$$\mathcal{H}(u) = Bf(u), \tag{6}$$

where $B \triangleq [b_{ij}] \in \mathbb{R}^{n \times r}$. Thus (1), (3) become

$$x_{k+1} = Ax_k + Bf(u_k) + w_k, \tag{7}$$

$$y_{\mathrm{H},k} = Cx_k. \tag{8}$$

Once basis functions are chosen, the identification problem can be viewed as a linear identification problem with the generalized input $f(u_k) \in \mathbb{R}^r$. A linear subspace algorithm such as N4SID [31] can now be used to estimate the state-space matrices and noise covariances. However, since $y_{\mathrm{H},k}$ is not measured, we use measurements of y_k for the identification of the Hammerstein subsystem.

Step Two: Wiener Identification

After the Hammerstein subsystem is identified, the Wiener nonlinearity $\mathcal{W}$ is estimated by solving a linear least squares problem. The function $\mathcal{W}$ is written in terms of its scalar-valued components as $\mathcal{W} = [\mathcal{W}_1 \cdots \mathcal{W}_l]^{\mathrm{T}}$, where, for $i = 1, \ldots, l, \mathcal{W}_i \colon \mathbb{R}_{\mathrm{H}}^l \to \mathbb{R}$. We assume that each component $\mathcal{W}_i$ can be expanded in terms of basis functions $g_1, \ldots, g_s$ as

$$\mathcal{W} = \begin{bmatrix} \sum_{j=1}^{s} \lambda_{1j} g_j \\ \vdots \\ \sum_{j=1}^{s} \lambda_{lj} g_j \end{bmatrix}, \tag{9}$$

where the λ_{ij}'s are coefficients of the basis-function expansion. Next, we define $g \colon \mathbb{R}_{\mathrm{H}}^l \to \mathbb{R}^s$ as $g \triangleq g_1 \cdots g_s]^{\mathrm{T}}$, and let $y \triangleq [y_1 \cdots y_N] \in \mathbb{R}^{l \times N}$ and $\psi \triangleq [g(\hat{y}_{\mathrm{H},1}) \cdots g(\hat{y}_{\mathrm{H,N}})] \in \mathbb{R}^{s \times N}$, where N is the number of available data points. Now, to estimate the Wiener nonlinearity we solve the least squares problem

$$\underset{\Lambda}{\mathrm{argmin}} \left\| y - \Lambda \psi \right\|, \tag{10}$$

where $\Lambda \triangleq [\lambda_{ij}] \in \mathbb{R}^{l \times s}$. The least-squares solution of (10) is

$$\Lambda = y \psi^{\dagger}, \tag{11}$$

where † denotes the Moore-Penrose generalized inverse.

Step Three: Basis-Function Optimization

A convenient choice of basis functions for f and g are radial basis functions, which can handle arguments of arbitrary dimension. For example, Laplacian, logistic, Gaussian, and thin-plate-spline radial basis functions have the form

$$f_j(u) = e^{-\alpha_j \|u - c_j\|_2}, \tag{12}$$

$$f_j(u) = \frac{1}{1 + e^{-\alpha_j \|u - c_j\|_2}}, \tag{13}$$

$$f_j(u) = e^{-\alpha_j \|u - c_j\|_2^2}, \tag{14}$$

$$f_j(u) = \|u - c_j\|_2^2 \log(\alpha_j \|u - c_j\|_2), \tag{15}$$

respectively. The parameters α_j and c_j determine the spread and center, respectively, of f_j. Similarly, the basis function g_j is defined in terms of the spread β_j and center d_j.

After the initial Hammerstein-Wiener identification (steps one and two), we optimize the radial basis functions with respect to the parameters α_j, c_j, β_j, and d_j. Letting $\alpha \triangleq [\alpha_1 \cdots \alpha_r]^{\mathrm{T}}$, $c \triangleq [c_1 \cdots c_r]^{\mathrm{T}}$, $\beta \triangleq [\beta_1 \cdots \beta_s]^{\mathrm{T}}$, and $d \triangleq [d_1 \cdots d_s]^{\mathrm{T}}$, the identification error is defined to be the mean-square error

$$E(\alpha, c, \beta, d) \triangleq \frac{1}{2} \sum_{k=1}^{N} (y_k - \hat{y}_k)^{\mathrm{T}} (y_k - \hat{y}_k), \tag{16}$$

where y_k is the output of the true system, $\hat{y}_k$ is the output of the identified Hammerstein-Wiener model, and N is the length of the data set.

TABLE 2 Gradient expressions for the Laplacian and logistic radial basis functions. By using these gradient expressions, we optimize the mean-square error E given by (16) with respect to the basis-function parameters α, c, β, and d. The type of basis functions chosen for the components of f and g may be different.

Gradient	Laplacian	Logistic
$\dfrac{\partial g_j}{\partial \hat{y}_{\mathrm{H},k}}$	$\beta_j e^{-\beta_j \|\hat{y}_{\mathrm{H},k} - d_j\|_2} \dfrac{(d_j - \hat{y}_{\mathrm{H},k})^{\mathrm{T}}}{\|\hat{y}_{\mathrm{H},k} - d_j\|_2}$	$\dfrac{\beta_j e^{-\beta_j \|\hat{y}_{\mathrm{H},k} - d_j\|_2}}{(1 + e^{-\beta_1 \|\hat{y}_{\mathrm{H},k} - d_1\|_2})^2} \dfrac{(d_j - \hat{y}_{\mathrm{H},k})^{\mathrm{T}}}{\|\hat{y}_{\mathrm{H},k} - d_j\|_2}$
$\dfrac{\partial g_j}{\partial \beta_j}$	$-e^{-\beta_j \|\hat{y}_{\mathrm{H},k} - d_j\|_2} \|\hat{y}_{\mathrm{H},k} - d_j\|_2$	$\dfrac{-\beta_j e^{-\beta_1 \|\hat{y}_{\mathrm{H},k} - d_j\|_2}}{(1 + e^{-\beta_j \|\hat{y}_{\mathrm{H},k} - d_j\|_2})^2} \|\hat{y}_{\mathrm{H},k} - d_j\|_2$
$\dfrac{\partial g_j}{\partial d_j}$	$\beta_j e^{-\beta_j \|\hat{y}_{\mathrm{H},k} - d_j\|_2} \dfrac{(\hat{y}_{\mathrm{H},k} - d_j)^{\mathrm{T}}}{\|\hat{y}_{\mathrm{H},k} - d_j\|_2}$	$\dfrac{\beta_j e^{-\beta_j \|\hat{y}_{\mathrm{H},k} - d_j\|_2}}{(1 + e^{-\beta_j \|\hat{y}_{\mathrm{H},k} - d_j\|_2})^2} \dfrac{(\hat{y}_{\mathrm{H},k} - d_j)^{\mathrm{T}}}{\|\hat{y}_{\mathrm{H},k} - d_j\|_2}$
$\dfrac{\partial f_j}{\partial \alpha_j}$	$-e^{-\alpha_j \|u_i - c_j\|_2} \|u_i - c_j\|_2$	$\dfrac{-\alpha_j e^{-\alpha_j \|u_i - c_j\|_2}}{(1 + e^{-\alpha_j \|u_i - c_j\|_2})^2} \|u_i - c_j\|_2$
$\dfrac{\partial f_j}{\partial c_j}$	$\alpha_j e^{-\alpha_j \|u_i - c_j\|_2} \dfrac{(\hat{y}_{\mathrm{H},k} - c_j)^{\mathrm{T}}}{\|u_i - c_j\|_2}$	$\dfrac{\alpha_j e^{-\alpha_j \|u_i - c_j\|_2})^2}{(1 + e^{-\alpha_j \|u_i - c_j\|_2})^2} \dfrac{(u_i - c_j)^{\mathrm{T}}}{\|u_i - c_j\|_2}$

TABLE 3 Gradient expressions for the Gaussian and thin-plate spline radial basis functions. By using these gradient expressions, we optimize the mean-square error E given by (16) with respect to the basis-function parameters α, **c**, β, and **d**.

Gradient	Gaussian	Thin-Plate Spline
$\dfrac{\partial g_j}{\partial \hat{y}_{H,k}}$	$2\beta_j e^{-\beta_j\|\hat{y}_{H,k}-d_j\|_2^2}(d_j-\hat{y}_{H,k})^\mathsf{T}$	$(\hat{y}_{H,k}-d_j)^\mathsf{T}[1/\beta_j+2\log(\beta_j\|\hat{y}_{H,k}-d_j\|_2)]$
$\dfrac{\partial g_j}{\partial \beta_j}$	$-e^{-\beta_j\|\hat{y}_{H,k}-d_j\|_2^2}\|\hat{y}_{H,k}-d_j\|_2^2$	$\dfrac{\|\hat{y}_{H,k}-d_j\|_2^2}{\beta_j}$
$\dfrac{\partial g_j}{\partial d_j}$	$2\beta_j e^{-\beta_j\|\hat{y}_{H,k}-d_j\|_2^2}(\hat{y}_{H,k}-d_j)^\mathsf{T}$	$(d_j-\hat{y}_{H,k})^\mathsf{T}[1/\beta_j+2\log(\beta_j\|\hat{y}_{H,k}-d_j\|_2)]$
$\dfrac{\partial f_j}{\partial \alpha_j}$	$-e^{-\alpha_j\|u_i-c_j\|_2^2}\|u_i-c_j\|_2^2$	$\dfrac{\|u_i-c_j\|_2^2}{\alpha_j}$
$\dfrac{\partial f_j}{\partial c_j}$	$2\alpha_j e^{-\alpha_j\|u_i-c_j\|_2^2}(u_i-c_j)^\mathsf{T}$	$(c_j-u_i)^\mathsf{T}[1/\alpha_j+2\log(\alpha_j\|u_i-c_j\|_2)]$

Let $\hat{A}$, $\hat{B}$, and $\hat{C}$ be the identified state space matrices for the Hammerstein-Wiener model, and let $\hat{y}_{H,k}$ be the output of the identified Hammerstein subsystem. Writing (16) in terms of g we have

$$E(\alpha, c, \beta, d) = \frac{1}{2}\sum_{k=1}^{N}(y_k - \Lambda g(\hat{y}_{H,k}))^\mathsf{T}(y_k - \Lambda g(\hat{y}_{H,k})), \quad (17)$$

with

$$\hat{y}_{H,k} = \hat{C}\hat{A}^k\hat{x}_0 + \sum_{i=0}^{k-1}\hat{C}\hat{A}^{k-i-1}\hat{B}f(u_i). \quad (18)$$

For $j = 1, \ldots, r$, the gradients of $E(\alpha, c, \beta, d)$ with respect to α_j and c_j are given by

$$\frac{\partial E}{\partial \alpha_j} = \sum_{k=1}^{N}(\Lambda g(\hat{y}_{H,k}) - y_k)^\mathsf{T}\Lambda \begin{bmatrix} \frac{\partial g_1}{\partial \hat{y}_{H,k}} \\ \vdots \\ \frac{\partial g_s}{\partial \hat{y}_{H,k}} \end{bmatrix}$$
$$\times \sum_{i=0}^{k-1}\hat{C}\hat{A}^{k-i-1}\hat{B}\hat{e}_{r,j}\frac{\partial f_j}{\partial \alpha_j} \quad (19)$$

and

$$\frac{\partial E}{\partial c_j} = \sum_{k=1}^{N}(\Lambda g(\hat{y}_{H,k}) - y_k)^\mathsf{T}\Lambda \begin{bmatrix} \frac{\partial g_1}{\partial \hat{y}_{H,k}} \\ \vdots \\ \frac{\partial g_s}{\partial \hat{y}_{H,k}} \end{bmatrix}$$
$$\times \sum_{i=0}^{k-1}\hat{C}\hat{A}^{k-i-1}\hat{B}\hat{e}_{r,j}\frac{\partial f_j}{\partial c_j}, \quad (20)$$

where $\hat{e}_{r,j}$ is the jth column of the $r \times r$ identity matrix. Similarly, for $j = 1, \ldots, s$, the gradients of $E(\alpha, c, \beta, d)$ with respect to β_j and d_j are given by

$$\frac{\partial E}{\partial \beta_j} = \sum_{k=1}^{N}(\Lambda g(\hat{y}_{H,k}) - y_k)^\mathsf{T}\Lambda \hat{e}_{s,j}\frac{\partial g_j}{\partial \beta_j} \quad (21)$$

and

$$\frac{\partial E}{\partial d_j} = \sum_{k=1}^{N}(\Lambda g(\hat{y}_{H,k}) - y_k)^\mathsf{T}\Lambda \hat{e}_{s,j}\frac{\partial g_j}{\partial d_j}. \quad (22)$$

In (19)–(22), the partial derivatives $(\partial g_j/\partial \hat{y}_{H,k})$, $(\partial g_j/\partial \beta_j)$, $(\partial g_j/\partial d_j)$, $(\partial f_j/\partial \alpha_j)$, and $(\partial f_j/\partial c_j)$ depend on the basis functions chosen for f and g. These expressions for Laplacian and logistic radial basis functions are given in Table 2, while the expressions for Gaussian and thin-plate spline basis functions are given in Table 3. Each basis function f_j or g_j can be chosen to be one of the four types of basis functions (12)–(15).

By using the gradient expressions (19)–(22) along with tables 2 and 3, a BFGS quasi-Newton optimization code is used to optimize $E(\alpha, c, \beta, d)$ with respect to the
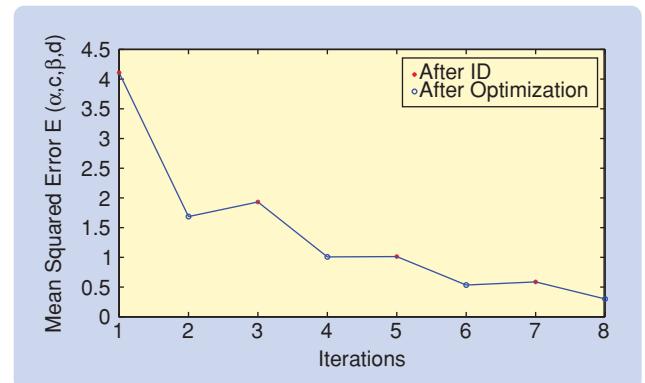
FIGURE 7 Cost versus number of iterations for iterative basis-function optimization. The red asterisks denote the cost after subspace-based Hammerstein-Wiener identification, while the blue circles denote the cost after the basis-function optimization. The decrease in the cost function is not monotonic due to the fact that the basis-function optimization procedure and subspace identification algorithm minimize different cost functions. Subspace identification minimizes the least squares error between the true states and the estimated states, whereas basis-function optimization minimizes the mean-square error at the output $E(\alpha, c, \beta, d)$.

basis-function parameters. Since the state-space matrices and the basis-function parameters are not estimated simultaneously, basis-function optimization and subspace-based Hammerstein-Wiener identification are performed alternately.

Convergence is not guaranteed for the iterative procedure described above. However, the iterative procedure converged in all of our tests. Figure 7 shows a typical plot of the cost function $E(\alpha, c, \beta, d)$ versus the number of iterations. The decrease in the cost function is not monotonic due to the fact that the basis-function optimization procedure and subspace identification algorithm minimize different cost functions.

SUBSPACE IDENTIFICATION WITH MISSING DATA

In this section, we show how subspace algorithms can be modified to accommodate missing data points. Consider the linear system

$$x_{k+1} = Ax_k + Bu_k + w_k, \tag{23}$$

$$y_k = Cx_k + Du_k + v_k, \tag{24}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, and $y_k \in \mathbb{R}^l$. To estimate A, B, C, and D in (23) and (24), we review results from [32]. The following definition is equivalent to Definition 1 in Section 1.4.2 of [32].

Definition 1
Let $P \in \mathbb{R}^{a \times d}$, $Q \in \mathbb{R}^{b \times d}$, and $R \in \mathbb{R}^{c \times d}$, such that $d > \max\{a, b, c\}$. Then *the oblique projection of the row space of P along the row space of Q onto the row space of R is defined as*

$$P/_Q R \triangleq P \begin{bmatrix} R \\ Q \end{bmatrix}^{\dagger} \begin{bmatrix} R \\ 0 \end{bmatrix}. \tag{25}$$

Next, let N be the number of measurements and let i be an integer such that $n \le i$ and $2i - 1 < N$. Define $Y_{0|2i-1} \in \mathbb{R}^{2li \times (N-2i+1)}$, $Y_p \in \mathbb{R}^{li \times (N-2i+1)}$, and $Y_f \in \mathbb{R}^{li \times (N-2i+1)}$ by

$$Y_{0|2i-1} \triangleq \begin{bmatrix} y_0 & y_1 & \cdots & y_{N-2i} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & y_i & \cdots & y_{N-i-1} \\ \hline y_i & y_{i+1} & \cdots & y_{N-i} \\ y_{i+1} & y_{i+2} & \cdots & y_{N-i+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{2i-1} & y_{2i} & \cdots & y_{N-1} \end{bmatrix} \tag{26}$$

$$= \begin{bmatrix} Y_{0|i-1} \\ \hline Y_{i|2i-1} \end{bmatrix} = \begin{bmatrix} Y_p \\ \hline Y_f \end{bmatrix}. \tag{27}$$

Alternatively,

$$Y_{0|2i-1} = \begin{bmatrix} y_0 & y_1 & \cdots & y_{N-2i} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & y_i & \cdots & y_{N-i-1} \\ \hline y_i & y_{i+1} & \cdots & y_{N-i} \\ y_{i+1} & y_{i+2} & \cdots & y_{N-i+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{2i-1} & y_{2i} & \cdots & y_{N-1} \end{bmatrix} \tag{28}$$

$$= \begin{bmatrix} Y_{0|i} \\ \hline Y_{i+1|2i-1} \end{bmatrix} = \begin{bmatrix} Y_{p^+} \\ \hline Y_{f^-} \end{bmatrix}, \tag{29}$$

where $Y_{p^+} \in \mathbb{R}^{l(i+1) \times (N-2i+1)}$ and $Y_{f^-} \in \mathbb{R}^{l(i-1) \times (N-2i+1)}$. The input block-Hankel matrices $U_{0|2i-1}$, U_p, U_f, U_{p^+}, and U_{f^-} are defined analogously with y replaced by u. The oblique projection matrix $\mathcal{O}_i \in \mathbb{R}^{li \times (N-2i+1)}$ is defined as

$$\mathcal{O}_i \triangleq Y_f /_{U_f} \begin{bmatrix} U_p \\ Y_p \end{bmatrix}, \tag{30}$$

and the extended observability matrix $\Gamma_i \in \mathbb{R}^{li \times n}$ is

$$\Gamma_i \triangleq \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{bmatrix}. \tag{31}$$

Finally, define the state sequences $X_p \in \mathbb{R}^{n \times (N-2i+1)}$, $X_f \in \mathbb{R}^{n \times (N-2i+1)}$ and $X_{f^-} \in \mathbb{R}^{n \times (N-2i+1)}$ as

$$X_p \triangleq [x_0 \quad x_1 \quad \cdots \quad x_{N-2i}], \tag{32}$$

$$X_f \triangleq [x_i \quad x_{i+1} \quad \cdots \quad x_{N-i}], \tag{33}$$

and

$$X_{f^-} \triangleq [x_{i+1} \quad x_{i+2} \quad \cdots \quad x_{N-i+1}]. \tag{34}$$

Definition 2
The input sequence $\{u_k\}_{k=1}^N$ is *persistently exciting* for (23), (24) if

$$\text{rank} \begin{bmatrix} X_p \\ U_{0|2i-1} \end{bmatrix} = 2mi + n. \tag{35}$$

Theorem 1 [32, Chap. 2, Theorem 2]
Assume that the input sequence $\{u_k\}_{k=1}^N$ is persistently exciting, let $w_k \equiv 0$ and $v_k \equiv 0$, and let the singular value decomposition of $\mathcal{O}_i$ be

$$\mathcal{O}_i = [U_1 \ U_2] \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{36}$$

$$= U_1 S_1 V_1^T, \tag{37}$$

where S_1 is nonsingular. Then, the following statements hold:

1) The matrix $\mathcal{O}_i$ can be factored as

$$\mathcal{O}_i = \Gamma_i X_f. \tag{38}$$

2) The order of the system (23), (24) is equal to the rank of S_1, that is, the number of nonzero singular values in (36).
3) There exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that

$$X_f = T S_1^{1/2} V_1^T. \tag{39}$$

Theorem 2 [32, Chap. 4, Theorem 12]

Assume that the input sequence $\{u_k\}_{k=1}^N$ is persistently exciting. Let the singular value decomposition of $\mathcal{O}_i$ be

$$\mathcal{O}_i = [U_1 \quad U_2] \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{40}$$

$$\approx U_1 S_1 V_1^T, \tag{41}$$

where $S_1 \in \mathbb{R}^{n \times n}$ is nonsingular. Then,

$$\lim_{N \to \infty} S_2 = 0. \tag{42}$$

Furthermore, define $\mathcal{O}_{i,\infty} \triangleq \lim_{N \to \infty} \mathcal{O}_i$ and $X_{f,\infty}^{KF} \triangleq \lim_{N \to \infty} X_f^{KF}$, where X_f^{KF} is the Kalman filter state sequence obtained when the initial conditions and initial covariance for the Kalman filter are chosen to be (4.27) and (4.28) of [32], respectively. Then the following statements hold:

1) The matrix $\mathcal{O}_{i,\infty}$ can be factored as

$$\mathcal{O}_{i,\infty} = \Gamma_i X_{f,\infty}^{KF}. \tag{43}$$

2) The order of (23), (24) is equal to the number of nonzero singular values in S_1 in (41).
3) There exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that

$$X_{f,\infty}^{KF} = T S_1^{1/2} V_1^T. \tag{44}$$

Using Theorem 2, define an estimate $\hat{X}_f$ of the state sequence X_f as $\hat{X}_f \triangleq S_1^{1/2} V_1^T$. An estimate $\hat{X}_{f-}$ of X_{f-} is obtained analogously. Next, the state space matrices are estimated by solving the least squares problem

$$\operatorname*{argmin}_{A,B,C,D} \left\| \begin{bmatrix} \hat{X}_{f-} \\ Y_{i|i} \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{X}_f \\ U_{i|i} \end{bmatrix} \right\|_F. \tag{45}$$

Missing Data

For the case of missing data, let q be a time step for which a measurement of u_q or y_q is unavailable. Then define the modified output block-Hankel matrix

$$Y_{0|2i-1} \triangleq \begin{bmatrix} y_0 & \cdots & y_{q-2i+1} & y_{q+1} & \cdots & y_{N-2i} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & \cdots & y_{q-i-1} & y_{q+i} & \cdots & y_{N-i-1} \\ \hline y_i & \cdots & y_{q-i} & y_{q+i+1} & \cdots & y_{N-i} \\ y_{i+1} & \cdots & y_{q-i+1} & y_{q+i+2} & \cdots & y_{N-i+1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{2i-1} & \cdots & y_{q-1} & y_{q+2i} & \cdots & y_{N-1} \end{bmatrix} \tag{46}$$

$$= \begin{bmatrix} \tilde{Y}_{0|i-1} \\ \tilde{Y}_{i|2i-1} \end{bmatrix} = \begin{bmatrix} \tilde{Y}_p \\ \tilde{Y}_f \end{bmatrix}. \tag{47}$$

Note that $\tilde{Y}_{0|2i-1}$ has the same form as the output block Hankel matrix (26), except that all of the columns of the $Y_{0|2i-1}$ containing y_q are omitted, that is $2i - 1$ columns are omitted. When several data points are missing, the block Hankel matrix is constructed similarly by omitting all columns that contain the time steps corresponding to the missing data. When N_q consecutive data points are missing, $\tilde{Y}_{0|2i-1}$ has $2i + N_q - 2$ fewer columns than $Y_{0|2i-1}$. However, when data points are missing at intervals of less than $2i$, $\tilde{Y}_{0|2i-1}$ is empty and no data are available for identification. $\tilde{Y}_{0|2i-1}$ can be partitioned alternatively as

$$\tilde{Y}_{0|2i-1} \triangleq \begin{bmatrix} y_0 & \cdots & y_{q-2i+1} & y_{q+1} & \cdots & y_{N-2i} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & \cdots & y_{q-i-1} & y_{q+i} & \cdots & y_{N-i-1} \\ \hline y_i & \cdots & y_{q-i} & y_{q+i+1} & \cdots & y_{N-i} \\ y_{i+1} & \cdots & y_{q-i+1} & y_{q+i+2} & \cdots & y_{N-i+1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{2i-1} & \cdots & y_{q-1} & y_{q+2i} & \cdots & y_{N-1} \end{bmatrix} \tag{48}$$

$$= \begin{bmatrix} \tilde{Y}_{0|i} \\ \tilde{Y}_{i+1|2i-1} \end{bmatrix} = \begin{bmatrix} \tilde{Y}_{p+} \\ \tilde{Y}_{f-} \end{bmatrix}. \tag{49}$$

The modified input block-Hankel matrices $\tilde{U}_{0|2i-1}$, $\tilde{U}_p$, $\tilde{U}_f$, $\tilde{U}_{p+}$, and $\tilde{U}_{f-}$ are defined analogously with y replaced by u. Finally, the modified state sequences $\tilde{X}_p$, $\tilde{X}_f$, and $\tilde{X}_{f-}$ are defined as

$$\tilde{X}_p \triangleq [x_0 \quad \cdots \quad x_{q-2i+1} \quad x_{q+1} \quad \cdots \quad x_{N-2i}], \tag{50}$$

$$\tilde{X}_f \triangleq [x_i \quad \cdots \quad x_{q-i} \quad x_{q+i+1} \quad \cdots \quad x_{N-i}], \tag{51}$$

and

$$\tilde{X}_{f-} \triangleq [x_{i+1} \quad \cdots \quad x_{q-i+1} \quad x_{q+i+2} \quad \cdots \quad x_{N-i+1}]. \tag{52}$$

These modified state sequences have a gap of $2i - 1$ time steps when one data point (u_q or y_q) is missing. $\tilde{X}_p$ is missing

states from time step $q - 2i + 2$ through time step q, while $\tilde{X}_f$ is missing states from time step $q - i + 1$ through time step $q + i$.

It is straightforward to check that theorems 1 and 2 hold when the modified block-Hankel matrices $\tilde{Y}_p$, $\tilde{U}_p$, $\tilde{Y}_f$, $\tilde{U}_f$, and the modified state sequence $\tilde{X}_f$ are used instead of Y_p, U_p, Y_f, U_f, and X_f, respectively. Thus by using these modified block Hankel matrices, subspace identification can be performed when data are missing.

The modified subspace algorithm for dealing with missing data points is used for linear subspace identification and the two Hammerstein-Wiener identification algorithms described above. However, if the missing data occur at intervals of less than $2i$, then identification is not possible since all columns of the Hankel matrices must be discarded.

IDENTIFICATION OF PERIODICALLY SWITCHING MODELS

To identify periodically switching models, consider the linear time-varying model

(a)

(b)
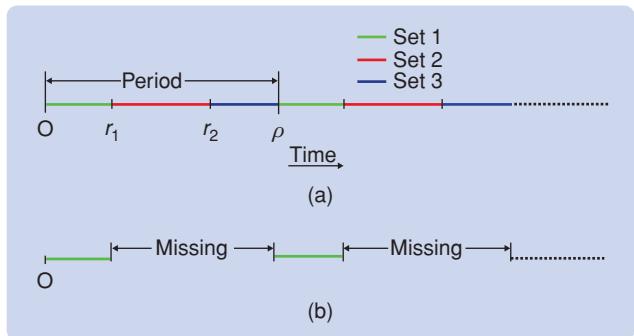
FIGURE 8 Identification of periodically switching models. The upper figure shows periodic switching among three sets of state-space matrices. The lower figure illustrates identification of the first set of state-space matrices, where data corresponding to the second and third sets of state-space matrices are viewed as missing.
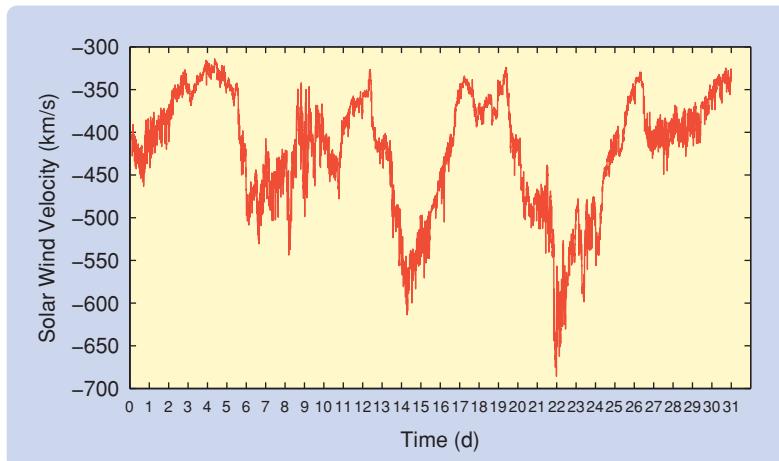
FIGURE 9 The velocity of the solar wind as a function of time for January 1999. This plot indicates that the velocity of the solar wind changes substantially over the course of a month. This change in the velocity implies a variable delay between the ACE measurements and inputs reaching the magnetosphere.

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \quad (53)$$
$$y_k = C_k x_k + D_k u_k + v_k, \quad (54)$$

where A_k, B_k, C_k, and D_k are assumed to be periodically switching with period ρ. The switching occurs at time steps $\nu\rho$, $\nu\rho + r_1$, $\nu\rho + r_2, \ldots, \nu\rho + r_{\tau-1}$, where ν is an integer, τ is the number of different sets of state-space matrices, and $r_0 = 0 < r_1 < r_2 < \cdots < r_{\tau-1} < \rho$ are the offset switching times. Since the state-space matrices remain constant between switching times, it follows that, for $\sigma = 1, \ldots, \tau - 1$,

$$A_{r_{\sigma-1}} = A_k, \quad r_{\sigma-1} \leq k \leq r_\sigma - 1 \quad (55)$$

and similarly for B_k, C_k, D_k.

To identify the σth set of state-space matrices, the data points corresponding to all of the other sets of matrices are viewed as missing data points as shown in Figure 8. Subspace identification with missing data discussed in the previous section is then used to estimate the state-space matrices. Since one missing datum causes a gap of $2i - 1$ state estimates in the estimated state sequence as seen in (51), the offset switching times r_σ must satisfy

$$r_\sigma - r_{\sigma-1} \geq 2i, \quad \sigma = 1, \ldots, \tau - 1. \quad (56)$$

Periodically switching Hammerstein-Wiener models are identified using the missing-data modification along with either the basis-function optimization algorithm or LS-SVMs. For basis-function optimization the gradient expressions (19)–(22) now involve products of the switching dynamics matrices $A_1, \ldots, A_\tau$. For example, consider three sets of switching matrices as shown in Figure 8 with $r_1 > 4$. Then, for $i = 4$ and $k = \rho - 2$ in (19)–(22), the term $\hat{C}\hat{A}^{k-i-1}\hat{B}$ becomes $\hat{C}_3\hat{A}_3^{\rho-2-r_2}\hat{A}_2^{r_2-r_1}\hat{A}_1^{r_1-4}\hat{B}_1$.

IDENTIFICATION USING ACE AND MAGNETOMETER DATA

Time Shifting of the Input Data

For system identification, the IMF and solar wind data measured by ACE are used as model inputs, while the ground-based magnetometer data are used as model outputs. In the region between ACE and the magnetopause, the solar wind is unobstructed and thus can be assumed to be constant. Hence, since the solar wind takes an additional 30 to 90 min to travel from ACE to the magnetopause, there is a time delay between the ACE measurements and the inputs entering the magnetosphere. This time delay, which depends on the solar wind velocity, is given by D_1/V_x, where V_x is the x component of the velocity of the solar wind, and D_1 is the distance between ACE and the magnetopause.

Figure 9 shows the velocity of the solar wind as measured by the ACE satellite for the month of January 1999. As seen from Figure 9 the solar wind velocity typically varies between 300 km/s and 900 km/s. To identify models that take into account the time delay, we advance the ACE data by time $t = D_1/V_x$. That is, each measurement is advanced by a different amount of time depending on the current solar wind velocity. Figure 10 shows a schematic of the ACE data shifting. Measurements made by the ACE satellite are thus inputs to the model at a time D_1/V_x into the future. The identified models thus have a real-time predictive capability of 30 to 90 min.

Input Selection and Model Switching

Physics suggests that combinations of components B_z and B_y of the IMF play a role in determining the size, shape, and dynamics of the magnetosphere. In particular, terms that are potentially useful include [7], [33] $B_s, V_x B_s, V_x^2 B_s, V_x B_z, V_x B_t,$ and $B_t V_x \sin^4(\theta/2)$, where $B_t \triangleq \sqrt{B_y^2 + B_z^2}, \theta \triangleq \cos^{-1}(B_z/B_t), V_x$ is the x-component of the velocity of the IMF, and

$$B_s \triangleq \begin{cases} B_z, & \text{if } B_z < 0, \\ 0, & \text{if } B_z \geq 0. \end{cases} \tag{57}$$

Testing all of these terms shows that $B_t V_x \sin^4(\theta/2)$ yields the least prediction error. In addition to the inputs through the Hammerstein nonlinearity, inputs that enter linearly into the dynamical system are also used. These linearly entering inputs are chosen from several candidate inputs using the error search algorithm described in the next section.

As seen from Figure 4, the size and shape of the magnetosphere on the daytime side and nighttime side of the Earth are significantly different. As a consequence of this asymmetry the flow patterns and the dynamics on the daytime side and nighttime side of the Earth are different. However, this asymmetry and the resulting differences in dynamics are fixed in a sun-fixed coordinate system, whereas the ground-based magnetometers rotate along with the Earth. The ground-based magnetometers thus measure magnetic perturbations in the daytime side and nighttime side alternately, with a 24-h periodicity. A periodically switching Hammerstein-Wiener model is used to capture this one-day periodicity and model the different dynamics governing the daytime side and nighttime side of the magnetospheric system.

Overfitting Determination

Candidate model inputs include B_z, B_y, B_x, plasma density, V_x, V_y, B_t, and $B_t V_x \sin^4(\theta/2)$. Since all of the above candidate inputs can be used for identification, there is a risk of overfitting [34], which can hamper the predictive capabilities of an empirical model. Overfitting is a result of either nonrepeatable effects such as noise or model/system mismatch. To

avoid overfitting, an error search algorithm is used to rank the inputs in the order of their importance. Specifically, the error search algorithm systematically adds one input at a time and evaluates the prediction error. The prediction error is evaluated as $\sum_{i=p}^{N} (y(i) - \hat{y}(i))^4$, where p is the first data point in the prediction region. A quartic error function is used instead of a quadratic function to emphasize the peak-finding ability of the model. If the prediction error degrades
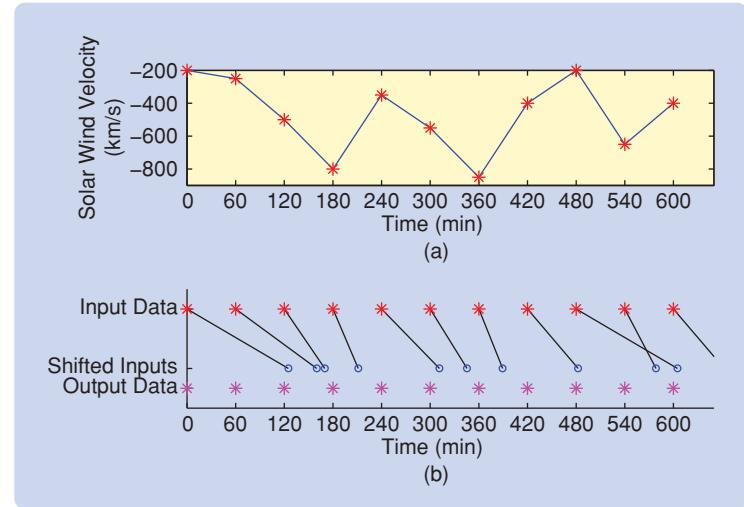
FIGURE 10 A schematic to demonstrate the procedure for time shifting ACE measurements. The upper plot (a) shows illustrative velocities, while the lower plot shows the corresponding time shifts. The x-axis in both plots denotes time in minutes, while in the lower plot (b) the red asterisks denote ACE measurements, the magenta asterisks denote magnetometer measurements, and the blue circles denote time-shifted inputs. For example, the measurement made at 0 min is shifted by about 120 min since the solar wind speed at 0 min is 200 km/s, while the measurement made at 180 min is shifted by about 30 min since the solar wind speed at that instant is 800 km/s. All data are fictitious for illustrative purposes.
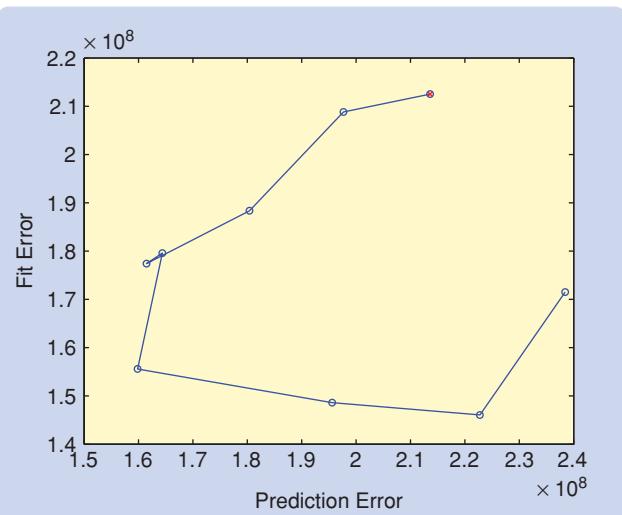
FIGURE 11 Illustration of the error search algorithm. Each branch in the plot represents an additional input in the model. Starting from the point marked with ×, in the first few branches the fit error and prediction error are both decreasing, but later the prediction error increases, indicating overfitting.
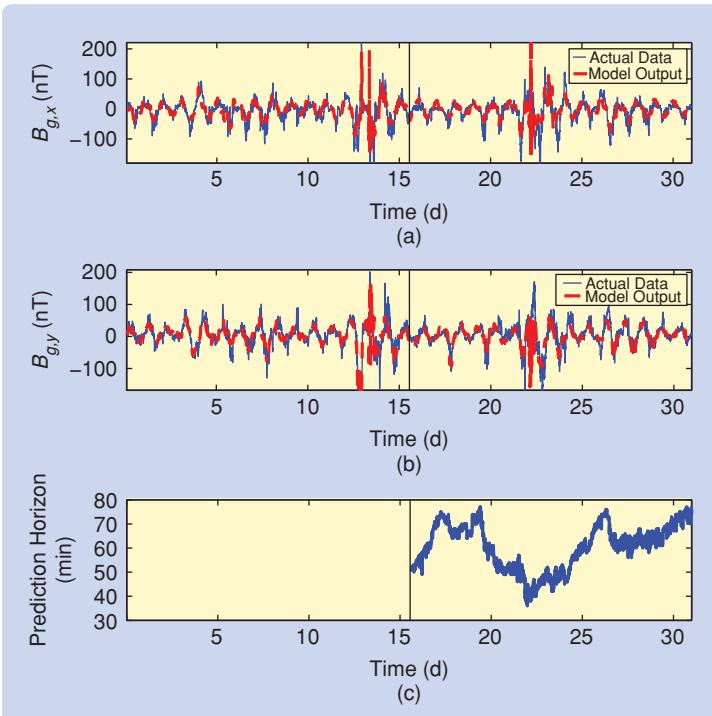
FIGURE 12 Measured and predicted data of the Thule magnetometer with the periodically switching Hammerstein-Wiener model identified using the basis-function optimization method. The x-component of the magnetometer reading $B_{g,x}$ and the y-component of the magnetometer reading $B_{g,y}$ are shown in (a) and (b), respectively. Data to the left of the vertical line are used for identification, while data to the right of the vertical line are used to assess the prediction ability of the identified model. The bottom plot (c) shows the prediction horizon in minutes for the corresponding model outputs, that is, the time by which the measurements made by the ACE satellite are shifted to obtain the predictions. The prediction efficiency for the x-component of the magnetometer data in the above plot is 0.9189.
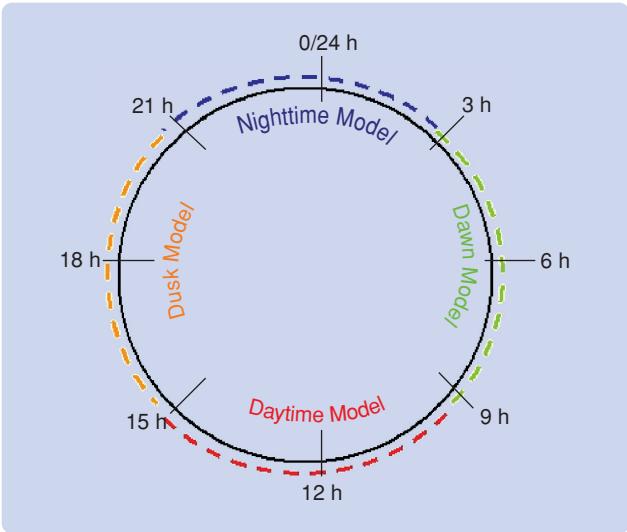
FIGURE 13 Switching times for the identified periodically switching Hammerstein-Wiener models. To capture the different dynamics during daytime and nighttime, the periodic model switches between four sets of state-space matrices. While the daytime and nighttime models represent the dynamics during periods of light and dark, respectively, the dawn and dusk models represent transition dynamics.

when an input is included, then that input is discarded, even if the fit error improves substantially. This test is performed for all of the candidate inputs. A typical outcome of the error search algorithm is shown in Figure 11, where each branch in the plot indicates an additional input to the model. Based on the error search algorithm the most effective linearly entering inputs are density, B_y, B_z, and B_t, whereas the input entering through the Hammerstein non-linearly is $B_t V_x \sin^4(\theta/2)$.

Regularization

We use regularized least-squares to obtain better estimates than the standard least-squares solution. In LS-SVM identification algorithm in [30], [35], Tikhonov regularization is incorporated into the subspace algorithm. Similarly, in the Hammerstein-Wiener identification method based on basis-function optimization, we use regularization in the least squares step to obtain improved estimates of the state-space matrices.

Identification Results

For illustration, data for the month of January 1999, with a sampling period of 15 min, are used to build periodically switching Hammerstein-Wiener models. Data from the first 15 days are used for identification, while data from the next 16 days are used for validation. The output of the periodically switching Hammerstein-Wiener model for THL, identified using the basis-function optimization method is shown in Figure 12. In this model, the dynamics order is chosen to be eight as determined by examining the singular values of $\mathcal{O}_i$ according to Theorem 2, and switching between four sets of state-space matrices occurs at regular time intervals of 6 h each as shown in Figure 13. As indicated by Figure 12, transients at the switching interface are typically not significant, and thus interpolation schemes are not employed. Furthermore, thin-plate spline radial basis functions are used to represent both the Hammerstein and Wiener nonlinearities. In the figure, data to the left of the black vertical line are used for identification, while the identified model is used to predict the data to the right of the vertical line. Identification is repeated with the data for the seventh day assumed to be missing. The model output is shown in Figure 14.

Model Validation

The prediction efficiencies E_p of the periodically switching Hammerstein models are calculated as [7]

$$E_\mathrm{p} \triangleq 1 - \sum_k (y_k - \hat{y}_k)^2 / \sigma_y^2, \qquad (58)$$

where σ_y^2 is the variance of the measured output y. From Table 4, the prediction efficiency for periodically switching

TABLE 4 Prediction efficiencies for the identified models of the Thule magnetometer, obtained from the periodically switching Hammerstein-Wiener methods. $E_{p,B}$ is the prediction efficiency in predicting the amplitude of the magnetic field, while $E_{p,\dot{B}}$ is the prediction efficiency in predicting the magnetic-field fluctuations. A prediction efficiency of $E_p = 1$ corresponds to perfect prediction.

Identification Method	dim f (r)	Order of $\mathcal{L}$ (n)	dim g (s)	$E_{p,B}$	$E_{p,\dot{B}}$
Basis-function optimization	49	8	31	0.9189	0.7389

Hammerstein-Wiener model identified using basis-function optimization is seen to be above 70% for both the amplitude B and the fluctuation dB/dt of the magnetic field. Next, periodically switching Hammerstein-Wiener models for the ground-based magnetometer stations KTN and STF are identified using the basis-function optimization method. The prediction efficiencies for the magnetometer stations listed in Table 1 as a function of the local time are shown in Figure 15.

Furthermore, we examine the frequency content of the residuals by computing the FFT of the prediction error. Figure 16 shows the frequency content of the residual between the measured data and the predictions of the Hammerstein-Wiener model for the Thule ground-based magnetometer identified using basis-function optimization. The spectrum of the residual does not indicate presence of prominent unmodeled linear effects. Additionally, the correlation coefficients of the residual with the inputs are all less than 0.15.
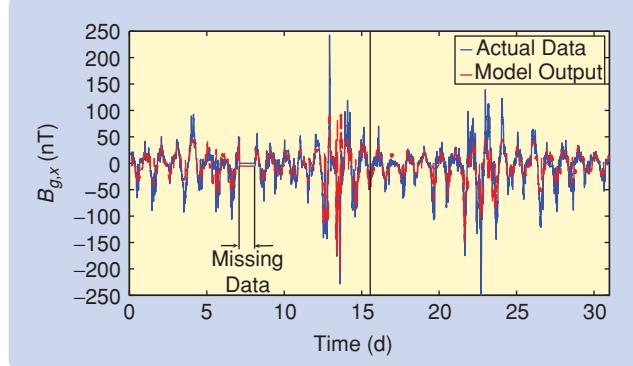
FIGURE 14 Measured and predicted data of the Thule magnetometer with the periodically switching Hammerstein-Wiener model identified using the basis-function optimization method. For this plot, the data for the seventh day are assumed to be missing. Data to the left of the vertical line are used for identification, while data to the right of the vertical line are used to assess the prediction ability of the identified model. The prediction efficiency in the above plot is 0.9142.

What Is Regularization?

To illustrate regularization, consider

$$z_k = \theta \phi_k + e_k, \tag{S1}$$

where $\phi_k \in \mathbb{R}^{n_\phi}$ and $z_k \in \mathbb{R}^{n_z}$ are known, $e_k \in \mathbb{R}^{n_z}$ is unknown measurement noise, and $\theta \in \mathbb{R}^{n_z \times n_\phi}$ is to be determined. We define the cost function $J_{ls}(\theta)$ to be the sum of squared errors

$$J_{ls}(\theta) \triangleq \sum_{k=1}^{N} \|\theta \phi_k - z_k\|_F^2, \tag{S2}$$

where N is the number of measurements. Defining $z \triangleq [z_1 \cdots z_N] \in \mathbb{R}^{n_z \times N}$ and $\phi \triangleq [\phi_1 \cdots \phi_N] \in \mathbb{R}^{n_\phi \times N}$, assuming that $\text{rank}(\phi) = n_\phi$, and minimizing (S2) yields the least-squares solution

$$\hat{\theta}_{ls} = z \phi^T (\phi \phi^T)^{-1}. \tag{S3}$$

When $\phi \phi^T$ is ill conditioned, some components of $\hat{\theta}_{ls}$ may be large. Ill conditioning of $\phi \phi^T$ may result from overparameterization, which occurs when θ has more components than needed.

Regularization [34], [35], [39], which is a modification of standard least squares, is useful when $\phi \phi^T$ is ill conditioned. Regularization involves augmenting the least-squares cost with a function of the unknown parameters. The most commonly used regularization method is Tikhonov regularization [39] (also called ridge regression) in which the regularized cost function is

$$J_{rls}(\theta) \triangleq \sum_{k=1}^{N} \|\theta \phi_k - z_k\|_2^2 + \gamma \|\theta\|_2^2, \tag{S4}$$

where $\gamma > 0$ is the regularization parameter. The regularized solution that minimizes (S4) is

$$\hat{\theta}_{rls} = z \phi^T (\phi \phi^T + \gamma I)^{-1}. \tag{S5}$$

By adding γI to $\phi \phi^T$ in (S5), γ has little effect on the significant eigenvalues of $\phi \phi^T$ (that is, eigenvalues λ such that $\lambda \gg \gamma$), while the small eigenvalues (satisfying $\lambda \ll \gamma$) are essentially set to γ. Since the condition number of $\phi \phi^T + \gamma I$ is approximately λ_{max}/γ, where λ_{max} is the largest eigenvalue of $\phi \phi^T$, γ is chosen to improve the condition number of $\phi \phi^T$ without affecting the large eigenvalues of $\phi \phi^T$.

In simplified terms, the penalty term $\gamma \|\theta\|_F^2$ in (S4), reduces the components of θ that are not useful in minimizing J_{ls}. However, regularization tends to increase bias in the estimates compared to standard least squares [34].

As noted in [30] subspace-based Hammerstein identification using basis functions is an overparameterized method. This overparameterization leads to ill conditioning of the matrices involved in the least-squares step for estimating the state space matrices A, B, and C.

CONCLUSIONS

A Hammerstein-Wiener identification algorithm is considered for application to the magnetospheric system. A modified subspace algorithm that allows missing data points is described and used to identify periodically switching models. To capture the periodically time-varying nature of the system, periodically switching Hammerstein-Wiener models are then identified using the two Hammerstein-Wiener identification methods. The inputs to the models are measurements from the ACE satellite, which is located at the first Lagrangian point between the Sun and the Earth, while the outputs of the model are ground-based magnetometer readings. The identified models are used to predict future magnetic fluctuations at 3 ground-based magnetometers and are validated using prediction efficiencies. Future work is expected to focus on additional lower latitude magnetometers and identification based on alternative model structures.

As discussed in "Real-time Web-based Magnetometer Data Prediction," identification using the basis-function optimization method is implemented to identify models and predict data for several ground-based magnetometer stations.
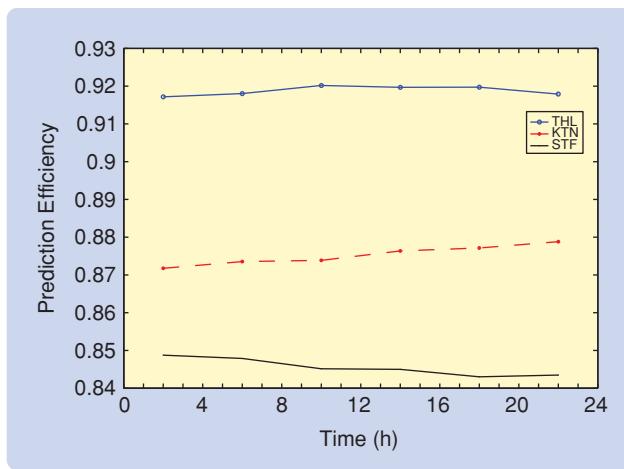
FIGURE 15 Prediction efficiencies as a function of local time for three ground-based magnetometer stations. For all three stations, Hammerstein-Wiener models are identified using the basis-function optimization method. The prediction efficiencies are best for the ground-based magnetometer at Thule, which is a high-latitude station.
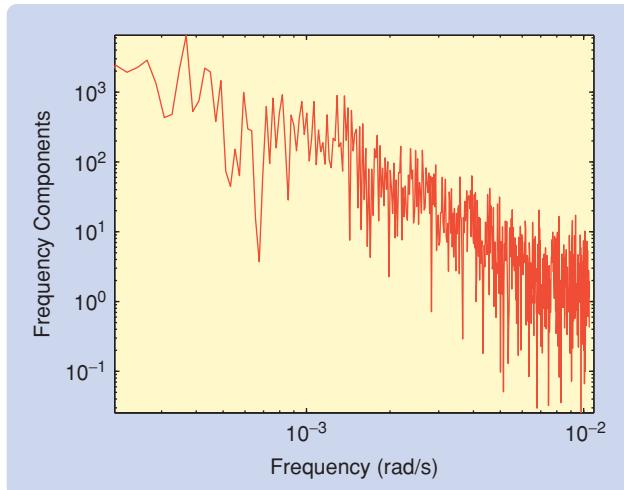
FIGURE 16 Frequency content of the residual between the actual data and the predictions made by the Hammerstein-Wiener model for the Thule ground-based magnetometer identified using the basis-function optimization method. The spectrum of the residual is not flat and shows significant content at low frequencies, which suggest that the identified models do not capture all the features of the actual system.

Real-Time Web-Based Magnetometer Data Prediction

To provide advance warning of future solar storms, a website for real-time magnetometer data prediction [40] is being developed. This website provides real-time prediction of ground-based magnetometer data for 30 to 90 min into the future depending on the currrent solar wind velocity.

The most recent ACE and ground-based magnetometer data are downloaded every minute from the Web site [41]. Using this data, periodically switching Hammerstein-Wiener models are identified once per day. A different periodically switching Hammerstein-Wiener model is identified for each magnetometer station. These models are then used with advanced ACE data to generate predictions of ground-based magnetometers, and the prediction plots are displayed on the website. These prediction plots are updated every 10 min. Thus, these predictions are generated continually, 24 hours a day, 365 days a year. All of the above steps are achieved through automated UNIX and Matlab scripts. For maintenance purposes extensive log files are also maintained.

Since the Sun rotates around its own axis with a periodicity of 27 days, the Earth is exposed to certain anomalous features of the solar surface every 27 days. Due to this periodicity some characteristics of the solar wind repeat every 27 days. To take advantage of this repeatability, the last 27 days of data are used for identifying the current model for web-based prediction.

Currently, the real-time prediction website provides predictions for the four ground-based magnetometers listed in Table S1. However, data for 32 magnetometers spread over the northern hemisphere are downloaded in real time. Predictions of magnetometer data for all of these magnetometer stations are expected to be added to the Web site.

Due to computational limitations, the optimization step of the basis-function optimization method is not used to construct the Web site models. Also, the error search algorithm that chooses model inputs as well as the tuning of the regularization parameter are performed offline.

REFERENCES

[1] D.H. Hathaway, *The Solar Wind, Solar Physics*. [Online]. Available: http://science.msfc.nasa.gov/ssl/pad/solar/sun_wind.htm

[2] S.F. Odenwald, *The 23rd Cycle*. New York: Columbia Univ. Press, 2001.

[3] I. Sandahl, "Does space weather really matter on the ground?" in *Proc. COSPAR Space Weather Study Using Multipoint Techniques*, vol. 12, Sept. 2000, pp. 349–357.

[4] S.-I. Akasofu, "The development of the auroral substorm," *Planet. Space Sci.*, vol. 101, pp. 273–282, 1964.

[5] K.G. Powell, P.L. Roe, T.J. Linde, T.I. Gombosi, and D.L. De Zeeuw, "A solution-adaptive upwind scheme for ideal magnetohydrodynamics," *J. Comput. Phys.*, vol. 154, pp. 284–309, 1999.

[6] J.G. Lyon, J.A. Fedder, and C.M. Mobarry, "The Lyon-Fedder-Mobarry (LFM) global MHD magnetospheric simulation code," *J. Atmos. Solar-Terr. Phys.*, vol. 66, pp. 1333–1350, 2004.

[7] R.S. Weigel, A.J. Klimas, and D. Vassiliadis, "Solar wind coupling to and predictability of ground magnetic fields and their time derivatives," *J. Geophys. Res.*, vol. 108, no. A7, pp. 16-1–16-14, 2003.

[8] D.N. Baker, R.S. Weigel, E.J. Rigler, R.L. McPherron, D. Vassiliadis, C.N. Arge, G.L. Siscoe, and H.E. Spence, "Sun-to-magnetosphere modeling: CISM forecast model development using linked empirical methods," *J. Atmos. Solar-Terr. Phys.*, vol. 66, pp. 1491–1497, 2004.

[9] D. Baker, "Statistical analyses in the study of solar wind-magnetosphere coupling," in *Solar Wind-Magnetosphere Coupling*, Y. Kamide and J. Slavin, Eds. Norwell, MA: Kluwer, 1986, pp. 17–38.

[10] D. Bauer and B. Ninness, "Asymptotic Properties of least-squares estimates of hammerstein-wiener models," *Int. J. Contr.*, vol. 75, no. 1, pp. 34–51, 2002.

[11] I.W. Hunter and M.J. Korenberg, "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models," *Biol. Cybern.*, vol. 55, pp. 135–144, 1986.

[12] Z. Hasiewicz, "Identification of a linear system observed through zero-memory non-linearity," *Int. J. Syst. Sci.*, vol. 18, no. 9, pp. 1595–1607, 1987.

[13] C.H. Chen and S.D. Fassois, "Maximum likelihood identification of stochastic wiener-hammerstein-type non-linear systems," *Mech. Syst. Signal Processing*, vol. 6, no. 2, pp. 135–153, 1992.

[14] E.-W. Bai, "An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems," *Automatica*, vol. 34, no. 3, pp. 333–338, 1998.

[15] C.-H. Chen and S.D. Fassois, "On the estimation of stochastic Wiener-Hammerstein-type systems with non-smooth non-linearity," in *Proc. American Contr. Conf.*, Albuquerque, NM, 1997, pp. 1494–1498.

[16] W. Greblicki, "Nonparametric approach to Wiener system identification," *IEEE Trans. Circuits Syst.*, vol. 44, no. 6, pp. 538–545, 1997.

[17] M. Lovera, T. Gustafsson, and M. Verhaegen, "Recursive subspace identification of linear and non-linear Wiener state-space models," *Automatica*, vol. 36, pp. 1639–1650, 2000.

[18] G.A. Pajunen, "Recursive identification of Wiener type nonlinear systems," in *Proc. American Contr. Conf.*, Boston, MA, June 1985, pp. 1365–1370.

[19] T. Wigren, "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model," *IEEE Trans. Automat. Contr.*, vol. 39, no. 11, pp. 2191–2206, 1994.

[20] D.H. Baldelli, R. Lind, and M. Brenner, "Nonlinear aeroelastic/aeroservoelastic modeling by block-oriented identification," *AIAA J. Guid. Contr. Dyn.*, vol. 28, pp. 1056–1064, 2005.

[21] V. Verdult. *Nonlinear System Identification: A State-Space Approach*. Enshede, NL: Twente Univ. Press, 2002. Available: http://doc.utwente.nl/38684/1/tooooo18.pdf/

[22] V. Verdult and M. Verhaegen, "Kernel Methods for subspace identification of multivariable LPV and bilinear systems," *Automatica*, vol. 41, no. 9, pp. 1557–1565, 2005.

[23] V. Verdult and M. Verhaegen, "Identification of multivariable linear parameter-varying systems based on subspace techniques," in *Proc. Conf. Dec. Contr.*, Sydney, Australia, Dec. 2000, pp. 1567–1572.

[24] V. Verdult, M. Verhaegen, C.T. Chou, and M. Lovera, "Efficient and systematic identification of MIMO bilinear state space models," in *Proc. Conf. Dec. Contr.*, Tampa, FL, Dec. 1998, pp. 1260–1265.

[25] J.J. Hench, "A technique for the identification of linear periodic state-space models," *Int. J. Contr.*, vol. 62, no. 2, pp. 289–301, 1995.

[26] M. Verhaegen and X. Yu, "A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems," *Automatica*, vol. 31, no. 2, pp. 201–216, 1995.

[27] S.L. Lacy and D.S. Bernstein, "Subspace identification of nonlinear systems with measured-input nonlinearities," *Int. J. Contr.*, vol. 78, no. 12, pp. 906–926, Aug. 2005.

[28] H. Palanthandalam-Madapusi, J.B. Hoagg, and D.S. Bernstein, "Basis-function optimization for subspace-based nonlinear identification of systems with measured-input nonlinearities," in *Proc. American Contr. Conf.*, Boston, MA, July 2004, pp. 4788–4793.

[29] H. Palanthandalam-Madapusi, A.J. Ridley, and D.S. Bernstein, "Identification and prediction of ionospheric dynamics using a Hammerstein-Wiener model," in *Proc. American Contr. Conf.*, Portland, OR, June 2005, pp. 5052–5057.

[30] I. Goethals, K. Pelckmans, J.A.K. Suykens, and B. De Moor, "Subspace identification of hammerstein systems using least squares support vector machines," *IEEE Trans. Automat. Contr.*, vol. 50, no. 10, pp. 1509–1519, 2005.

[31] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, Jan. 1994.

[32] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Norwell, MA: Kluwer, 1996.

[33] S.-I. Akasofu, "Predicting geomagnetic storms as a space weather project," in *Proc. COSPAR Space Weather Study Using Multipoint Techniques*, vol. 12, Sept. 2000, pp. 3–20.

[34] O. Nelles. *Nonlinear System Identification*. New York: Springer-Verlag, 2001.

[35] I. Goethals, K. Pelckmans, J.A.K. Suykens, and B. De Moor, "Identification of MIMO Hammerstein model using least squares support vector machines," *Automatica*, vol. 41, no. 7, pp. 1263–1272, July 2005.

[36] F. Six. *What is the magnetosphere?* MSFC Space Plasma Physics, NASA. [Online]. Available: http://science.nasa.gov/ssl/pad/sppb/edu/magnetosphere/bullets.html

[37] E.R. Christian and A.J. Davis. "*Advanced Composition Explorer.*" [Online]. Available: http://www.srl.caltech.edu/ACE/

[38] A.J. Ridley, "Estimation of the uncertainty in timing the relationship between magnetospheric and solar wind processes," *J. Atmos. Solar-Terr. Phys.*, vol. 62, pp. 757–771, 2000.

[39] A.N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Sov. Math. Doklady*, vol. 4, pp. 1035–1038, 1963.

[40] H.J. Palanthandalam-Madapusi, A.J. Ridley, and D.S. Bernstein. *Real-time Web-based Prediction of Magnetometer Data*. [Online]. Available: http://amie.engin.umich.edu/~hpalanth

[41] USDOC/NOAA/NESDIS/National Geophysical Data Center. Available: http://www.ngdc.noaa.gov/

AUTHOR INFORMATION

Harish J. Palanthandalam-Madapusi (hpalanth@ umich.edu) received the B.E. degree from the University of Mumbai in mechanical engineering in 2001. From October 2001 to July 2002, he was a research engineer at the Indian Institute of Technology, Bombay. He received the Ph.D. degree from the Aerospace Engineering Department at the University of Michigan in 2007. His interests are in the areas of system identification, data analysis, data assimilation, and estimation. He is currently an assistant professor in the Mechanical and Aerospace Engineering Department at Syracuse University.

Dennis S. Bernstein is a professor in the Aerospace Engineering Department at the University of Michigan. He is editor-in-chief of *IEEE Control Systems Magazine* and author of *Matrix Mathematics* (Princeton University Press, 2005). His interests are in system identification and adaptive control for aerospace applications.

Aaron J. Ridley received the B.S. degree from Eastern Michigan University in physics in 1992 and the Ph.D. degree from the University of Michigan in atmospheric and space sciences in 1997. From 1997 to 1999, he was a research scientist at the Southwest Research Institute, San Antonio, Texas. He is currently an associate professor in the Department of Atmospheric, Oceanic, and Space Sciences at the University of Michigan. His interests include magnetospheric physics, the upper atmosphere, and high-latitude electrodynamical data assimilation.