AGU100 ADVANCING EARTH AND SPACE SCIENCE

Space Weather

RESEARCH ARTICLE
10.1029/2018SW001953

Special Section:
Space Weather Capabilities Assessment

Key Points:
- The SWMF model has been running in experimental real-time mode at CCMC for several years, and all saved output is available
- The comparison against real-time Dst is quite good, especially when a few hours after cold restarts are removed from the comparison
- It is necessary to include an inner magnetospheric drift physics model to reproduce Dst; a real-time run without one does much worse

Correspondence to:
M. Liemohn,
liemohn@umich.edu

Citation:
Liemohn, M., Ganushkina, N. Y., De Zeeuw, D. L., Rastaetter, L., Kuznetsova, M., Welling, D. T., et al. (2018). Real-time SWMF at CCMC: Assessing the Dst output from continuous operational simulations. *Space Weather*, *16*, 1583–1603. https://doi.org/10.1029/2018SW001953

Received 9 MAY 2018
Accepted 22 SEP 2018
Accepted article online 29 SEP 2018
Published online 15 OCT 2018

©2018. American Geophysical Union. All Rights Reserved.

Real-Time SWMF at CCMC: Assessing the Dst Output From Continuous Operational Simulations

Mike Liemohn[1], Natalia Yu. Ganushkina[1,2], Darren L. De Zeeuw[1], Lutz Rastaetter[3], Maria Kuznetsova[3], Daniel T. Welling[1,4], Gabor Toth[1], Raluca Ilie[5], Tamas I. Gombosi[1], and Bart van der Holst[1]

[1]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA, [2]Finnish Meteorological Institute, Helsinki, Finland, [3]Community Coordinated Modeling Center, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [4]Now at Department of Physics, University of Texas at Arlington, Arlington, TX, USA, [5]Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, IL, USA

Abstract The ground-based magnetometer index of Dst is a commonly used measure of near-Earth current systems, in particular the storm time inner magnetospheric current systems. The ability of a large-scale, physics-based model to reproduce, or even predict, this index is therefore a tangible measure of the overall validity of the code for space weather research and space weather operational usage. Experimental real-time simulations of the Space Weather Modeling Framework (SWMF) are conducted at the Community Coordinated Modeling Center (CCMC). Presently, two configurations of the SWMF are running in real time at CCMC, both focusing on the geospace modules, using the Block Adaptive Tree Solar wind-type Roe Upwind Solver magnetohydrodynamic model, the Ridley Ionosphere Model, and with and without the Rice Convection Model. While both have been running for several years, nearly continuous results are available since April 2015. A 27-month interval through July 2017 is used for a quantitative assessment of Dst from the model output compared against the Kyoto real-time Dst. Quantitative measures are presented to assess the goodness of fit including contingency tables and a receiver operating characteristic curve. It is shown that the SWMF run with the inner magnetosphere model is much better at reproducing storm time values, with a correlation coefficient of 0.69, a prediction efficiency of 0.41, and Heidke skill score of 0.57 (for a −50-nT threshold). A comparison of real-time runs with and without the inner magnetospheric drift physics model reveals that nearly all of the storm time Dst signature is from current systems related to kinetic processes on closed magnetic field lines.

Plain Language Summary As society becomes more dependent on technologies susceptible to adverse space weather, it is becoming increasingly critical to have numerical models capable of running in real time to nowcast/forecast the conditions in the near-Earth space environment. One such model is available at the Community Coordinated Modeling Center and has been running for several years, allowing for an assessment of the quality of the result. Comparisons are made against globally compiled index of near-Earth space storm activity, including numerous statistical quantities and tests. The skill of the model is remarkable, especially when a few hours after each of the *cold restarts* of the model are removed from the comparison. It is also shown that a global model alone is not that good at reproducing this storm index; a regional model for the inner part of geospace is necessary for good data-model agreement.

1. Introduction

Predictive models of disturbances in the space environment that lead to hazardous space weather effects are becoming more mature, advancing beyond the early stages of development and allowing for sophisticated assessments of the accuracy and reliability of these tools. Numerous centers and other institutions have codes running online and in real time, making nowcasts and forecasts of the full range of space, from solar photons to thermospheric neutrals. Sometimes it is the case that a national agency will take up the mantle of space weather prediction in that country. Many such *space weather centers* exist, including ones in Belgium (e.g., Berghmans et al., 2005; Hochedez et al., 2005), Turkey (e.g., Tulunay et al., 2006), the United Kingdom (e.g., Murray et al., 2017), and the United States (e.g., Manoj & Maus, 2012; Simpson, 2003). Tsagouri et al. (2013) compiled a comprehensive list of such centers, the quantities predicted, and the lead times expected. Several of these models have been available for a decade or more, providing years of space weather prediction results.

A number of online nowcast and forecast codes have undergone systematic evaluations of their performance. One of the earliest such studies is that by Lundstedt et al. (2002), who developed a neural net predictor of Dst and made it available online. Costello (1997) introduced a real-time Kp prediction method, which was run online by the Space Weather Modeling Center for many years. Another early predictor model available online is that of Wing et al. (2005), who presented a real-time forecast model for predicting the Kp index. Several other studies then followed that assessed the accuracy of online, real-time geomagnetic index prediction schemes, including the Temerin and Li (2006) Dst prediction model; the Saiz et al. (2008) model, which predicts Dst using only interplanetary magnetic field (IMF) Bz as the sole input; NARMAX Dst model (Boynton et al., 2011) and AL model (Amariutei & Ganushkina, 2012); the Rice University study analyzing prediction schemes for Dst, AE, and Kp (Bala et al., 2009; Bala & Reiff, 2012, 2014); the Tobiska et al. (2013) Dst prediction model using the Anemomilos method; and the Revallo et al. (2014) model using a neural network approach. Other real-time, online models have been developed and assessed for other space weather quantities, for example, predicting solar flares and the subsequent activity (e.g., Devos et al., 2014; Murray et al., 2017), solar wind propagation and shock/interplanetary coronal mass ejection (ICME) arrival (e.g., Jian et al., 2016; Liu & Qin, 2015; Oler, 2004; Reiss et al., 2016), solar energetic particle events (e.g., Marsh et al., 2015), radiation belt electrons (e.g., Boynton et al., 2015; Wei et al., 2011), kiloelectron volt-energy electrons in the inner magnetosphere (e.g., Ganushkina, Amariutei, et al., 2015), and ionospheric densities (e.g., Codrescu et al., 2012; Schunk et al., 2012; Tsagouri & Belehaki, 2015; Zhu et al., 2012).

Other studies of forecasting accuracy assessment examined numerical experiments that replicated real-time usage of a model. Investigations have been conducted for solar extreme ultraviolet and X-ray irradiance, such as Henney et al. (2012), who showed good predictions of the daily $F_{10.7}$ value 7 days in advance. Many studies have considered solar wind propagation, in particular CME/ICME shock arrival times at Earth, including relatively early studies by Fry et al. (2003) and Mozer and Briggs (2003), as well as quite a few during the last solar maximum (e.g., Howard & Tappin, 2010; Norquist & Meeks, 2010; Macneice et al., 2011; Qin et al., 2009; Turner & Li, 2011), and some very recent analyses of shock propagation and arrival (e.g., Savani et al., 2017), using an archival time interval but running the codes as if they were conducting a real-time simulation. Some geospace models have also been used in this replicated real-time state, including Dst predictions (e.g., O'Brien & McPherron, 2000), ionospheric densities (e.g., Wang et al., 2001), convection patterns (e.g., Bekerat et al., 2003), and geosynchronous electron fluxes (e.g., Balikhin et al., 2016; Horne et al., 2013).

A final class of studies of relevance to the present work includes those that identify the predictive features of intense or severe space weather. This goes back to Gonzalez et al. (1994) classifying the typical driving conditions of different levels of magnetic storms, and Bell et al. (1997) characterizing the common elements of superstorms. More recently, Tsubouchi and Kubo (2010) used probability distributions to develop method of predicting intense space weather a month in advance, giving good baseline accuracy for future operational models. Riley (2012) and Riley and Love (2017) used power law and lognormal distribution analysis to quantify the severity level of the once-in-a-century storm. Others in this category include the probabilistic prediction of strong southward IMF events by Zhang and Moldwin (2015), the solar wind speed probabilistic predictions by Bussy-Virat and Ridley (2016), and the severe geomagnetic disturbance prediction schemes of Ayala Solares et al. (2016) and Balan et al. (2017).

One thing missing from all of the above studies is an examination of a coupled framework simulation of geospace, including not only a global magnetohydrodynamic (MHD) model but also the inclusion of an inner magnetospheric drift physics code. Such a code exists at the Community Coordinated Modeling Center (CCMC), where a set of geospace coupled models within the Space Weather Modeling Framework (SWMF) has been running as part of their experimental real-time runs. While this is not a robust operational model configuration with error-resolving input streams to produce an unbroken output of predictions, it runs nearly continuously with only rare stoppages. It began in 2007, with a real-time version of just an MHD model coupled to an ionospheric potential solver, but then in 2011 another version with a drift physics model came online. This new version was the geospace model selected as part of the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) geospace code challenge (Pulkkinen et al., 2013). This new version has been running nearly continuously since 2014. Also note that this code is not the same as the operational SWMF geospace model running at NOAA SWPC. The SWPC version has been improved relative to this one running at CCMC, and the cold restart issue is not part of the SWPC code

setup. The main reason to use the CCMC version is that it has been running for a longer time than the SWPC version, and there is a comparison available with a simpler version of the SWMF.

Now that SWMF has been running in an experimental real-time setting at CCMC for several years, it is useful to conduct a quantitative assessment of its performance and accuracy. In this study, we will focus on its ability to predict magnetic storm severity. The next section describes the numerical setup of the SWMF code suite running there. Section 3 then presents a couple of example storm intervals, while section 4 shows a statistical analysis of the full interval of results available from the code. Section 5 discusses the importance and caveats of these results and places them into context of similar studies.

2. Experimental Real-Time Runs of SWMF at CCMC

The SWMF is a collection of models brought together within a structure of couplers and common utilities to provide a unified approach to space environment modeling (Toth et al., 2005, 2012). Here we refer to the SWMF code suite as the SWMF with three physics domain modules included: a MHD model for the global magnetospheric solution, a kinetic drift physics model to better capture the inner magnetospheric energy-dependent processes, and an ionospheric electrodynamics code to solve the Poisson equation for the electrostatic potential distribution. This type of code configuration has been running online at the CCMC in an experimental real-time mode in nearly continuous operation since 2014. Another version, with a lower MHD grid resolution without the drift physics code included, has been running for a decade, but this other simulation will not be the focus of this study. Below, the models within SWMF are described, the CCMC run setup is given, and an example of code output is presented along with a statement on result availability.

2.1. SWMF Model Suite

The SWMF model suite has three physics models included: an MHD code, a drift physics code, and an electrodynamics solver. For the MHD code, these simulations use the Block Adaptive Tree Solar wind-type Roe Upwind Solver (BATS-R-US) model (Powell et al., 1999). For these simulations, this code solves the ideal MHD equations with the inclusion of semirelativistic effects, including the Boris correction to reduce the speed of light and other wave speeds (Gombosi et al., 2002), implicit time stepping (Toth et al., 2006), and passive outflow from the ionosphere (Welling & Liemohn, 2014). It includes robust methods for removing magnetic flux divergence (Toth & Roe, 2002) and reducing numerical diffusion (Ridley et al., 2010; Toth et al., 2011). While the code includes other features, such as multispecies and multifluid equation sets, anisotropic pressure calculations, Hall term inclusion, and anomalous resistivity, these augmentations to BATS-R-US are not included in the simulations presented here.

The drift physics code is the Rice Convection Model, or RCM (Harel et al., 1981; Toffoletto et al., 2003), which solves the motion of energetic charged particles through the inner magnetosphere. It solves a range of energies for electrons and two ion species, protons, and singly charged oxygen. The bounce- and pitch angle-averaged phase space density is solved in a latitude-longitude grid at the northern ionospheric footpoints of the magnetic field lines, allowing the flux tube volume to change as needed for magnetic reconfigurations, with an associated adiabatic change in energy for the particles. This code is coupled into the SWMF (De Zeeuw et al., 2004) by taking its outer boundary condition and the magnetic field configuration from the MHD model and then passes its pressure values within the RCM domain space back to BATS-R-US, which are used to nudge the MHD solution to these more correct values. This nudging is done with a relaxation time constant of 20 s, determined by testing to balance stability and accuracy. The coupling is done every 10 s throughout the simulation.

The ionospheric electrodynamics solver is the Ridley Ionosphere Model (RIM), which takes the field-aligned currents from BATS-R-US to get a pattern of the electric potential throughout the ionosphere (Ridley et al., 2004; Ridley & Liemohn, 2002). Conductances in RIM are a function of the incoming field-aligned current intensity and location, plus daily-varying values for the dayside and nightside conductances. The electric fields from RIM are applied at the inner boundary of the MHD code to yield tangential velocities there, plus they are passed to the RCM for ExB drift calculations in the inner magnetosphere. This coupling is conducted every time step.

2.2. Real-Time Run Setup

The run setup used for the real-time experimental SWMF simulations at CCMC are as follows. BATS-R-US is run with ~1 million grid cells, with a smallest grid size of 0.25 R_E and largest cells of 8 R_E. The outer boundaries are

set at 32 R_E in the $+X$ upstream direction, 224 R_E in the $-X$ downstream direction, and 128 R_E in the $\pm Y$ and Z directions. The inner boundary is set to a 2.5 R_E sphere with cut Cartesian grid cells near the boundary. The mass density at the inner boundary is set to 28 + 0.1 × CPCP amu/cm^3, where CPCP is the cross polar cap potential in kilovolts determined from the maximum and minimum values from RIM. The code is run with a Sokolov solver (Sokolov et al., 2001) and a Koren flux limiter (Koren, 1993) with $\beta = 1.2$ to ensure robustness and 0.01 Boris factor to limit the wave speeds, especially over the Earth's magnetic poles. The nominal time step is set to 5 s, but this is adaptable to maintain numerical stability; this is a maximum value, and the actual time step could be reduced when the density or pressure changes exceed some limits. This time step reduction happens, especially during active times.

The RCM is run with 25 energy channels for electrons, 75 channels for protons, and 75 channels for oxygen ions. These are isotropic *fluids* with a specified adiabatic invariant, ranging (for the ions) from below a kiloelectron volt to several tens of kiloelectron volts at the outer boundary (i.e., up to a few hundreds of kiloelectron volts near the peak of the ring current pressure maximum). An ion loss lifetime of 10-hr *e*-folding rate is uniformly applied through the simulation domain, which extends from 1.01 R_E to the invariant latitude mapping to 95% of the distance to the dayside subsolar magnetopause. The outer boundary is therefore dynamic and, on the nightside, can vary from near geosynchronous orbit (~7 R_E) to double this distance (~15 R_E). The ionospheric footpoint spatial grid is nonuniform, but step sizes are typically less than 2° in latitude and less than half an hour in local time. The RCM is run with a 10-s time step, taking its outer boundary from the single-fluid MHD results by assuming a number density split of 90% H^+ and 10% O^+. The BATS-R-US density and pressure values in the overlap region with RCM are nudged toward the drift physics model's values with a 20-s relaxation timescale.

RIM is a two-dimensional Poisson equation solver. The lower latitude boundary is set at 35° with a zero potential value imposed there. Above this, a preconditioned generalized minimum residual solver is used to get the potential everywhere on a 1° resolution spherical grid. There is no time stepping in RIM, but it is coupled to BATS-R-US every 10 s.

This setup configuration achieves real-time solutions on 64 processors. The time step is typically in the ~1-s range, but sometimes during active geomagnetic conditions it can shrink to 10 or 100 times smaller. We note that this problem has been since resolved with more optimized parallel processing. When the code lagged significantly behind real time, then it was occasionally reset, resulting in a few hours of lost simulation time and another few hours of *ramp-up time* to get the simulation back to good comparison with observations. The effects of these restarts will be explored in section 3.

This run configuration is essentially the same as that used in the recent SWPC model assessment challenge (Pulkkinen et al., 2013). It is not exactly the same, as there have been several small code modifications since then to improve the accuracy and performance of the simulation result. The SWMF input values are taken from the real-time solar wind data stream from the Advanced Composition Explorer, in particular the IMF vector and the solar wind density, velocity vector, and temperature.

2.3. Example Output and Plot Availability

The output from these runs is available in several places. First, they can be accessed at the CCMC through their experimental real-time simulation webpage. This is the most versatile of the options because it enables the user to use the standard CCMC plotting software to explore the full range of output parameters and display styles from each of the codes within the SWMF suite (i.e., BATS-R-US, RCM, and RIM). Output is also available through CCMC's Integrated Space Weather Analysis (iSWA) website, for which many cygnets (an iSWA web application object, i.e., a clickable icon, that produces a predefined figure of user-defined time interval and size) have been created to display the output from this and other experimental real-time simulations. Two examples of the output available at iSWA are shown in Figure 1, displaying 2-D output slices from BATS-R-US during a relatively quiet time with nominal solar wind input. The features seen in the panels of Figure 1 are what is expected for a quiet time magnetosphere, that is, with the subsolar magnetopause near 10 R_E and the inner edge of the nightside plasma sheet near geosynchronous altitude. The straightforward interface of iSWA allows users to browse a set of predetermined yet standard plot options, assembling an analysis page of a user-defined set of cygnets, not only from SWMF but also from any other model or observational data set in the 300+ cygnet collection, for detailed examination of a particular time interval. Through

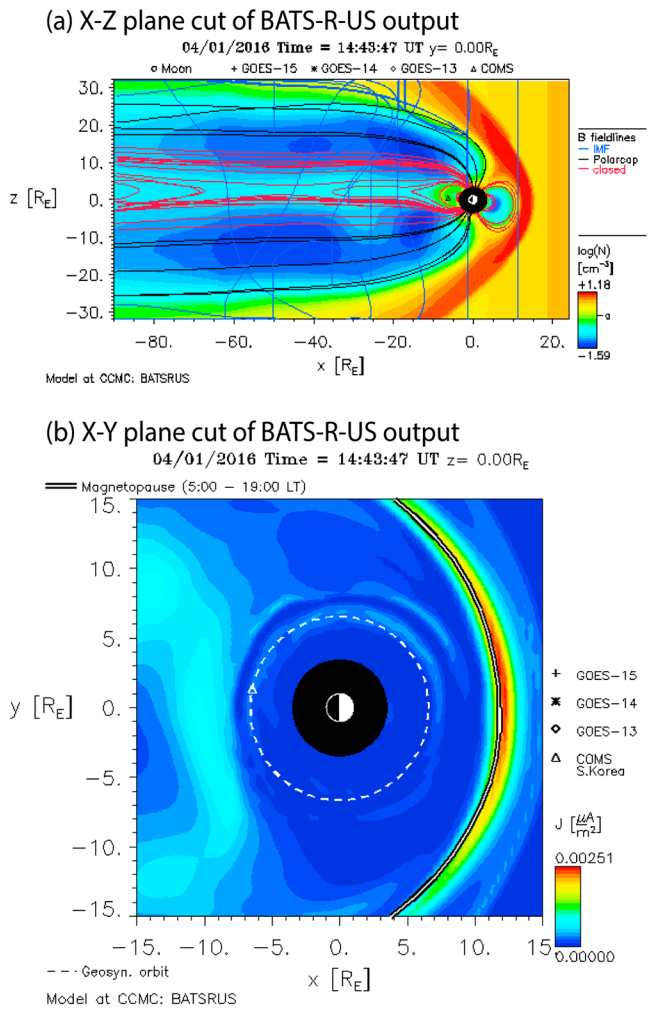

Figure 1. Example output from the experimental real-time SWMF simulations at CCMC, showing two plots from the BATS-R-US code during a relatively quiet interval with nominal solar wind values: (a) number density in the *X-Z* plane and (b) magnitude of current density in the *X-Y* plane. The overlaid lines in Figure 1a show the magnetic field, color coded by their connectivity topology (blue, IMF; black, open; red: closed). The white dashed circle in Figure 1b shows the location of geosynchronous orbit. SWMF = Space Weather Modeling Framework; CCMC = Community Coordinated Modeling Center; BATS-R-US = Block Adaptive Tree Solar wind-type Roe Upwind Solver; IMF = interplanetary magnetic field; GOES = Geostationary Operational Environmental Satellite.

iSWA, users have full access to the ever-growing history of saved images for each of these plot types, including the ability to create small movies.

Yet another place where output is available is a dedicated page at the University of Michigan (U-M) Center for Space Environment Modeling (CSEM). The CSEM site allows users to see the latest images of the full set of iSWA plots of the SWMF output, from both experimental real-time versions (i.e., with and without RCM). This site also shows a comparison of the model values for Dst displayed against the observed real-time Dst values from the Kyoto World Data Center (WDC) for geomagnetism. Figure 2 shows an example set of these Dst comparisons for 3 months in 2015. This interval was chosen because it contains a relatively large number of moderate storms, showcasing the ability of SWMF to capture the time series of measured geomagnetic activity.

3. Assessment of Output Quality

The Dst comparisons in Figure 2 are merely qualitative. To quantitatively assess the ability of the SWMF experimental real-time simulations to reproduce magnetic storms, a detailed analysis of this Dst comparison was conducted.

3.1. Analysis Methodology

The SWMF experimental real-time simulations have been reliably running online at CCMC since 19 April 2015. The values for this study extend from that day until 17 July 2017, providing 27 months of data for a statistical comparison against the real-time observations of geomagnetic activity over the same interval. In particular, the code produces a model-derived Dst index equivalent, which is compared here against the real-time Dst values provided by the Kyoto WDC.

The calculation of Dst from the SWMF follows the methodology of Yu et al. (2010). Two Biot-Savart law integrals are computed to obtain the model-derived value, one over the full domain space of the BATS-R-US MHD model and another over the 2-D domain space of the RIM ionospheric current distribution. The perturbation is calculated at the center of the Earth, as this has been shown to be equivalent to a summation around a low-latitude ring on the Earth's surface (Carovillano & Siscoe, 1973). The field-aligned currents in the gap region are omitted, which were found to be only a very weak contributor to a Dst calculation (e.g., Dubyagin et al., 2014; Yu et al., 2010).

While there are other ground-based magnetometer-derived globally compiled indices of geomagnetic activity available, Dst will be the only one considered here (see Mayaud 1980 for a full explanation of this and other common indices). Several studies have shown that the various Dst equivalent indices are, essentially equivalent, with errors between them in the range of 10 nT during quiet times and ~20% during active times (see, e.g., Katus & Liemohn, 2013; Love & Gannon, 2009; Wanliss & Showalter, 2006). The 1-min indices are not used here because the timing of the upstream solar wind conditions is not known to that accuracy; therefore, a comparison on that timescale is inappropriate. While the SWMF produces Dst output on a 1-min cadence, these input timing uncertainties would make the comparison metrics worse. The 1-hr resolution of real-time Dst is sufficient for the purpose of this comparison, especially when considering the validity of the model output over a >2-year interval.

Two approaches to the quantitative analysis will be conducted. The first set is a collection of statistical metrics that measure the quality of the model output to match the observed numbers at each time. The first in this set is a comparison of the full range of the calculated and measured index, which provides a comparison of

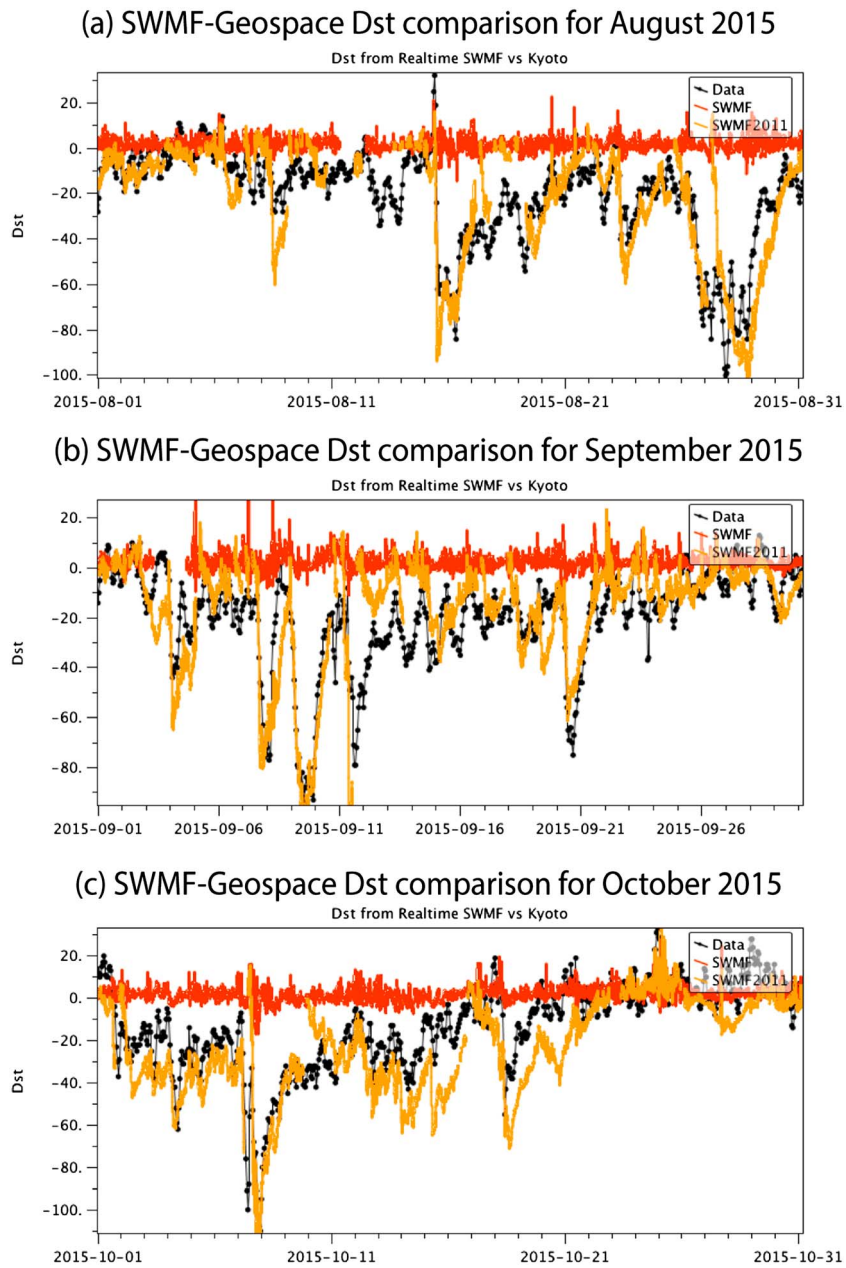

Figure 2. Example output from the CCMC experimental real-time runs of SWMF, showing Dst comparisons available at the CSEM website, for 3 months: (a) August 2015, (b) September 2015, and (c) October 2015. The black line with dots is the real-time Dst time series from the Kyoto WDC, the yellow line is the SWMF result with the RCM inner magnetosphere model included, and the red line is the earlier version of the experimental real-time code without the RCM. SWMF = Space Weather Modeling Framework; CCMC = Community Coordinated Modeling Center; CSEM = Center for Space Environment Modeling; WDC = World Data Center; RCM = Rice Convection Model.

the extremes of the modeled and observed data sets. The second is the average of all values in each set, which provides a comparison of the typical values in each set (Taylor, 1997, p. 93). The subtraction of these two means is called the mean error, a value that shows the bias in the model results. The third is standard deviation σ, a measure of the variability of any single value within each set from the mean. These first three statistics should be similar between the modeled and observed data to declare the comparison *good*. The distribution of Dst is not Gaussian, however, so the standard deviation does not have its common meaning of defining percentiles around the mean. To address this, the fourth metric to consider are specific percentiles of each data set. The chosen ones for this assessment are the 5%, 10%, and 50%

(median) values. Because the Dst index is so skewed with a long tail at very negative values, this choice for the percentiles highlights the active times, which are presumably of more interest to users of these model results. The fifth metric in the first set is the Pearson correlation coefficient (R), which determines the linearity of the fit between the two data sets. Because one is trying to emulate the other, it is expected that the two data sets should be linearly correlated. A good value of R varies with the data set and the interpretation, but it is generally accepted that $R > 0.5$ is considered good and $R > 0.7$ would be exceptionally noteworthy. Another way of interpreting R is to consider the coefficient of determination (R^2), which quantifies the amount of variation in one set captured by the variation in the other. Following the Pearson good values, a coefficient of determination above 0.25 is considered good while a value above 0.5 is exceptional. The sixth metric is the root-mean-square error (RMSE) between modeled and observed data, calculated with the same formula as model-output standard deviation but with the mean of the modeled Dst replaced by the observed Dst value for each time in the series:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(M_i - A_i)^2} \tag{1}$$

where M is the model value and A is the observed value at each time i in the series. This metric represents the variation of any single modeled Dst value away from the observed Dst value at any chosen time. A good value would be below the standard deviations of both the modeled Dst and observed Dst. The seventh and final metric is the prediction efficiency (PE), which is related to RMSE and the observed data's σ, like this (Wilks, 2011, p. 305):

$$\mathrm{PE} = 1 - \frac{\Sigma(M_i - A_i)^2}{\Sigma\left(A_i - \overline{A}\right)^2} \tag{2}$$

where the $\overline{A}$ is the mean of the observed data set. A PE of 1 means the model perfectly matches the data, and a value above zero indicates that the model is better at reproducing the observations than a random sampling from the observational data set (i.e., PE > 0 is a good value). PE is essentially a comparison against a null model, with the null model being a multiyear average of the observational data set.

The second set of metrics are from event-based analyses, quantifying the ability of the model to predict magnetic storms, defined here as an observed Dst value below some threshold, x_{thres}. For any hourly time in the data set, the modeled and observed Dst will either be above or below x_{thres}, creating a binary contingency table (see Murphy, 1996, for a comprehensive review of the development of the 2×2 contingency table). If both data sets are below x_{thres}, then the time is counted as a correct prediction, that is, a hit (H). Similarly, if both are above x_{thres}, then the time is counted as a correct negative (N). The other two categories represent incorrect predictions, either being misses (M) when the observed value crossed below x_{thres} but not the model value or a false alarm (F), when the model value crossed below x_{thres} but not the observed one.

The H, M, F, and N values can be combined to create metrics of the goodness of the model at predicting events in the observations. The first is the probability of detection (POD), defined as

$$\mathrm{POD} = \frac{H}{H+M} \tag{3}$$

The POD ranges from 0 to 1, depending on the relative amount of hits and misses, giving the proportion of time when the observations were in *event state* that the model was also in event state. A high value is desired. The second is the probability of false detection (POFD), given by the equation

$$\mathrm{POFD} = \frac{F}{F+N} \tag{4}$$

This quantity also ranges from 0 to 1, giving the proportion of time when the observations were in *nonevent state* that the model was also in nonevent state. A good model will yield a low POFD. The third quantity from the contingency table is the threat score (TS), which combines all of the values except the correct negatives, like this

$$\mathrm{TS}\,\frac{H}{H+M+F} \tag{5}$$

Also ranging from 0 to 1, TS shows the proportion of times when both modeled and observed data were in event state as a fraction of all times when either model or observation was in that state. TS is a useful quantity when N is much larger than the other table entries. High values of TS are good, even though it will always be less than POD. The fourth value is the false alarm ratio (FAR), exchanging N in the denominator of POFD for H

$$\mathrm{FAR} + \frac{F}{F+H} \tag{6}$$

Like the others, FAR also ranges from 0 to 1 and gives the fraction of times when the model was in the event state but the observations were not relative to the total number of times the model was in the event state. A low FAR is desirable. The fifth contingency table value to be considered here is the Heidke skill score (HSS), defined as (Heidke, 1926)

$$\mathrm{HSS} = \frac{2[(H{\cdot}N) - (M{\cdot}F)]}{(H+M)(M+N) + (H+F)(F+N)} \tag{7}$$

This is a measure of the fraction of correctly predicted times beyond x_{thres} after eliminating those correct predictions that could be from random chance. Of the various skill scores, it is a reasonable choice because it is truly equitable (see, e.g., Hogan & Mason, 2012) and widely used. The HSS range is from a best value of 1, in which all of the values are in the H and N categories, down to a worst value of -1, when all of the values are evenly split between F and M. A value zero corresponds to a random prediction with the same frequency of events as observed, while a negative value is worse than random forecast. Note that HSS can be undefined when the observations have no times in the event state and the model correctly predicts this outcome. Therefore, care must be taken in choosing an appropriate x_{thres} within the expected range of the two data sets that contains observed events and therefore tests the skill of the model. HSS and other skill score metrics have been used in short-range weather forecasting for nearly a century (e.g., Muller, 1944) and are still commonly used today (e.g., Wilks, 2011, chapter 8). The fifth quantity to be considered from the contingency table is the frequency bias or simply bias

$$\mathrm{Bias} = \frac{H+F}{H+M} \tag{8}$$

The bias does not report on the skill of the model to predict the observations, but rather on the diagonal asymmetry of the matrix. A value of one is symmetric, while values well below this indicate the model systematically underestimates the data and values well above this show that the model systematically overestimates the data. The sixth and final quantity from the contingency table is the measure of significance, χ^2,

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\left(E_{ij} - O_{ij}\right)^2}{E_{ij}} \tag{9}$$

where i and j cycle through the rows and columns of the contingency table, respectively, O_{ij} is the value in that cell of the table (i.e., H, M, F, or N), and E_{ij} is defined as

$$E_{ij} = \frac{T_i{\cdot}T_j}{T} \tag{10}$$

where T_i is the total in row i (i.e., either $H + M$ or $F + N$), T_j is the total in column j (so, either $H + F$ or $M + N$), and T is the sum of all times in the data set ($T = H + M + F + N$). The significance of this χ^2 value depends on the degrees of freedom of the contingency table, which is defined as (# of rows $-$ 1) $\times$ (# of columns $-$ 1). For our 2×2 binary table, this is 1. The value of χ^2 can be converted to a probability P that the table could happen purely by chance. To reach the 95% significance level, χ^2 should be greater than 3.84, and to reach the 99% highly significant classification, χ^2 should be greater than 6.63.

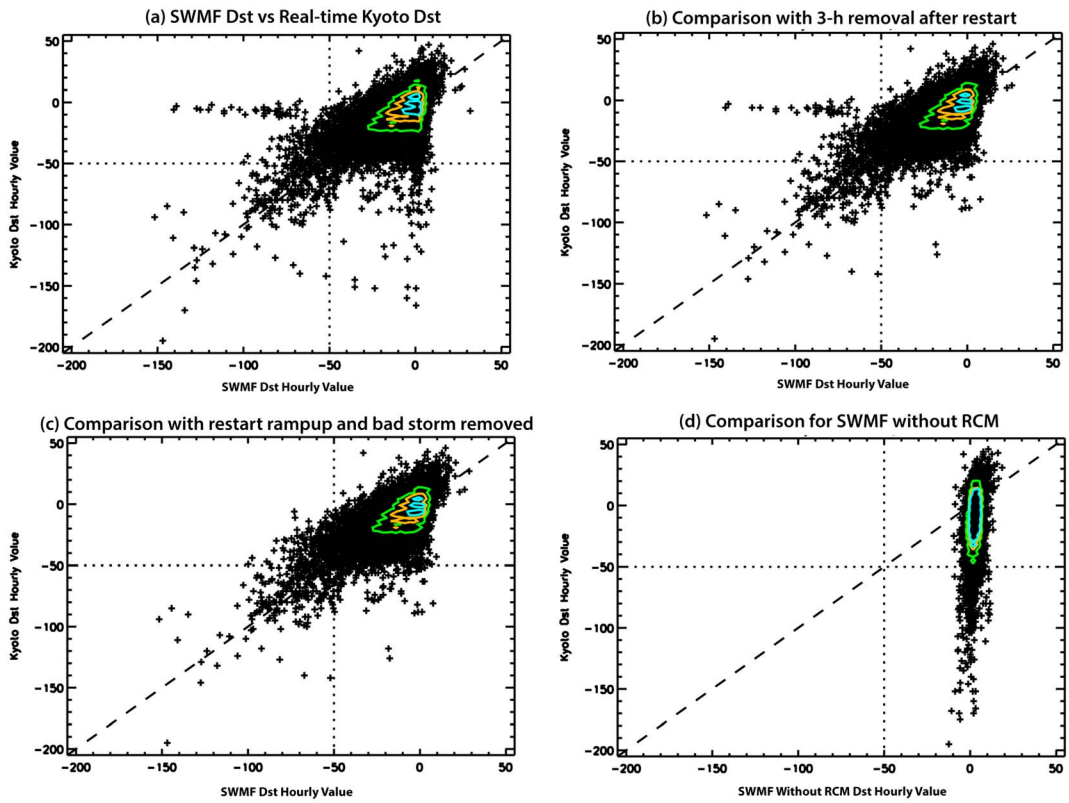

Figure 3. Scatterplots of the observed real-time Dst time series (*y* axis values) against a prediction Dst time series (*x* axis values). (a) The prediction is the SWMF model output Dst values. (b) The same values but with the removal of a 3-hr window after each restart of the code. (c) The same values as in Figure 3b but now also removing 2 days in March 2017. (d) The model is the SWMF code without RCM included in the simulation setup. The two dashed lines show the −50-nT threshold values used for the first set of contingency table analysis (Table 2). The green, orange, and cyan contour lines show the regions of 40, 80, and 120 values within a 2.6 × 2.6-nT bin. SWMF = Space Weather Modeling Framework; RCM = Rice Convection Model.

In the analysis below, two thresholds will be used, −50 and −30 nT. Defined by Gonzalez et al. (1994), the first is the cutoff for a moderate magnetic storm, while the second is the cutoff for a weak magnetic storm. While the 2+ years of the study interval are during solar maximum, solar cycle 24 is not a particularly active maximum and so the number of correct negatives, *N*, is expected to be large.

In addition, a receiver operating characteristic (ROC) curve (also called a relative operating characteristic) is generated to judge the quality of the model output at predicting any arbitrarily set activity level (Mason, 1982; Swets, 1973). By plotting the POD against the probability of false detection for a large set of threshold settings (varied from +10 to −80 nT, in this case), the resulting curve can be compared against a unity slope line to assess the model's goodness versus a randomly generated prediction. The closer the ROC curve shifts toward the upper left quadrant (high POD and low POFD), the better the model is at predicting any level of expected activity.

3.2. Statistical Analysis Results

Figure 3a shows a scatterplot of the observed real-time Kyoto Dst index against the modeled Dst from the experimental real-time SWMF simulation. A unity-slope dashed line is drawn for reference, along with two dotted lines at the −50-nT nominal storm thresholds. Values in the lower left quadrant are hits, those in the lower right quadrant are misses, the ones in the upper left quadrant are the false alarms, and the points in the upper right region are correctly identified negative values. To better see the distribution within the all-black region of overlapping points, the plot region was divided into 100 bins along each axis (i.e., 2.6-nT width) and the green, orange, and cyan contours demark the bins with more than 40, 80, and 120 values in them.

Table 1
Statistics of the Dst Comparison

	All available times	With a 3-hr restart gap	Removing bad storm
Usable number of values	15125	13650	13590
Range of Kyoto Dst values in set (nT)	[−195, 47]	[−195, 46]	[−195, 46]
Range of SWMF Dst values in set (nT)	[−152, 96]	[−152, 29]	[−152, 29]
Average and σ of Kyoto Dst in set (nT)	−11.7, 18.1	−11.3, 17.4	−11.3, 17.5
Average and σ of SWMF Dst in set (nT)	−11.7, 16.7	−12.9, 16.7	−12.8, 16.2
Percentile (5, 10, 50) of Kyoto Dst in set, nT	−42, −32, −10	−41, −31, −10	−41, −31, −10
Percentile (5, 10, 50) of SWMF Dst in set (nT)	−43.5, −31.9, −7.1	−45.2, −33.2, −8.6	−44.2, −32.9, −8.5
Pearson correlation coefficient (R)	0.59	0.67	0.69
Coefficient of determination (R^2)	0.35	0.44	0.47
Root-mean-square error (nT)	15.8	14.0	13.4
Prediction efficiency	0.24	0.35	0.41

Note. SWMF = Space Weather Modeling Framework.

Several features should be pointed out in the scatterplot in Figure 3a. First, most of the values are relatively close to the unity-slope line, indicating that, at least qualitatively, the model is reasonably reproducing the observed Dst time series.

A second feature to notice is the cluster of points at low values for the modeled Dst but relatively large observed Dst (in the lower right quadrant). These are from the occasional code restarts, which are initialized with a nominal input setting, resulting in a fairly low pressure inner magnetosphere. It takes a few hours of simulation to build up the pressure and the resulting Dst to realistic values. These data points can be eliminated by removing from the comparison a few hours after each restart from the comparison, essentially extending the missing-output gap a bit longer. Note that this also removes a cluster of correct negatives from the upper right quadrant of the figure.

Figure 3b shows the scatterplot again, this time with a 3-hr window removed from the comparison after each restart of the SWMF code. Most of the plot is left unaffected, with the main difference between Figures 3a and 3b being the omission of many of the more egregious misses, as well as some correct negatives. Nearly all of the modeled values that were significantly below its corresponding observed value were indeed the result of a fresh restart of the code, catching it during the ramp-up period. Several settings for the length of this window were examined, and 3 hr seems to be a good trade-off that removes nearly all of the spuriously low modeled Dst values while retaining as many of the data points in the comparison set as possible.

A third feature of note in Figure 3a is the cluster of points at low values for the observed Dst and relatively high values for the modeled Dst (in the upper left quadrant). These come from a single ~50-hr interval on 13–15 March 2017 when the model predicted a magnetic storm that was not recorded in the observed Dst time series. The real-time solar wind conditions appear to be nominal, quiet time values during this false-alarm event, yet the code output includes elevated solar wind driving, creating a storm. The diagnosis is that an incorrect solar wind file was used for these few days, although there is no clear indication of how such an error occurred. That is, the code responded to the solar wind input that it was given, even though this input was not the true real-time solar wind stream for this interval.

Because the March 2017 *bad storm* event is such an unusual anomaly in the data set, we are presenting results with this interval removed. Figure 3c shows the values from Figure 3b (with the 3-hr restart ramp-up times removed) but now also has the values from 13 to 15 March 2017 removed as well. The *false alarm* cluster is now gone. The final panel in Figure 3 shows results for a different SWMF code configuration and will be discussed in section 3.5.

Table 1 lists the statistics of the data-model comparison of the Dst time series. Table 1 shows that the useable number of hourly values is well above 10,000, even with the exclusion of a 3-hr window after restarts and the bad storm in March 2017. It shows that the full range of the SWMF Dst values is close to that of the observed real-time Dst values, with the extreme minimum being less severe in the model output than in the measurements. The mean and standard deviation of the observed and modeled Dst number sets are very close. The 5%, 10%, and 50% values of the observed and numerical data sets are listed in Table 1. It is seen that the 5% values are more negative for the model results than for the real-time Kyoto Dst sets, but that the 50% value

Table 2
Dst Comparison Contingency Table Values With an Event Threshold of −50 nT or Below (Moderate and Intense Storm Detection)

	All available times	With a 3-hr restart gap	Removing bad storm
Usable number of values	15125	13650	13590
Number of hits	273	266	266
Number of misses	207	131	131
Number of false alarms	282	279	245
Number of correct negatives	14363	12974	12948
Probability of detection	0.57	0.67	0.67
Probability false detection	0.019	0.021	0.019
Threat score	0.36	0.39	0.41
False alarm ratio	0.51	0.51	0.48
Heidke skill score	0.51	0.55	0.57
Frequency bias	1.16	1.37	1.29
Measure of significance, χ^2	3970.	4235.	4520.

shows the reverse of this, with the Kyoto medians being more negative than the model result medians. This indicates that the model results are more dispersed than the observed values, despite the smaller full range of the modeled values. The Pearson linear correlation coefficient is found to be 0.59 for the scatterplot shown in Figure 3a (all times included), 0.67 for the scatterplot in Figure 3b (excluding 3 hr after each restart), and 0.69 for the scatterplot in Figure 3c (excluding the restart ramp-up times and the bad storm interval). The last number is particularly good, resulting in a coefficient of determination of 0.47, which means that the model-output Dst can explain nearly half of the variation in the observed real-time Dst time series. The RMSE between the data and model Dst values are 15.8, 14.0, 13.4 nT for the three SWMF output sets, which are comparable to the uncertainty in the Dst index itself of ~12 nT as found by Katus and Liemohn (2013). Finally, the PE is well above zero for all three versions of the model output, reaching 0.41 for the set with both the restart ramp-up time and the bad storm removed.

Contingency table entries are made by counting the values in each quadrant of Figures 3a–3c. These values, for a Dst storm threshold of −50 nT, are given in Table 2. In addition to the count values for *H*, *M*, *F*, and *N* in the contingency table, the calculated values of POD, POFD, FAR, TS, HSS, FB, and χ^2 are listed as well. All of these numbers in Table 2 indicate that the model is performing well, with probabilities of detection and HSSs above 0.5, hit rates above 0.3, and probability of false detection values well below 0.1. The measures of significance are in the thousands, indicating that these values are well beyond what could reasonably be attained by chance.

Comparing the first two data columns of Table 2, it is seen that five of the seven calculated statistics (POD, FAR, TS, HSS, and χ^2) are equal or better for the modeled data with the 3-hr window after restarts removed. The only two that are worse are the POFD and the bias, which are slightly higher. Because the POFD value includes correct negatives in the denominator (see equation (4)), the removal of some points in this quadrant has a very small negative influence on the quantity (i.e., increasing it). The bias increased because the number of misses was reduced, a value in the denominator of this quantity (see equation (7)). Overall, the values in the first two data columns of Table 2 show that the removal of these restart ramp-up windows improves the data-model comparison.

The numbers improve even more with the removal of the bad storm interval. While this is only 60 values being cut from the data, the POFD, TS, FAR, HSS, bias, and χ^2 values are all slightly better. Only the POD, which only includes hits and misses in the calculation, was unaffected by the removal of the 13–15 March 2017 interval.

A different contingency table is found by changing the threshold to the weak storm value of −30 nT. The counts and derived quantities for the three SWMF data sets shown in Figures 3a–3c are listed in Table 3. The derived values are, mostly, a bit worse than the corresponding entries in Table 2, when the threshold was set to the moderate storm level of −50 nT: TS is ~10% lower; POD, HSS, and χ^2 are ~20% lower; POFD is about 3 times higher (but still small, well below 0.1); FAR is very slightly lower; and the bias is closer to unity for the lower threshold. This indicates that the model is slightly better at capturing the Dst time series during

Table 3
Dst Comparison Contingency Table Values With an Event Threshold of −30 nT or Below (Including Weak Storm Detection)

	All available times	With a 3-hr restart gap	Removing bad storm
Usable number of values	15125	13650	13590
Number of hits	873	857	857
Number of misses	951	720	720
Number of false alarms	827	824	786
Number of correct negatives	12474	11249	11227
Probability of detection	0.48	0.54	0.54
Probability false detection	0.062	0.068	0.065
Threat score	0.33	0.36	0.36
False alarm ratio	0.49	0.49	0.48
Heidke skill score	0.43	0.46	0.47
Frequency bias	0.93	1.07	1.04
Measure of significance, χ^2	2788.	2917.	2997.

moderate storms than during weak storm intervals. In addition, similar to Table 2, the results are, in general, better with the removal of the 3-hr window after each restart of the code (except POFD, for the same reason as given above) and even better with the additional removal of the bad storm interval.

A final plot to present here is the ROC curve, shown in Figure 4 for the three versions of SWMF output. The curves *start* in the upper right corner, where the first threshold of +10 nT has both a POD and a POFD close to 1. As the threshold is lowered to more active values, the progression is a systematic lowering of both of these quantities. Comparing with the POD and POFD entries in Tables 2 and 3, the location on this curve for *storm time* values of Dst is very far to the left, with POFD values near zero and POD values in the 0.5 to 0.7 range. The expectation of a *good* model is that the POFD drops faster than the POD as a function of threshold values. All three ROC curves are always above the unity-slope line, which satisfies this measure and reveals that the SWMF model does well at capturing *events* in the observed Dst regardless of the activity threshold level. For any POFD value, the blue curve is above the red one, showing that the removal of the restart ramp-up times improves the prediction Dst for all threshold settings. Futhermore, the green curve is above the blue one, showing that the removal of the bad storm interval improves the model's ability to reproduce the observed Dst values even more. The nonmonotonicity of the curves at POFD values below 0.1 is due to the non-Gaussian distribution of the observed and modeled Dst values during active geomagnetic times (Dst below −30 nT). The curve continues to the −80 nT final threshold applied for this analysis, at which point the POFD value is very close to zero.

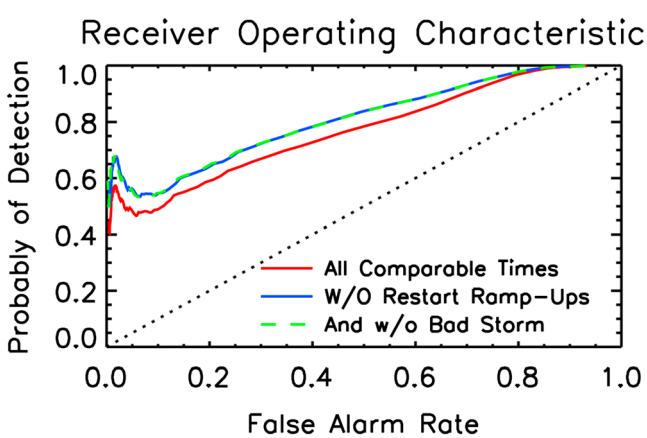

Figure 4. Receiver operating characteristic curves for the SWMF with all points included (red curve), with a 3-hr window removed after restarts (blue curve), and with the additional removal of the bad storm values in March 2017 (dashed green curve). The black dashed line shows the unity slope, for reference. SWMF = Space Weather Modeling Framework.

3.3. Assessing the Restart Ramp-Up Time

Thus far, the results presented have focused on the exclusion of a 3-hr window after each restart of the code. Figure 5 explains why this length of window was chosen for the tables and plots given in section 3.2. Here the window length of the removed interval after restarts is varied from 0 to 7 hr and, using a Dst event threshold of −50 nT, the resulting contingency table quantities of POD, POFD, FAR, TS, and HSS are shown. The zero-window-length values are the same as the *all available times* values listed in Table 2.

It is seen that all of the values increase as the restart ramp-up interval that was removed increases from 0 to 3 hr. Three of the four quantities reach a maximum in the 3- to 5-hr range and then start decreasing. The reason for the increase is obvious; the times during the magnetospheric ramp-up are being removed and this process removes misses from the comparison. The subsequent decrease for larger restart interval removals, however, is because most of these restarts are occurring during active times, and the longer intervals are starting to remove hits from the database rather than misses. These three (POD, TS, and HSS) are ones that we want to maximize. The remaining quantities (POFD and FAR), which are ones that we want to

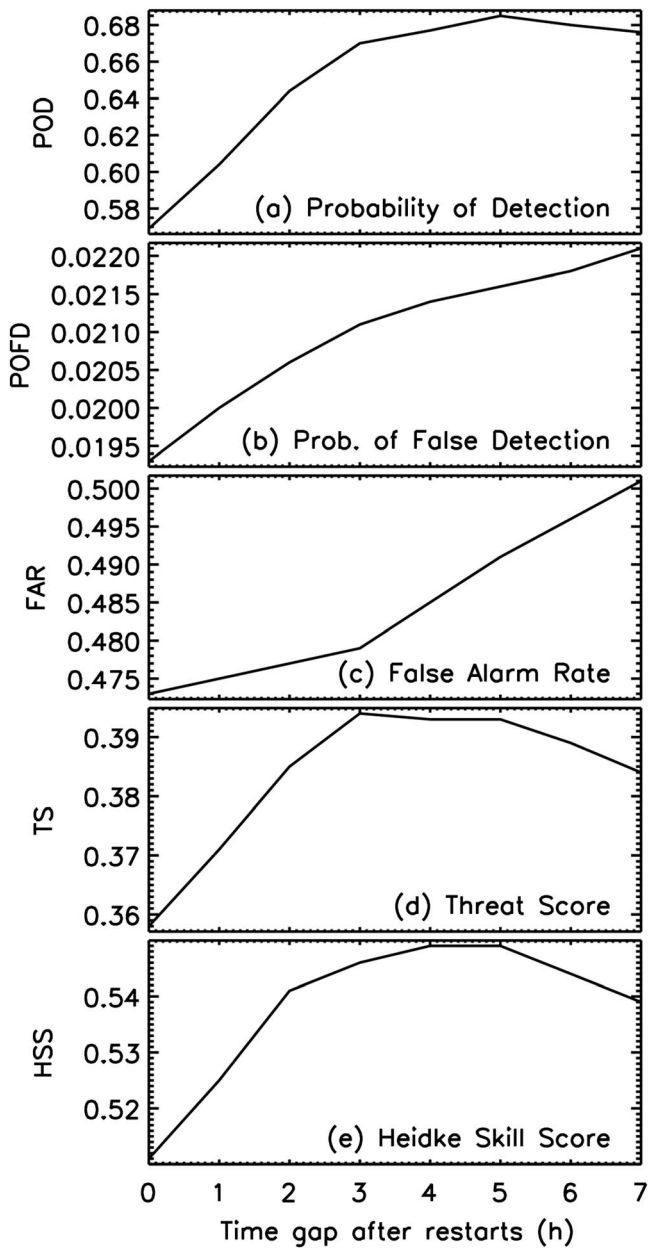

Figure 5. Contingency table metrics as a function of the number of hourly values removed after a restart of the code, for a Dst threshold of −50 nT: (a) POD, (b) POFD, (c) FAR, (d) TS, and (e) HSS. POD = probability of detection; POFD = probability of false detection; FAR = false alarm ratio; TS = threat score; HSS = Heidke skill score.

minimize, monotonically increase with the length of this window. Therefore, there is no preferred window length that optimizes all five of these quantities. A trade-off must be achieved between these competing variations with window length. A window length of 3 hr was chosen as a compromise that optimized most of these parameters yet retained as much of the full model output set as possible. Note that the *y* axis ranges on the panels in Figure 5 are quite narrow; the number of points being removed from the data set is small (~10%) compared to the total.

3.4. Comparison Against Persistence

A useful comparison to conduct is to assess the data-model goodness of fit against persistence of the data itself. Persistence is the ability of the observed data set to be used to predict itself at some later time. At zero time lag, the autocorrelation of the data is, of course, perfect. The similarity starts to erode with time shift, and eventually the ability of an observed Dst value should fall below the ability of the SWMF model.

Figure 6 explores when the model becomes better than persistence, plotting contingency table parameters (for a Dst event threshold of −50 nT) as a function of time shift between observed real-time Dst values. The black curve in each panel is the time-lagged autocorrelation of the Kyoto real-time Dst value, with a shift from 0 to 24 hr, while the blue, red, and green horizontal lines indicate the values of these parameters from the three data-model comparisons listed in Table 2. At zero time lag, the autocorrelation compares the observed values against themselves, so the correlation is perfect. At 1-hr lag, the statistics are still very good, because Dst does not change very quickly most of the time. As the time lag increases, the metrics of how well the observed Dst can predict itself continue to get worse. Because they all start at perfection for zero time lag and then get worse with increasing time lag, at some point, all of these metrics will cross over the value for the data-model comparison. The better the model is at reproducing the observed Dst values, then the shorter the time lag where this crossover will occur. In Figure 6a, for POD, the crossover point at which the model becomes better than persistence occurs at a 4- to 7-hr shift. In Figures 6b and 6c, for POFD and FAR, the crossover point is farther out at 10 to 12 hr. In both Figures 6d and 6e (TS and HSS), the crossover occurs in the 7- to 8-hr range.

3.5. Results for the Run Without RCM

As mentioned very briefly at the beginning of section 2, CCMC also has a version of SWMF running in experimental real-time mode, since 2007, that does not include an inner magnetospheric drift physics model. This older real-time run is different from the newer SWMF experimental real-time simulation in two other ways—the grid resolution is a bit more coarse (~700 k cells instead of ~1 million) and the inner boundary condition density setting is a static 28 amu/cm^3 rather than including a dependence on the cross polar cap potential. The big difference, however, is the omission of the RCM inner magnetosphere model from the setup configuration. One good aspect of this run is that it is robust and rarely requires a restart of the code. This version, however, is not nearly as good at capturing the variation of Dst as the other, newer SWMF simulation with the RCM included.

An example output from this version of the model is shown in Figure 2—the red curve in each panel that remains close to zero throughout the 3-month interval. A data-model scatterplot, shown in Figure 3d, reveals that the model never achieves a Dst value below −30 nT; that is, it never predicts even a weak magnetic storm in this index. The Pearson correlation coefficient is 0.33, which is statistically significant but not

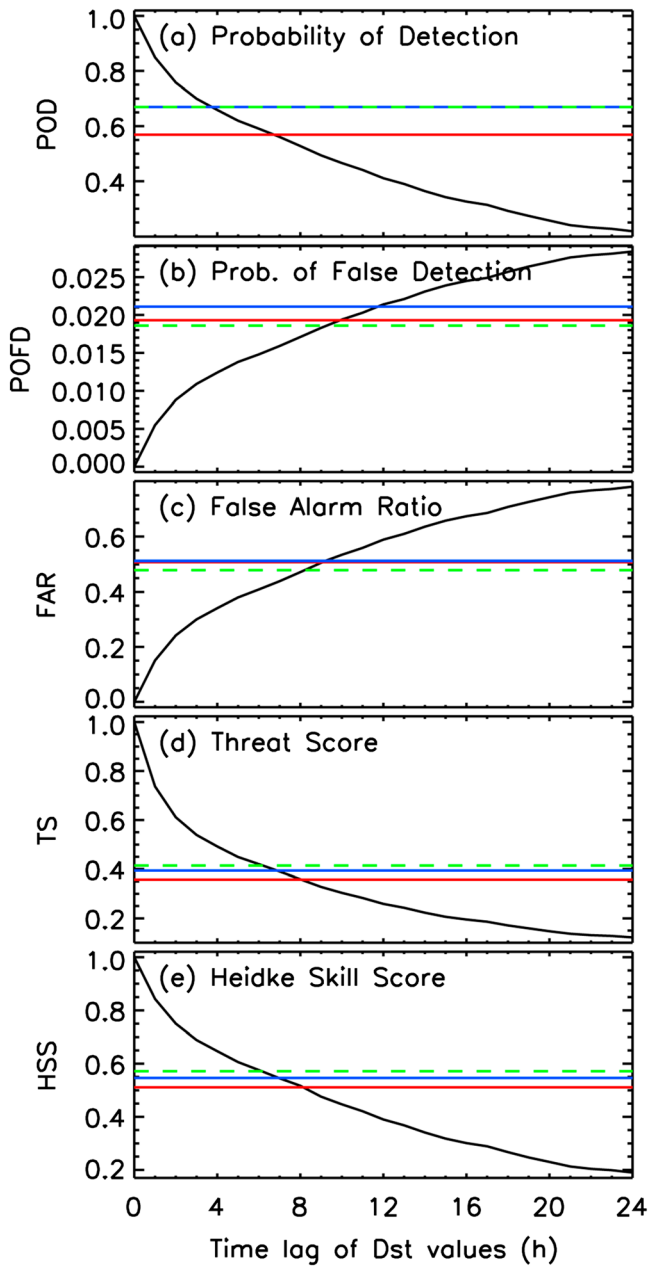

Figure 6. Contingency table metrics for the persistence calculation of the real-time Kyoto Dst values (i.e., their ability to predict themselves): (a) POD, (b) POFD, (c) FAR, (d) TS, and (e) HSS. The red horizontal line is the value for the SWMF comparison against the observed values, the blue line is the value for the SWMF code with a 3-hr window removed after each restart of the code, and the green-dashed line is the value for the SWMF code with both the restart ramp-up and the bad storm removed. POD = probability of detection; POFD = probability of false detection; FAR = false alarm ratio; TS = threat score; HSS = Heidke skill score; SWMF = Space Weather Modeling Framework.

suitable for predicting the observed values. A particularly low metric is the PE, which, at −0.54, indicates that this Dst output time series is worse than random guessing. It is not worth showing contingency table values for this comparison—because there are no hits or false alarms, all of the derived values are zero.

A useful comparison to make between the two versions of the SWMF experimental real-time outputs is their linear fits against the real-time Kyoto Dst time series. Given the linear regression formula (e.g., Taylor, 1997, p. 181),

$$y_{\text{fit}} = A + B{\cdot}x \tag{11}$$

where A is the intercept, B is the slope, and x is the independent variable, which is the Kyoto real-time Dst time series for this calculation (note that this is flipped from the axes shown in Figure 3). The fits of the two runsets reveal the amount of low-latitude surface level magnetic perturbation captured by each model configuration. Because the main difference is the inclusion of an inner magnetospheric drift physics model, a comparison between these two values reveals the contribution of near-Earth kinetic physics to the storm time Dst index.

Regression values are given in Table 4. For the older SWMF setup without RCM included, the intercept and slope values are 3.07 nT and 0.036 (nT/nT), respectively. For the newer SWMF setup (i.e., the one with the RCM), we use the results for which the restart ramp-up times and the bad storm interval have been removed (i.e., the scatterplot shown in Figure 3c). The intercept and slope values are −5.52 nT and 0.640 (nT/nT), respectively. The slope is less than unity, indicating that the model is underpredicting the observed values, but not nearly as much as the older SWMF setup.

A comparison of these regression curves for a few key Dst values (−30, −50, and −100 nT) is also listed in Table 4. The difference is large, showing that nearly all of the Dst value is due to kinetic drift physics of kiloelectron volt-energy particles on closed magnetic field lines.

4. Discussion

The Dst comparison results presented above demonstrate that the SWMF code running in experimental real-time mode at the CCMC reproduces this integrative metric reasonably well during both quiet times and storm intervals. Only one parameter was considered in the data-model comparisons: real-time Dst. As a measure of the near-Earth low-latitude and near-equatorial current systems, accurate reproduction of Dst is a measure of the goodness of the code to reproduce large-scale dynamics of geospace and, in particular, the plasma pressure at the near-Earth edge of the nightside plasma sheet and formation of and systematic interplay between the tail, partial ring, banana, and symmetric ring currents during active times. Currents in the inner magnetosphere contributing to Dst play a critical role in distorting the magnetic field governing radiation belt drift motion and are important for other space weather effects (e.g., Ganushkina, Liemohn, et al., 2015, Ganushkina et al., 2017, 2018). The comparisons in section 3 quantify how well this version of the SWMF code does at capturing storm physics. One metric that was not that good was the FAR, which was near 0.5 for storm-level threshold settings.

It is useful to compare the metrics calculated here with those determined for other online Dst prediction studies. Using their three differential equations and dozens of free parameters, the scheme of Temerin and Li

Table 4
Regression for the Newer and Older SWMF Configurations Against the Real-Time Kyoto Dst

	Newer SWMF (with RCM), with removal of restart ramp-ups and bad storm interval—seen in Figure 3c	Older SWMF (without RCM)—seen in Figure 3d
Intercept	−5.52 nT	+3.07 nT
Slope	0.640 [nT/nT]	0.036 [nT/nT]
Fit value at −30 nT	−24.7 nT	+2.0 nT
Fit value at −50 nT	−37.5 nT	+1.3 nT
Fit value at −100 nT	−69.5 nT	−0.5 nT

Note. SWMF = Space Weather Modeling Framework; RCM = Rice Convection Model.

(2006) yielded R = 0.96, PE = 0.91, and RMSE = 6.7 nT. Saiz et al. (2008), with their scheme of predicting Dst < -100 nT intense storm events from IMF Bz only, obtained a POD from 0.24 to 0.48, depending on the model parameters. With the NARMAX model, Boynton et al. (2011) were able to achieve R = 0.84 for their model-predicted output (their one-step-ahead method produced R = 0.98). Bala and Reiff (2012), using their artificial neural network scheme and the Boyle Index for the upstream solar wind input, found 6-hr lead time metrics of R = 0.80 and RMSE = 10.3 nT. Using a different neural network algorithm, the Dst prediction model of Revallo et al. (2014) had R = 0.74 and PE = 0.44.

In general, all of these models, expect that from the Saiz et al. (2008) study, are better than the SWMF experimental real-time output from CCMC. They were tuned to predict Dst, however, while the SWMF has not been fine tuned in this way. Instead, it is trying to solve the physics of the geospace system, with Dst as a byproduct of the solution. The output is far more extensive than just a single integrative perturbation index, with MHD quantities throughout the magnetosphere, energy-dependent plasma values in the inner magnetosphere, and ionospheric electrodynamic quantities over the globe. All of these output values can, and should, be evaluated for their potential application to space weather needs.

A source of error in this comparison are the uncertainties in the observations, both in the real-time Dst time series and in the upstream solar wind conditions that are driving the real-time SWMF simulation. For the former, several previous studies have examined the similarities and differences between Dst and the various global indices that mimic its compilation technique (Katus & Liemohn, 2013; Wanliss & Showalter, 2006). By comparing different indices against each other, they concluded that there is a ~10-nT random error in these values during quiet times, when the index is close to zero, growing to be ~20% of the index value during disturbed times (Dst < -50 nT). There are also systematic offsets between the various storm-monitoring global indices of ~5 nT and for a few index-index comparisons, up to 10 nT. Therefore, any data-model comparison that is better than ~15 nT is very reasonable. This is the case here for this version of SWMF.

For the latter, there are uncertainties in the upstream condition not only from the in situ measurements of the plasma and field but also from the propagation of those values from the satellite to the magnetopause. During active times, the uncertainties in the measurements increase because of enhanced background radiation from solar energetic particles. The level of contamination during strong solar energetic particle events is highly dependent on sensor design (see Gilbert et al. (2014), for a detailed description of this effect). There are also a number of different methods for propagating the solar wind from measurement location to Earth (e.g., Bargatze et al., 2005; Haaland et al., 2010; Pulkkinen & Rastätter, 2009; Weimer & King, 2008; Weimer et al., 2002, 2003). These combine to yield a fractional uncertainty of 20% on the values and a timing uncertainty of ±15 min (e.g., Cash et al., 2016; Crooker et al., 1982; Horbury et al., 2001; Kelly et al., 1986; Richardson et al., 1998; Ridley, 2000). In addition, Zheng et al. (2013) discussed the lessons learned about ICME forecasting at the CCMC and Posner et al. (2014) emphasized the necessity of a robust fleet of well-instrumented spacecraft assets for supporting space weather nowcasting. Barnard et al. (2017) highlight the problems from having too sparse of a solar wind observation set, arguing that there are inconsistencies when using STEREO-A and STEREO-B imaging velocities, unexplained by current understanding of ICMEs, that invalidate their combined use in forecasting. Multiple and more robust solar wind measurements far upstream of Earth, with full anticoincidence detector technology to mitigate background noise, would reduce this source of uncertainty in space weather forecasting. Defining the error in the resulting SWMF model results would require an ensemble run set. Because this is only an experimental real-time simulation, only one input stream is considered. This is an excellent topic for future work.

Another source of error is with the assumptions built in to the SWMF modeling approach. The version of BATS-R-US employed for this experimental real-time run uses a single-fluid ideal MHD equation set, yet there are versions of this code with more sophisticated mathematics, such as Hall terms (Tóth et al., 2008), multiple fluids for each ion species (e.g., Glocer, Tóth, Ma, et al., 2009; Ilie et al., 2015), and anisotropic pressure equations (Meng, Toth, Sokolov, et al., 2012; Meng, Toth, Liemohn, et al., 2012). Ridley et al. (2010) conducted a systematic investigation of the influence of numerical scheme and grid size on the accuracy of the magnetospheric solution, finding substantial differences between the choices of solvers, limiters, and mesh descriptions in BATS-R-US. Indeed, there are new numerical schemes now implemented in BATS-R-US, like the fifth-order solver by Chen et al. (2016). The code also uses a passive inner boundary condition (see Welling & Liemohn, 2014), even though more sophisticated ionospheric outflow models exist (Glocer, Toth, Gombosi, et al., 2009; Welling et al., 2016). The inner magnetospheric drift physics model in this configuration is the RCM, which assumes plasma isotropy, and, in this version, has a constant charge exchange attenuation factor of 10 hr. There are other models that resolve pitch angle and include additional physical processes yet still have electrodynamic magnetosphere-ionosphere coupling (e.g., Fok et al., 2001, 2014; Glocer et al., 2013; Ilie et al., 2012; Liemohn et al., 2005; Welling et al., 2015). The ionospheric electrodynamics model in this version assumes a Poisson equation solution for the current closure and electric potential, and the conductance model is rather limited in its validity (see Welling et al., 2017). This aspect of the solution could be improved with self-consistent precipitation effects on the ionosphere and a full 3-D calculation of the ionosphere-thermosphere system (e.g., Perlongo et al., 2017). Futhermore, there are other regional codes that could be added to the solution, such as kinetic modeling of the magnetic reconnection regions with a particle-in-cell code (Chen et al., 2017; Daldorff et al., 2014). In summary, there are many reasons on the numerical side why this comparison is not perfect, and with additional computing resources devoted to real-time simulations, a more robust calculation of geospace could be conducted.

One recent study that has considered an extended run-time simulation of the SWMF model is Haiducek et al. (2017). They conducted several simulations of a full month interval, January 2005, a time with several storms and many substorms. One of their runsets was very similar to the configuration used here, while two others had higher grid resolution in the BATS-R-US simulation domain, one with and one without inclusion of the RCM. They found that the nominal run setting was also quite good at reproducing the Dst time series (final Dst, in their case, rather than real-time Dst), and that the use of a higher-resolution grid provided only marginal improvement to the data-model comparison.

The poor reproduction of Dst from the run without an inner magnetospheric drift physics model is not at all surprising. The lack of this module means that it does not develop a large plasma pressure in the near-Earth nightside magnetosphere, even during strong solar wind driving. This has been known since the initial coupling of the RCM with the BATS-R-US MHD model by De Zeeuw et al. (2004). That study, however, only considered steady idealized solar wind inputs, reporting on the relaxation of the model output to an asymptotic solution. No comparison was made against observations. At CCMC, this version without an inner magnetosphere model has been nearly continuously run for a decade. Moreover, it is not the coarse numerical grid causing this lack of inner magnetospheric pressure buildup. The run without RCM in Haiducek et al. (2017) used a high-resolution grid in the MHD domain, and the difference between that run and its counterpart with RCM included is vast. Examining Figure 5b of Haiducek et al. (2017) reveals that there are essentially no values of Dst less than -30 nT produced by their *Hi-res w/o RCM* simulation. Figure 6b of that study shows that there are, in fact, a few such values during storm sudden commencement, but during the main phase and recovery phase of storms, this code yields very little low-latitude magnetic perturbation on the surface of the Earth.

This result from Haiducek et al. (2017) can help us interpret the results in the present study. Specifically, it is not the slightly lower grid resolution that is causing the older SWMF result to produce near-zero Dst but rather the absence of an inner magnetospheric drift physics model. The analysis presented here, of several years from the CCMC simulation output set, solidly confirms that accurately simulating the kinetic physics of the inner magnetosphere is essential to reproduce the Dst index time series, especially during storm intervals.

Other studies have examined the breakdown of which current systems contribute to the Dst perturbation at Earth's surface. Most of these studies find that the majority contributors to Dst during storms are near-Earth magnetospheric currents, specifically the symmetric ring current, partial ring current, banana current, or the

inner portion of the cross-tail current (e.g., Dubyagin et al., 2014; Ganushkina et al., 2004, 2010; Liemohn et al., 2001, 2015; Turner et al., 2000). Using the same code as the CCMC experimental real-time runs, with nearly the same numerical configuration, Liemohn et al. (2011, 2013) analyzed an ICME and a CIR storm, respectively, finding that the SWMF with a drift physics code produces strong currents inside of 10 R_E, enough to account for nearly all of the storm time Dst signature. While the present study has not examined the current systems in detail from the real-time simulations, it is expected to be the same and the comparison of the two real-time versions of the SWMF code (with and without a drift physics model) verifies that the previous analyses should hold true.

It has been emphasized many times that this simulation is only experimental. Much effort is needed to convert any code from research quality to the robust requirements of operational usage. This has been done for a version of the SWMF model by the NOAA SWPC. The version of the code that has been converted to 24/7 real-time use by SWPC has the same three numerical model components as the run presented here. Therefore, the validation presented here should be applicable to the NOAA SWPC version, with this study's findings serving as a lower limit to the quality of the SWPC output. The experimental real-time runs are CCMC that have been going for a longer time span than the SWPC version, though, so analysis of these results is useful to assess the nowcasting capabilities of this model. Note, however, that the primary motivation for SWPC to implement their version of this model is the prediction of geomagnetically induced currents. That version is focused on accurately modeling large time variations of the ground magnetic perturbation from space currents, which are substorm related rather than storm related and are not captured by the 1-hr Dst index. That is, this study is a good first step but does not fully validate their version for their purposes.

The comparisons against observational persistence provide guidance for solar wind prediction capabilities. It was shown above that persistence out to 4–8 hr, depending on the parameter, is better than the SWMF code output. This poses a challenge to real-time solar wind observations or models. That is, given this version of SWMF, it would be necessary to have predictive capabilities of the solar wind of this timescale (4–8 hr) in order for this geospace model to be better than simply using the latest observed value. Furthermore, persistence cannot reveal the presence of an approaching shock front and related disturbed solar wind conditions. If it is the onset of activity that is desired to be predicted, then numerical models are indeed better than persistence.

5. Conclusions

A particular configuration of the SWMF model has been running at the CCMC for several years. The results have been analyzed and compared against a single data set, the real-time Dst index, to assess the ability of this code to simulate the timing and intensity of magnetic storms. In addition to considering example intervals and scatterplots of the observed and simulated Dst values against each other, statistical parameters were calculated to quantify the goodness of the reproduction. Furthermore, contingency tables were calculated for several settings of the event threshold, in particular −50 and −30 nT, for moderate and weak magnetic storms, respectively. A ROC curve was calculated by varying the threshold across a wide range of possibilities.

The conclusion is that this version of the code does reasonably well, offering promise toward first principles-based forecasting of geospace. It does even better when a short, few-hour window of output is removed following each restart of the code and the additional removal a bad storm interval where the solar wind input was clearly wrong. The Pearson correlation coefficient is 0.69, which yields a coefficient of determination of 0.47—that is, nearly half of the variation of the Kyoto real-time Dst is captured by the SWMF model. One metric that is not particularly good is the FAR, which is ~0.5 for storm thresholds, indicating that there are observed storm conditions for only about half of the time when the model calculates storm-level Dst values. The code does well regardless of the Dst threshold setting chosen, with the ROC curve being always above the unity slope line. For storm time thresholds of −30 nT or lower, the probability of false detection is essentially zero and both the POD and the HSS are over 0.5. It was shown that the code is actually a bit better at predicting the Dst values during moderate storms than during weak storms.

This assessment of the output from the experimental real-time run of the SWMF demonstrates that the model is a useful numerical tool for simulating the low-latitude magnetic perturbation, a signal dominated by currents in near-Earth space, during magnetic storm conditions. While it is not a robust space weather

operational code, the CCMC provides the code output as an open resource to both the space physics research community and space weather users. Several websites are available for accessing the output from this runset.

Finally, a comparison was made against another configuration of the SWMF that is also running in experimental real-time mode at the CCMC, but without an inner magnetospheric drift physics model included in the setup. Within the 2+ year interval considered in this study, this other version never predicts a storm-level value of Dst below −30 nT. A comparison of linear fit values quantifies that nearly all of the storm time Dst value is from energy-dependent drift physics of particles on closed magnetic field lines. This confirms the necessity of accurate kinetic physics in the inner magnetosphere in order to simulate storms.

Acknowledgments
The authors would like to thank the U.S. government for sponsoring this research, in particular research grants from NASA (NNX11AO60G, NNX14AC02G, NNX16AG66G, NNX17AI48G, and NNX17AB87G) and NSF (AGS-1102863 and AGS-1414517). The part of the research done by M. Liemohn, N. Ganushkina, and Bart van der Holst has received funding from the European Union Horizon 2020 Research and Innovation Programme under grant agreement 637302 PROGRESS. The simulations were conducted on the computing facilities at NASA GSFC, and the run output is freely available on their website (https://ccmc.gsfc.nasa.gov/cgi-bin/SWMFpred.cgi) and the CCMC iSWA interactive tool (https://ccmc.gsfc.nasa.gov/iswa/). The real-time solar wind data were provided by NOAA SWPC (http://www.swpc.noaa.gov/products/real-time-solar-wind). The authors thank the World Data Center in Kyoto, Japan, for the real-time Dst values (http://wdc.kugi.kyoto-u.ac.jp/dst_realtime/presentmonth/index.html).

References

Amariutei, O. A., & Ganushkina, N. Y. (2012). On the prediction of the auroral westward electrojet index. *Annales Geophysicae*, *30*(5), 841–847. https://doi.org/10.5194/angeo-30-841-2012

Ayala Solares, J. R., Wei, H.-L., Boynton, R. J., Walker, S. N., & Billings, S. A. (2016). Modeling and prediction of global magnetic disturbance in near-Earth space: A case study for Kp index using NARX models. *Space Weather*, *14*, 899–916. https://doi.org/10.1002/2016SW001463

Bala, R., & Reiff, P. (2012). Improvements in short-term forecasting of geomagnetic activity. *Space Weather*, *10*, S06001. https://doi.org/10.1029/2012SW000779

Bala, R., & Reiff, P. (2014). Validating the Rice neural network and the Wing Kp real-time models. *Space Weather*, *12*, 417–425. https://doi.org/10.1002/2014SW001075

Bala, R., Reiff, P. H., & Landivar, J. E. (2009). Real-time prediction of magnetospheric activity using the Boyle index. *Space Weather*, *7*, S04003. https://doi.org/10.1029/2008SW000407

Balan, N., Ebihara, Y., Skoug, R., Shiokawa, K., Batista, I. S., Tulasi Ram, S., et al. (2017). A scheme for forecasting severe space weather. *Journal of Geophysical Research: Space Physics*, *122*, 2824–2835. https://doi.org/10.1002/2016JA023853

Balikhin, M. A., Rodriguez, J. V., Boynton, R. J., Walker, S. N., Aryan, H., Sibeck, D. G., & et al. (2016). Comparative analysis of NOAA REFM and SNB3GEO tools for the forecast of the fluxes of high-energy electrons at GEO. *Space Weather*, *14*, 22–31. https://doi.org/10.1002/2015SW001303

Bargatze, L. F., McPherron, R. L., Minamora, J., & Weimer, D. (2005). A new interpretation of Weimer et al.'s solar wind propagation delay technique. *Journal of Geophysical Research*, *110*, A07105. https://doi.org/10.1029/2004JA010902

Barnard, L. A., deKoning, C., Scott, C. J., Owens, M. J., Wilkinson, J., & Davies, J. A. (2017). Testing the current paradigm for space weather prediction with heliospheric imagers. *Space Weather*, *15*, 782–803. https://doi.org/10.1002/2017SW001609

Bekerat, H. A., Schunk, R. W., & Scherliess, L. (2003). Evaluation of statistical convection patterns for real-time ionospheric specifications and forecasts. *Journal of Geophysical Research*, *108*(A12), 1413. https://doi.org/10.1029/2003JA009945

Bell, J. T., Gussenhoven, M. S., & Mullen, E. G. (1997). Super storms. *Journal of Geophysical Research*, *102*, 14,189–14,198. https://doi.org/10.1029/96JA03759

Berghmans, D., Van der Linden, R. A. M., Vanlommel, P., Warnant, R., Zhukov, A., Robbrecht, E., et al. (2005). Solar activity: Nowcasting and forecasting at the SIDC. *Annales Geophysicae*, *23*(9), 3115–3128. https://doi.org/10.5194/angeo-23-3115-2005

Boynton, R. J., Balikhin, M. A., & Billings, S. A. (2015). Online NARMAX model for electron fluxes at GEO. *Annales Geophysicae*, *33*(3), 405–411. https://doi.org/10.5194/angeo-33-405-2015

Boynton, R. J., Balikhin, M. A., Billings, S. A., Sharma, A. S., & Amariutei, O. A. (2011). Data derived NARMAX Dst model. *Annales Geophysicae*, *29*(6), 965–971. https://doi.org/10.5194/angeo-29-985-2011

Bussy-Virat, C. D., & Ridley, A. J. (2016). Twenty-four hour predictions of the solar wind speed peaks by the probability distribution function model. *Space Weather*, *14*, 861–873. https://doi.org/10.1002/2016SW001437

Carovillano, R. L., & Siscoe, G. L. (1973). Energy and momentum theorems in magnetospheric processes. *Reviews of Geophysics*, *11*(2), 289.

Cash, M. D., Witters Hicks, S., Biesecker, D. A., Reinard, A. A., de Koning, C. A., & Weimer, D. R. (2016). Validation of an operational product to determine L1 to Earth propagation time delays. *Space Weather*, *14*, 93–112. https://doi.org/10.1002/2015SW001321

Chen, Y., Toth, G., Cassak, P., Jia, X., Gombosi, T. I., Slavin, J. A., et al. (2017). Global three-dimensional simulation of Earth's dayside reconnection using a two-way coupled magnetohydrodynamics with embedded particle-in-cell model: Initial results. *Journal of Geophysical Research: Space Physics*, *122*, 10,318–10,335. https://doi.org/10.1002/2017JA024186

Chen, Y., Toth, G., & Gombosi, T. I. (2016). A fifth-order finite difference scheme for hyperbolic equations on block-adaptive curvilinear grids. *Journal of Computational Physics*, *305*, 604–621. https://doi.org/10.1016/j.jcp.2015.11.003

Codrescu, M. V., Negrea, C., Fedrizzi, M., Fuller-Rowell, T. J., Dobin, A., Jakowski, N., et al. (2012). A real-time run of the Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics (CTIPe) model. *Space Weather*, *10*, S02001. https://doi.org/10.1029/2011SW000736

Costello, K. A. (1997). Moving the Rice MSFM into a real-time forecast mode using solar wind driven forecast models. Ph.D. dissertation, Rice University, Houston, Texas.

Crooker, N. U., Siscoe, G. L., Russell, C. T., & Smith, E. J. (1982). Factors controlling degree of correlation between ISEE 1 and ISEE 3 interplanetary magnetic field measurements. *Journal of Geophysical Research*, *87*(A4), 2224–2230.

Daldorff, L. K. S., Toth, G., Gombosi, T. I., Lapenta, G., Amaya, J., Markidis, S., & et al. (2014). Two-way coupling of a global Hall magnetohydrodynamics model with a local implicit particle-in-cell model. *Journal of Computational Physics*, *268*, 236–254. https://doi.org/10.1016/j.jcp.2014.03.009

De Zeeuw, D. L., Sazykin, S., Wolf, R. A., Gombosi, T. I., Ridley, A. J., & Toth, G. (2004). Coupling of a global MHD code and an inner magnetosphere model: Initial results. *Journal of Geophysical Research*, *109*, A12219. https://doi.org/10.1029/2003JA010366

Devos, A., Verbeeck, C., & Robbrecht, E. (2014). Verification of space weather forecasting at the Regional Warning Center in Belgium. *Journal of Space Weather and Space Climate*, *4*, A29.

Dubyagin, S., Ganushkina, N. Y., Kubyshkina, M., & Liemohn, M. W. (2014). Contribution from different current systems to SYM and ASY midlatitude indices. *Journal of Geophysical Research: Space Physics*, *119*, 7243–7263. https://doi.org/10.1002/2014JA020122

Fok, M.-C., Buzulukova, N. Y., Chen, S.-H., Glocer, A., Nagai, T., Valek, P., & et al. (2014). The comprehensive inner magnetosphere-ionosphere model. *Journal of Geophysical Research: Space Physics*, *119*, 7522–7540. https://doi.org/10.1002/2014JA020239

Fok, M.-C., Wolf, R. A., Spiro, R. W., & Moore, T. E. (2001). Comprehensive computational model of the Earth's ring current. *Journal of Geophysical Research*, *106*(A5), 8417–8424.

Fry, C. D., Dryer, M., Smith, Z., Sun, W., Deehr, C. S., & Akasofu, S.-I. (2003). Forecasting solar wind structures and shock arrival times using an ensemble of models. *Journal of Geophysical Research*, *108*(A2), 1070. https://doi.org/10.1029/2002JA009474

Ganushkina, N., Liemohn, M., Kubyshkina, M., Ilie, R., & Singer, H. (2010). Distortions of the magnetic field by storm-time current systems in Earth's magnetosphere. *Annales Geophysicae*, *28*(1), 123–140.

Ganushkina, N. Y., Amariutei, O. A., Welling, D., & Heynderickx, D. (2015). Nowcast model for low-energy electrons in the inner magnetosphere. *Space Weather*, *13*, 16–34. https://doi.org/10.1002/2014SW001098

Ganushkina, N. Y., Liemohn, M. W., & Dubyagin, S. (2018). Current systems in the Earth's magnetosphere. *Reviews of Geophysics*, *56*, 309–332. https://doi.org/10.1002/2017RG000590

Ganushkina, N. Y., Liemohn, M. W., Dubyagin, S., Daglis, I., Dandouras, I., De Zeeuw, D. L., et al. (2015). Defining and resolving current systems in geospace. *Annales Geophysicae*, *33*(11), 1369–1402. https://doi.org/10.5194/angeo-33-1369-2015

Ganushkina, N. Y., Jaynes, A., & Liemohn, M. W. (2017). Space weather effects produced by the ring current particles. *Space Science Reviews*, *212*(3-4), 1315–1344. https://doi.org/10.1007/s11214-017-0412-2

Ganushkina, N. Y., Pulkkinen, T. I., Kubyshkina, M. V., Singer, H. J., & Russell, C. T. (2004). Long-term evolution of magnetospheric current systems during storms. *Annales Geophysicae*, *22*(4), 1317–1334.

Gilbert, J. A., Gershman, D. J., Gloeckler, G., Lundgren, R. A., Zurbuchen, T. H., Orlando, T. M., et al. (2014). Characterization of background sources in space-based time-of-flight mass spectrometers. *Review of Scientific Instruments*, *85*, 091301. https://doi.org/10.1063/1.4894694

Glocer, A., Fok, M., Meng, X., Toth, G., Buzulukova, N., Chen, S., & et al. (2013). CRCM + BATS-R-US two way coupling. *Journal of Geophysical Research: Space Physics*, *118*, 1635–1650. https://doi.org/10.1002/jgra.50221

Glocer, A., Toth, G., Gombosi, T., & Welling, D. (2009). Modeling ionospheric outflows and their impact on the magnetosphere, initial results. *Journal of Geophysical Research*, *114*, A05216. https://doi.org/10.1029/2009JA014053

Glocer, A., Tóth, G., Ma, Y., Gombosi, T., Zhang, J.-C., & Kistler, L. M. (2009). Multifluid Block-Adaptive-Tree Solar wind Roe-type Upwind Scheme: Magnetospheric composition and dynamics during geomagnetic storms—Initial results. *Journal of Geophysical Research*, *114*, A12203. https://doi.org/10.1029/2009JA014418

Gombosi, T. I., Toth, G., De Zeeuw, D. L., Hansen, K. C., Kabin, K., & Powell, K. G. (2002). Semi-relativistic magnetohydrodynamics and physics-based convergence acceleration. *Journal of Computational Physics*, *177*(1), 176–205.

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B. T., & et al. (1994). What is a geomagnetic storm? *Journal of Geophysical Research*, *99*(A4), 5771–5792.

Haaland, S., Munteanu, C., & Mailyan, B. (2010). Solar wind propagation delay: Comment on "Minimum variance analysis-based propagation of the solar wind observations: Application to real-time global magnetohydrodynamic simulations" by A. Pulkkinen and L. Raststätter. *Space Weather*, *8*, S06005. https://doi.org/10.1029/2009SW000542

Haiducek, J. D., Welling, D. T., Ganushkina, N. Y., Morley, S. K., & Ozturk, D. S. (2017). SWMF global magnetosphere simulations of January 2005: Geomagnetic indices and cross-polar cap potential. *Space Weather*, *15*, 1567–1587. https://doi.org/10.1002/2017SW001695

Harel, M., Wolf, R. A., Reiff, P. H., Spiro, R. W., Burke, W. J., Rich, F. J., & et al. (1981). Quantitative simulation of a magnetospheric substorm. 1. Model logic and overview. *Journal of Geophysical Research*, *86*(A4), 2217.

Heidke, P. (1926). Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungdienst (calculation of the success and goodness of strong wind forecasts in the storm warning service). *Geografika Annaler*, *8*, 301–349.

Henney, C. J., Toussaint, W. A., White, S. M., & Arge, C. N. (2012). Forecasting $F_{10.7}$ with solar magnetic flux transport modeling. *Space Weather*, *10*, S02011. https://doi.org/10.1029/2011SW000748

Hochedez, J.-F., Zhukov, A., Robbrecht, E., Van der Linden, R., Berghmans, D., Vanlommel, P., et al. (2005). Solar weather monitoring. *Annales Geophysicae*, *23*(9), 3149–3161. https://doi.org/10.5194/angeo- 23-3149-2005

Hogan, R. J., & Mason, I. B. (2012). Deterministic forecasts of binary events. In I. T. Jolliffe, & D. B. Stephenson (Eds.), *Forecast Verification* (pp. 31–60). Chichester, UK: Wiley-Blackwell. https://doi.org/10.1002/9781119960003.ch3

Horbury, T. S., Burgess, D., Fränz, M., & Owen, C. J. (2001). Prediction of Earth arrival times of interplanetary southward magnetic field turnings. *Journal of Geophysical Research*, *106*(A12), 30,001–30,009.

Horne, R. B., Glauert, S. A., Meredith, N. P., Boscher, D., Maget, V., Heynderickx, D., & et al. (2013). Space weather impacts on satellites and forecasting the Earth's electron radiation. *Space Weather*, *11*, 169–186. https://doi.org/10.1002/swe.20023

Howard, T. A., & Tappin, S. J. (2010). Application of a new phenomenological coronal mass ejection model to space weather forecasting. *Space Weather*, *8*, S07004. https://doi.org/10.1029/2009SW000531

Ilie, R., Liemohn, M. W., Toth, G., Ganushkina, N. Y., & Daldorff, L. K. S. (2015). Assessing the role of oxygen on ring current formation and evolution through numerical experiments. *Journal of Geophysical Research: Space Physics*, *120*, 4656–4668. https://doi.org/10.1002/2015JA021157

Ilie, R., Liemohn, M. W., Toth, G., & Skoug, R. (2012). Kinetic model of the inner magnetosphere with arbitrary magnetic field. *Journal of Geophysical Research*, *117*, A04208. https://doi.org/10.1029/2011JA017189

Jian, L. K., Macneice, P. J., Mays, M. L., Taktakishvili, A., Odstrcil, D., Jackson, B., et al. (2016). Validation for global solar wind prediction using Ulysses comparison: Multiple coronal and heliospheric models installed at the Community Coordinated Modeling Center. *Space Weather*, *14*, 592–611. https://doi.org/10.1002/2016SW001435

Katus, R. M., & Liemohn, M. W. (2013). Similarities and differences in low- to middle-latitude geomagnetic indices. *Journal of Geophysical Research: Space Physics*, *118*, 5149–5156. https://doi.org/10.1002/jgra.50501

Kelly, T. J., Crooker, N. U., Siscoe, G. L., Russell, C. T., & Smith, E. J. (1986). On the use of a sunward libration-point-orbiting spacecraft as an interplanetary magnetic field monitor for magnetospheric studies. *Journal of Geophysical Research*, *91*(A5), 5629–5636. https://doi.org/10.1029/JA091iA05p05629

Koren, B. (1993). A robust upwind discretisation method for advection, diffusion and source terms. In C. Vreugdenhil & B. Koren (Eds.), *Numerical methods for advection-diffusion problems* (pp. 117). Braunschweig: Vieweg.

Liemohn, M. W., De Zeeuw, D. L., Ganushkina, N. Y., Kozyra, J. U., & Welling, D. T. (2013). Magnetospheric cross-field currents during the January 6-7, 2011, high-speed stream-driven interval. *Journal of Atmospheric and Solar-Terrestrial Physics*, *99*, 78–84. https://doi.org/10.1016/j.jastp.2012.09.007

Liemohn, M. W., De Zeeuw, D. L., Ilie, R., & Ganushkina, N. Y. (2011). Deciphering magnetospheric cross-field currents. *Geophysical Research Letters*, *38*, L20106. https://doi.org/10.1029/2011GL049611

Liemohn, M. W., Katus, R. M., & Ilie, R. (2015). Statistical analysis of storm-time near-Earth current systems. *Annales Geophysicae*, *33*(8), 965–982. https://doi.org/10.5194/angeo-33-965-2015

Liemohn, M. W., Kozyra, J. U., Clauer, C. R., & Ridley, A. J. (2001). Computational analysis of the near-Earth magnetospheric current system. *Journal of Geophysical Research*, *106*(A12), 29,531–29,29542.

Liemohn, M. W., Ridley, A. J., Brandt, P. C., Gallagher, D. L., Kozyra, J. U., Mitchell, D. G., et al. (2005). Parametric analysis of nightside conductance effects on inner magnetospheric dynamics for the 17 April 2002 storm. *Journal of Geophysical Research*, *110*, A12S22. https://doi.org/10.1029/2005JA011109

Liu, H.-L., & Qin, G. (2015). Improvements of the shock arrival times at the Earth model STOA. *Journal of Geophysical Research: Space Physics*, *120*, 5290–5297. https://doi.org/10.1002/2015JA021072

Love, J. J., & Gannon, J. L. (2009). Revised Dst and the epicycles of magnetic disturbance: 1958–2007. *Annales Geophysicae*, *27*(8), 3101–3131.

Lundstedt, H., Gleisner, H., & Wintoft, P. (2002). Operational forecasts of the geomagnetic *Dst* index. *Geophysical Research Letters*, *29*(24), 2181. https://doi.org/10.1029/2002GL016151

Macneice, P., Elliott, B., & Acebal, A. (2011). Validation of community models: 3. Tracing field lines in heliospheric models. *Space Weather*, *9*, S10003. https://doi.org/10.1029/2011SW000665

Manoj, C., & Maus, S. (2012). A real-time forecast service for the ionospheric equatorial zonal electric field. *Space Weather*, *10*, S09002. https://doi.org/10.1029/2012SW000825

Marsh, M. S., Dalla, S., Dierckxsens, M., Laitinen, T., & Crosby, N. B. (2015). SPARX: A modeling system for solar energetic particle radiation space weather forecasting. *Space Weather*, *13*, 386–394. https://doi.org/10.1002/2014SW001120

Mason, I. B. (1982). A model for assessment of weather forecasts. *Australian Meteorological Magazine*, *30*, 291–303.

Mayaud, P. N. (1980). *Derivation, meaning, and use of geomagnetic indices*, *Geophys. Monogr. Ser.* (Vol. 22, pp. 154). Washington, DC: American Geophysical Union. https://doi.org/10.1029/GM022

Meng, X., Toth, G., Liemohn, M. W., Gombosi, T. I., & Runov, A. (2012). Pressure anisotropy in global magnetospheric simulations: A magnetohydrodynamics model. *Journal of Geophysical Research*, *117*, A08216. https://doi.org/10.1029/2012JA017791

Meng, X., Toth, G., Sokolov, I. V., & Gombosi, T. I. (2012). Classical and semirelativistic magnetohydrodynamics with anisotropic ion pressure. *Journal of Computational Physics*, *231*(9), 3610–3622. https://doi.org/10.1016/j.jcp.2011.12.042

Mozer, J. B., & Briggs, W. M. (2003). Skill in real-time solar wind shock forecasts. *Journal of Geophysical Research*, *108*(A6), 1262. https://doi.org/10.1029/2003JA009827

Muller, R. H. (1944). Verification of short-range weather forecasts (a survey of the literature). *Bulletin of the American Meteorological Society*, *25*, 18–27.

Murphy, A. H. (1996). The Finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, *11*(1), 3–20.

Murray, S. A., Bingham, S., Sharpe, M., & Jackson, D. R. (2017). Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather*, *15*, 577–588. https://doi.org/10.1002/2016SW001579

Norquist, D. C., & Meeks, W. C. (2010). A comparative verification of forecasts from two operational solar wind models. *Space Weather*, *8*, S12005. https://doi.org/10.1029/2010SW000598

O'Brien, T. P., & McPherron, R. L. (2000). Forecasting the ring current index Dst in real time. *Journal of Atmospheric and Solar Terrestrial Physics*, *62*, 1295–1299. https://doi.org/10.1016/S1364-6826(00)00072-9

Oler, C. (2004). Prediction performance of space weather forecast centers following the extreme events of October and November 2003. *Space Weather*, *2*, S08001. https://doi.org/10.1029/2004SW000076

Perlongo, N., Ridley, A., Liemohn, M. W., & Katus, R. M. (2017). The effect of ring current electron scattering rates on magnetosphere-ionosphere coupling. *Journal of Geophysical Research: Space Physics*, *122*, 4168–4189. https://doi.org/10.1002/2016JA023679

Posner, A., Hesse, M., & Cyr, O. C. S. (2014). The main pillar: Assessment of space weather observational asset performance supporting nowcasting, forecasting, and research to operations. *Space Weather*, *12*, 257–276. https://doi.org/10.1002/2013SW001007

Powell, K., Roe, P., Linde, T., Gombosi, T., & De Zeeuw, D. L. (1999). A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *Journal of Computational Physics*, *154*, 284–309. https://doi.org/10.1006/jcph.1999.6299

Pulkkinen, A., & Rastätter, L. (2009). Minimum variance analysis-based propagation of the solar wind observations: Application to real-time global magnetohydrodynamic simulations. *Space Weather*, *7*, S12001. https://doi.org/10.1029/2009SW000468

Pulkkinen, A., Rastätter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., et al. (2013). Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations. *Space Weather*, *11*, 369–385. https://doi.org/10.1002/swe.20056

Qin, G., Zhang, M., & Rassoul, H. K. (2009). Prediction of the shock arrival time with SEP observations. *Journal of Geophysical Research*, *114*, A09104. https://doi.org/10.1029/2009JA014332

Reiss, M. A., Temmer, M., Veronig, A. M., Nikolic, L., Vennerstrom, S., Schöngassner, F., & et al. (2016). Verification of high-speed solar wind stream forecasts using operational solar wind models. *Space Weather*, *14*, 495–510. https://doi.org/10.1002/2016SW001390

Revallo, M., Valach, F., Hejda, P., & Bochníček, J. (2014). A neural network Dst index model driven by input time histories of the solar wind–magnetosphere interaction. *Journal of Atmospheric and Solar-Terrestrial Physics*, *110–111*, 9–14. https://doi.org/10.1016/j.jastp.2014.01.011

Richardson, J. D., Dashecskiy, F., & Paularena, K. I. (1998). Solar wind plasma correlations between L1 and Earth. *Journal of Geophysical Research*, *103*(A7), 14,619–14,629.

Ridley, A. J. (2000). Estimation of the uncertainty in timing the relationship between magnetospheric and solar wind processes. *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(9), 757–771.

Ridley, A. J., Gombosi, T. I., & De Zeeuw, D. L. (2004). Ionospheric control of the magnetosphere: Conductance. *Annales Geophysicae*, *22*(2), 567–584.

Ridley, A. J., Gombosi, T. I., Sokolov, I. V., Toth, G., & Welling, D. T. (2010). Numerical considerations in simulating the global magnetosphere. *Annales Geophysicae*, *28*(8), 1589–1614.

Ridley, A. J., & Liemohn, M. W. (2002). A model-derived description of the penetration electric field. *Journal of Geophysical Research*, *107*(A8), 1151. https://doi.org/10.1029/2001JA000051

Riley, P. (2012). On the probability of occurrence of extreme space weather events. *Space Weather*, *10*, S02012. https://doi.org/10.1029/2011SW000734

Riley, P., & Love, J. J. (2017). Extreme geomagnetic storms: Probabilistic forecasts and their uncertainties. *Space Weather*, *15*, 53–64. https://doi.org/10.1002/2016SW001470

Saiz, E., Cid, C., & Cerrato, Y. (2008). Forecasting intense geomagnetic activity using interplanetary magnetic field data. *Annales Geophysicae*, *26*(12), 3989–3998. https://doi.org/10.5194/angeo-26-3989-2008

Savani, N. P., Vourlidas, A., Richardson, I. G., Szabo, A., Thompson, B. J., Pulkkinen, A., et al. (2017). Predicting the magnetic vectors within coronal mass ejections arriving at Earth: 2. Geomagnetic response. *Space Weather*, *15*, 441–461. https://doi.org/10.1002/2016SW001458

Schunk, R. W., Gardner, L., Scherliess, L., & Zhu, L. (2012). Problems associated with uncertain parameters and missing physics for long-term ionosphere-thermosphere forecasting. *Radio Science*, *47*, RS0L23. https://doi.org/10.1029/2011RS004911

Simpson, S. (2003). From research model to forecasting tool. *Space Weather*, *1*(1), 1009. https://doi.org/10.1029/2003SW000029
Sokolov, I. V., Zhang, H.-M., & Sakai, J. I. (2001). Simple and efficient godunov scheme for computational relativistic gas dynamics. *Journal of Computational Physics*, *172*, 209–234. https://doi.org/10.1006/jcph.2001.6821
Swets, J. A. (1973). The relative operating characteristic in psychology. *Science*, *182*(4116), 990–1000.
Taylor, J. R. (1997). *An introduction to error analysis*. Mill Valley, CA, USA: University Science Books.
Temerin, M., & Li, X. (2006). Dst model for 1995–2002. *Journal of Geophysical Research*, *111*, A04221. https://doi.org/10.1029/2005JA011257
Tobiska, W. K., Knipp, D., Burke, W. J., Bouwer, D., Odstrcil, D., Hagan, M. P., et al. (2013). The Anemomilos prediction methodology for Dst. *Space Weather*, *11*, 490–508. https://doi.org/10.1002/swe.20094
Toffoletto, F., Sazykin, S., Spiro, R., & Wolf, R. (2003). Inner magnetospheric modeling with the Rice Convection Model. *Space Science Reviews*, *107*(1/2), 175–196.
Toth, G., De Zeeuw, D. L., Gombosi, T. I., & Powell, K. G. (2006). A parallel explicit/implicit time stepping scheme on block-adaptive grids. *Journal of Computational Physics*, *217*(2), 722–758.
Tóth, G., Ma, Y. J., & Gombosi, T. I. (2008). Hall magnetohydrodynamics on block adaptive grids. *Journal of Computational Physics*, *227*(14), 6967–6984. https://doi.org/10.1016/j.jcp.2008.04.010
Toth, G., Meng, X., Gombosi, T. I., & Ridley, A. J. (2011). Reducing numerical diffusion in magnetospheric simulations. *Journal of Geophysical Research*, *116*, A07211. https://doi.org/10.1029/2010JA016370
Toth, G., & Roe, P. L. (2002). Divergence- and curl-preserving prolongation and restriction formulas. *Journal of Computational Physics*, *180*(2), 736–750.
Toth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C. R., De Zeeuw, D. L., et al. (2005). Space Weather Modeling Framework: A new tool for the space science community. *Journal of Geophysical Research*, *110*, A12226. https://doi.org/10.1029/2005JA011126
Toth, G., van der Holst, B., Sokolov, I. V., De Zeeuw, D. L., Gombosi, T. I., Fang, F., et al. (2012). Adaptive numerical algorithms in space weather modeling. *Journal of Computational Physics*, *231*(3), 870–903.
Tsagouri, I., & Belehaki, A. (2015). Ionospheric forecasts for the European region for space weather applications. *Journal of Space Weather and Space Climate*, *5*, A9. https://doi.org/10.1051/swsc/2015010
Tsagouri, I., Belehaki, A., Bergeot, N., Cid, C., Delouille, V., Egorova, T., et al. (2013). Progress in space weather modeling in an operational environment. *Journal of Space Weather and Space Climate*, *3*, A17.
Tsubouchi, K., & Kubo, Y. (2010). Quantitative assessment of the probability forecast for the geomagnetic storm occurrence. *Space Weather*, *8*, S12007. https://doi.org/10.1029/2010SW000614
Tulunay, E., Senalp, E. T., Radicella, S. M., & Tulunay, Y. (2006). Forecasting total electron content maps by neural network technique. *Radio Science*, *41*, RS4016. https://doi.org/10.1029/2005RS003285
Turner, D. L., & Li, X. (2011). Using spacecraft measurements ahead of Earth in the Parker spiral to improve terrestrial space weather forecasts. *Space Weather*, *9*, S01002. https://doi.org/10.1029/2010SW000627
Turner, N. E., Baker, D. N., Pulkkinen, T. I., & McPherron, R. L. (2000). Evaluation of the tail current contribution to Dst. *Journal of Geophysical Research*, *105*(A3), 5431–5439. https://doi.org/10.1029/1999JA000248
Wang, W., Killeen, T. L., Burns, A. G., & Reinisch, B. W. (2001). A real-time model-observation comparison of F2 peak electron densities during the upper atmospheric research collaboratory campaign of October 1997. *Journal of Geophysical Research*, *106*(A10), 21,077–21,082. https://doi.org/10.1029/2000JA000211
Wanliss, J. A., & Showalter, K. M. (2006). High-resolution global storm index: Dst versus SYM-H. *Journal of Geophysical Research*, *111*, A02202. https://doi.org/10.1029/2005JA011034
Wei, H.-L., Billings, S. A., Surjalal Sharma, A., Wing, S., Boynton, R. J., & Walker, S. N. (2011). Forecasting relativistic electron flux using dynamic multiple regression models. *Annales Geophysicae*, *29*, 415–2420. https://doi.org/10.5194/angeo-29-415-2011
Weimer, D. R., & King, J. H. (2008). Improved calculations of interplanetary magnetic field phase front angles and propagation time delays. *Journal of Geophysical Research*, *113*, A01105. https://doi.org/10.1029/2007JA012452
Weimer, D. R., Ober, D. M., Maynard, N. C., Burke, W. J., Collier, M. R., McComas, D. J., et al. (2002). Variable time delays in the propagation of the interplanetary magnetic field. *Journal of Geophysical Research*, *107*(A8), 1210. https://doi.org/10.1029/2001JA009102
Weimer, D. R., Ober, D. M., Maynard, N. C., Collier, M. R., McComas, D. J., Ness, N. F., et al. (2003). Predicting interplanetary magnetic field (IMF) propagation delay times using the minimum variance technique. *Journal of Geophysical Research*, *108*(A1), 1026. https://doi.org/10.1029/2002JA009405
Welling, D. T., Anderson, B. J., Crowley, G., Pulkkinen, A. A., & Rastätter, L. (2017). Exploring predictive performance: A reanalysis of the geospace model transition challenge. *Space Weather*, *15*, 192–203. https://doi.org/10.1002/2016SW001505
Welling, D. T., Barakat, A. R., Eccles, J. V., Schunk, R. W., & Chappell, C. R. (2016). Coupling the generalized polar wind model to global magnetohydrodynamics. In C. R. Chappell, R. W. Schunk, P. M. Banks, J. L. Burch, & R. M. Thorne (Eds.), *Magnetosphere-Ionosphere Coupling in the Solar System* (pp. 179–194). Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9781119066880.ch14
Welling, D. T., Jordanova, V. K., Glocer, A., Toth, G., Liemohn, M. W., & Weimer, D. R. (2015). The two-way relationship between ionospheric outflow and the ring current. *Journal of Geophysical Research: Space Physics*, *120*, 4338–4353. https://doi.org/10.1002/2015JA021231
Welling, D. T., & Liemohn, M. W. (2014). Outflow in global magnetohydrodynamics as a function of a passive inner boundary source. *Journal of Geophysical Research: Space Physics*, *119*, 2691–2705. https://doi.org/10.1002/2013JA019374
Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, *International Geophysics Series* (Vol. 100). Amsterdam, The Netherlands: Elsevier.
Wing, S., Johnson, J. R., Jen, J., Meng, C.-I., Sibeck, D. G., Bechtold, K., et al. (2005). Kp forecast models. *Journal of Geophysical Research*, *110*, A04203. https://doi.org/10.1029/2004JA010500
Yu, Y., Ridley, A. J., Welling, D. T., & Tóth, G. (2010). Including gap region field-aligned currents and magnetospheric currents in the MHD calculation of ground-based magnetic field perturbations. *Journal of Geophysical Research*, *115*, A08207. https://doi.org/10.1029/2009JA014869
Zhang, X. Y., & Moldwin, M. B. (2015). Probabilistic forecasting analysis of geomagnetic indices for southward IMF events. *Space Weather*, *13*, 130–140. https://doi.org/10.1002/2014SW001113
Zheng, Y., Macneice, P., Odstrcil, D., Mays, M. L., Rastaetter, L., Pulkkinen, A., et al. (2013). Forecasting propagation and evolution of CMEs in an operational setting: What has been learned. *Space Weather*, *11*, 557–574. https://doi.org/10.1002/swe.20096
Zhu, L., Schunk, R., Scherliess, L., & Eccles, V. (2012). Importance of data assimilation technique in defining the model drivers for the space weather specification of the high-latitude ionosphere. *Radio Science*, *47*, RS0L24. https://doi.org/10.1029/2011RS004936