Global Magnetohydrodynamic Simulations: Performance Quantification of Magnetopause Distances and Convection Potential Prediction

Agnit Mukhopadhyay[1], Xianzhe Jia[1], Daniel Welling[2], and Michael Liemohn[1]

[1]University of Michigan
[2]University of Michigan,University of Texas at Arlington

November 25, 2022

Abstract

The performance of three global magnetohydrodynamic (MHD) models in estimating the Earth's magnetopause location and ionospheric cross polar cap potential (CPCP) have been presented. Using the Community Coordinated Modeling Center's Run-on-Request system and extensive database on results of various magnetospheric scenarios simulated for a variety of solar weather patterns, the aforementioned model predictions have been compared with magnetopause standoff distance estimations obtained from six empirical models, and with cross polar cap potential estimations obtained from the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) Model and the Super Dual Auroral Radar Network (SuperDARN) observations. We have considered a range of events spanning different space weather activity to analyze the performance of these models. Using a fit performance metric analysis for each event, the models' reproducibility of magnetopause standoff distances and CPCP against empirically-predicted observations were quantified, and salient features that govern the performance characteristics of the modeled magnetospheric and ionospheric quantities were identified. Results indicate mixed outcomes for different models during different events, with almost all models underperforming during the extreme-most events. The quantification also indicates a tendency to underpredict magnetopause distances in the absence of an inner magnetospheric model, and an inclination toward over predicting CPCP values under general conditions.

Global Magnetohydrodynamic Simulations: Performance Quantification of Magnetopause Distances and Convection Potential Prediction

Agnit Mukhopadhyay[1*], Xianzhe Jia[1], Daniel T. Welling[2], Michael W. Liemohn[1]

[1]University of Michigan, United States, [2]University of Texas at Arlington, United States

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

We use the CRediT (Contributor Roles Taxonomy) categories (Brand et al., 2015) for providing the following contribution description. AM led the conceptualization, designed the methodology, conducted the investigation, performed data visualization and formal analysis, and wrote the original draft. XJ provided resources and supervised the initial conceptualization and methodology design. DW and ML assisted in conceptualization and formal analysis, provided the resources, funding acquisition, supervision, and aided in project administration. All authors have contributed toward the revision and editing of the manuscript.

Keywords

Ionospheric potential, magnetopause standoff distance, global MHD modeling, Space weather, Model validation and Analysis, NASA CCMC

Abstract

Word count: 199

The performance of three global magnetohydrodynamic (MHD) models in estimating the Earth's magnetopause location and ionospheric cross polar cap potential (CPCP) have been presented. Using the Community Coordinated Modeling Center's Run-on-Request system and extensive database on results of various magnetospheric scenarios simulated for a variety of solar weather patterns, the aforementioned model predictions have been compared with magnetopause standoff distance estimations obtained from six empirical models, and with cross polar cap potential estimations obtained from the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) Model and the Super Dual Auroral Radar Network (SuperDARN) observations. We have considered a range of events spanning different space weather activity to analyze the performance of these models. Using a fit performance metric analysis for each event, the models' reproducibility of magnetopause standoff distances and CPCP against empirically-predicted observations were quantified, and salient features that govern the performance characteristics of the modeled magnetospheric and ionospheric quantities were identified. Results indicate mixed outcomes for different models during different events, with almost all models underperforming during the extreme-most events. The quantification also indicates a tendency to underpredict magnetopause distances in the absence of an inner magnetospheric model, and an inclination toward over predicting CPCP values under general conditions.

Contribution to the field

This article quantifies the performance of three global MHD models' estimation of magnetopause standoff distances (MPSD) and ionospheric cross polar cap potential (CPCP). Both quantities are important in understanding the instantaneous state of the near-Earth space environment. We have considered a range of events spanning different space weather activity to analyze the performance of these models. The modeling results have been compared against several empirical magnetopause models, and potential measurements from SuperDARN and AMIE. Our study finds that while models are able to reasonably estimate MPSD when compared against empirical values, the CPCP is overpredicted in most cases.

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: No human studies are presented in this manuscript.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Global Magnetohydrodynamic Simulations: Performance Quantification of Magnetopause Distances and Convection Potential Predictions

Agnit Mukhopadhyay [1], **Xianzhe Jia** [1], **Daniel Welling** [2] **and Michael Liemohn** [1]

[1]*Climate and Space Sciences and Engineering Department, University of Michigan, Ann Arbor, MI, USA*

[2]*Department of Physics, University of Texas at Arlington, TX, USA*

Correspondence*:
Agnit Mukhopadhyay
agnitm@umich.edu

2 ABSTRACT

3   The performance of three global magnetohydrodynamic (MHD) models in estimating the Earth's
4 magnetopause location and ionospheric cross polar cap potential (CPCP) have been presented.
5 Using the Community Coordinated Modeling Center's Run-on-Request system and extensive
6 database on results of various magnetospheric scenarios simulated for a variety of solar weather
7 patterns, the aforementioned model predictions have been compared with magnetopause standoff
8 distance estimations obtained from six empirical models, and with cross polar cap potential
9 estimations obtained from the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) Model
10 and the Super Dual Auroral Radar Network (SuperDARN) observations. We have considered a
11 range of events spanning different space weather activity to analyze the performance of these
12 models. Using a fit performance metric analysis for each event, the models' reproducibility of
13 magnetopause standoff distances and CPCP against empirically-predicted observations were
14 quantified, and salient features that govern the performance characteristics of the modeled
15 magnetospheric and ionospheric quantities were identified. Results indicate mixed outcomes
16 for different models during different events, with almost all models underperforming during the
17 extreme-most events. The quantification also indicates a tendency to underpredict magnetopause
18 distances in the absence of an inner magnetospheric model, and an inclination toward over
19 predicting CPCP values under general conditions.

20 Keywords: magnetopause standoff distance, polar cap potential, ionospheric potential, global MHD, metric, space weather, RMSE,
21 underprediction, overprediction.

1   INTRODUCTION

22 The global state of the terrestrial magnetosphere may be broadly characterized by two categories of physical
23 identifiers: (a) geomagnetic indices which indicate variations in the near-Earth space environment due to
24 activity (e.g. Dst, Sym-H, Kp, AE; Pulkkinen et al., 2011; Glocer et al., 2016; Liemohn et al., 2018), and
25 (b) physical quantities that help describe the morphology and energy balance in the magnetosphere (ground
26 magnetic perturbations dB/dt and ΔB, field aligned currents, polar cap potential; Rastätter et al., 2011;

27 Pulkkinen et al., 2013; Honkonen et al., 2013; Anderson et al., 2017; Welling et al., 2017). In the latter set,
28 the cross polar cap potential (CPCP) and magnetopause standoff distances (MPSD) are two widely used
29 physical quantities that simultaneously help define the structure and state of the magnetospheric system.
30 The MPSD, defined as the nearest subsolar point of the magnetopause to the Earth's surface (e.g. Fairfield,
31 1971; Elsen and Winglee, 1997; Gombosi, 1998), has been a predominant measure in studying compression
32 of the Earth's dayside magnetosphere (e.g. Welling et al., 2020), while providing an instantaneous value
33 of the energy imparted on the terrestrial magnetic system by the solar wind (e.g. Lin et al., 2010). The
34 CPCP, on the other hand, acts as an instantaneous indicator of the amount of energy flowing into the
35 Earth's magnetosphere-ionosphere system from the solar wind (e.g. Boyle et al., 1997; Burke et al., 1999;
36 Russell et al., 2001; Liemohn and Ridley, 2002; Ridley and Liemohn, 2002; Ridley, 2005; Ridley et al.,
37 2010), and is frequently used in conjunction with field aligned currents (FACs) to describe ionospheric
38 electrodynamics (e.g. Reiff et al., 1981; Siscoe et al., 2002a,b; Ridley et al., 2004; Khachikjan et al., 2008;
39 Mukhopadhyay et al., 2020). Observationally, these two quantities are difficult to measure globally, with
40 MPSD estimates largely depending on satellite crossings of the magnetopause over a distributed period of
41 time (e.g. Shue et al., 1997), and CPCP depending on incomplete global coverage of the hemisphere using
42 ground-based observations and/or in-situ measurements from space (e.g. Gao, 2012). These quantities are,
43 therefore, measured using physics-driven empirical (e.g. Shue et al., 1997; Petrinec and Russell, 1993;
44 Boyle et al., 1997) or assimilative techniques (e.g. Kihn and Ridley, 2005). Since most of these techniques
45 were created for different initial conditions (e.g. Lin et al., 2010; Gao, 2012), comparison of multiple such
46 models against first-principles-based global models or each other is a daunting task. This task is made
47 especially precarious when studying extreme events, as most of these techniques were not designed to
48 simulate extreme conditions (e.g. Welling et al., 2017; Mukhopadhyay et al., 2020).

49   Several empirical models have been developed to estimate the MPSD. Physically, the size and shape
50 of the magnetopause can be estimated based on the dynamic and static pressure of the solar wind (e.g.
51 Kivelson and Russell, 1995) along with sufficient knowledge of the interplanetary magnetic field. This is
52 the primary basis of these models that estimate MPSD by assuming a general shape of the magnetopause.
53 The most commonly used magnetopause models such as the Shue et al. (1997, 1998) models or the Petrinec
54 and Russell (1993, 1996) model use trigonometric functions and solar wind parameters to describe the
55 MPSD. Later models such as the Liu et al. (2015) model have attempted to include additional pressure and
56 magnetic field components of the solar wind using predicted values from first-principles-based models in
57 addition to satellite crossing data in order to improve on these empirical models. A performance analysis of
58 many such models was presented by Lin et al. (2010) to compare their model against a range of empirical
59 models dating back to 1993. More recently, Staples et al. (2020) conducted a thorough analysis of MPSD
60 model performance, especially during extreme driving.

61   In contrast to MPSD models, the CPCP which is defined as the difference between the maxima and minima
62 of the ionospheric potential (e.g. Boyle et al., 1997) is largely derived from instantaneous observations
63 of ionospheric and/or ground-based quantities. The four most commonly used techniques to estimate
64 the ionospheric CPCP are: (1) polar observations by the Defense Meteorological Satellite Program (e.g.
65 Hairston and Heelis, 1996), (2) the polar cap index (e.g. Troshichev et al., 1996), (3) measurements by the
66 Super Dual Auroral Radar Network (SuperDARN; e.g. Khachikjan et al., 2008), and (4) the Assimilative
67 Mapping of Ionospheric Electrodynamics (AMIE) technique (e.g. Ridley and Kihn, 2004). An extensive
68 comparison of the general features, advantages and limitations of these datasets could be found in the work
69 by Gao (2012).

With the advent of physics-driven space weather prediction over the last couple of decades, validation of global first-principles-based models has become a common exercise in the space science community to identify and improve on our physical understanding of the near-Earth system (e.g. Pulkkinen et al., 2011, 2013; Rastätter et al., 2011). Compared to other space weather indices and/or space-based plasma quantities, fewer studies have compared the performance of MPSD and CPCP values from global models until recently (Collado-Vega et al., 2019; Burleigh et al., 2019; Mukhopadhyay et al., 2018, 2019). This is partly because, contrary to space weather indices (e.g. Glocer et al., 2013) and most other space weather quantities like FACs (e.g. Anderson et al., 2017) or ΔB (e.g. Welling et al., 2017), both MPSD and CPCP are measured by multiple methods and datasets. This means that a metric analysis of these quantities modeled after the GEM Challenges, which compared globally-modeled results against singular observational datasets, will not yield meaningful results.

In this study, an attempt to quantitatively compare globally-simulated MPSD and CPCP against multiple observationally-derived datasets has been undertaken. Three global magnetospheric models – the Space Weather Modeling Framework (SWMF), the Lyon-Fedder-Mobarry (LFM) model and the Open General Geospace Circulation Model (OpenGGCM) have been simulated through the NASA Community Coordinated Modeling Center (CCMC) website for seven space weather events. The global results are compared against six empirical MPSD models and two CPCP datasets. The performance analysis conducted in Pulkkinen et al. (2011), Rastätter et al. (2011) and Honkonen et al. (2013), one of the few validation studies to have compared MPSD and CPCP against the Lin et al. (2010) model and SuperDARN respectively, were used as a basis to select events and construct a metric performance analysis. However, to better serve the primary aim of the study, a new metric, Exclusion Parameter in addition to modified versions of the Root-Mean-Square Error and Maximum Amplitude Ratio has been used to dissociate physics-driven deficiencies in each model that impact the prediction of MPSD and CPCP. Results indicate global models to be overpredicting CPCP, while reasonably estimating MPSD values.

2 METHODOLOGY

2.1 Global Models & Event Selections

Three global models have been compared in this study – (1) SWMF, (2) LFM model and (3) OpenGGCM. The SWMF is a true framework containing a number of physics-based models (Tóth et al., 2005, 2012) and is operationally used in space weather prediction (e.g. Cash et al., 2018). It employs the BATS-R-US model (Powell et al., 1999) to simulate the global magnetospheric domain using conservative MHD equations. BATS-R-US is dynamically coupled to an inner magnetospheric model like Rice Convection Model (Wolf et al., 1982) which provides realistic ring current pressure and density (De Zeeuw et al., 2004; Glocer et al., 2016; Welling et al., 2018). The global and inner magnetospheric components are connected to the Ridley Ionosphere Model (RIM) which solves for the ionospheric electrodynamics using a prescribed empirical conductance model (Ridley et al., 2004; Mukhopadhyay et al., 2020).

The LFM model (Lyon et al., 2004; Merkine et al., 2003; Merkin et al., 2005a,b) is another global model that is actively used throughout the space science community. The MHD component employs a 3D stretched spherical grid to solve for semi-conservative MHD equations in the magnetospheric domain, which is then coupled with a magnetosphere-ionosphere coupler/solver (MIX). MIX solves for the ionospheric electric potential using a semi-empirical auroral conductance module that is driven using MHD inputs (Fedder et al., 1995; Wiltberger et al., 2001). Although the model is capable of additional coupling to an inner magnetospheric module (Pembroke et al., 2012), this coupling is not yet fully available on the CCMC website, and, therefore, was not utilized in the simulations conducted for this study.

112 OpenGGCM (Raeder et al., 2001, 2008) employs a non-uniform static Cartesian grid to solve the semi-
113 conservative resistive MHD equations in the GSE coordinate system. It is coupled with the Coupled
114 Thermosphere-Ionosphere Model (CTIM; e.g. Connor et al., 2016) to solve for the ionospheric potential
115 using both first-principle based and empirical methods. OpenGGCM provides auroral precipitation and
116 ionospheric FACs to CTIM, and receives the potential as an inner boundary condition. In spite of its
117 capability (Cramer et al., 2017), like LFM, there is no coupled inner magnetospheric model for OpenGGCM
118 available through the CCMC website, and therefore only OpenGGCM with coupled CTIM was used in this
119 study.

120 Seven geospace events, listed in Table 1, were chosen for the study. The selected events vary in strength
121 and magnetospheric structure as indicated by the minimum Dst and maximum AE reached during the
122 course of each event. Each event has been studied at least once in previous work. (Pulkkinen et al., 2011;
123 Honkonen et al., 2013; Miyoshi et al., 2006; Yermolaev et al., 2008). All global models have been executed
124 through the CCMC website (http://ccmc.gsfc.nasa.gov/) and receive as input the solar wind value at L1.
125 The ionospheric CPCP of the MHD models, made available as DPhi on the CCMC website, was used. The
126 features and settings of the global models were kept as similar to each other as possible. All models were
127 run with solar wind parameters provided by ACE and/or WIND, depending on availability. The simulation
128 results have been listed in the dataset provided with this manuscript, and have been made available through
129 the CCMC website using the CCMC-assigned run names.

2.2   Dataset for Data-Model Comparison

2.2.1   Magnetopause Standoff Distance Models

131 All magnetopause models used in this study have been listed in Table 2 along with a summary of their
132 fitting details with the solar wind. In this study, the results of Lin et al. (2010) were primarily used to select
133 the list of empirical models. In order to better evaluate MPSD models, Lin et al. (2010) used the standard
134 deviation $\sigma(d)$ to compare their model's performance with existing models against 246 satellite crossings
135 of the magnetopause with 5 min average solar wind parameters (see Table 10 in Lin et al., 2010). The
136 present study has included only those empirical models that predicted with a standard deviation lesser than
137 $\sim$1. In addition to the above, a later model developed by Liu et al. (2015) has also been used.

2.2.2   Cross Polar Cap Potential Models

140 Observations from SuperDARN and assimilated results from AMIE have been used to derive CPCP for
141 this study. SuperDARN is a network of radars that measures line-of-sight ionospheric convection velocities
142 with a ground-based network of radars and then infers functional forms of the electrostatic potential, as a
143 function of the colatitude and longitude (Ruohoniemi and Baker, 1998). For more detail on SuperDARN's
144 estimation technique of the CPCP, please refer to Khachikjan et al. (2008). AMIE assimilates many types
145 of data from both ground-based and space-based instruments and produces estimates of several ionospheric
146 parameters including the potential in the polar cap (Richmond and Kamide, 1988). In the version used in
147 this study (Kihn and Ridley, 2005), only ground magnetometer data have been used to predict the potential.

2.3   Performance Metrics

149 To undertake this comparative analysis, we have used the following three performance metrics: (1)
150 Root-Mean-Square Error (RMSE), (2) Maximum Amplitude Ratio (MAR), and (3) Exclusion Parameter
151 (EP). RMSE and MAR have been defined similarly to the metrics defined in Pulkkinen et al. (2011) and
152 Honkonen et al. (2013), in order to quantify the error in the simulated results. The metric EP has been
153 introduced specifically for this study in order to better quantify model-model comparisons. In the following,
154 results from the empirical magnetopause models and ionospheric results from AMIE and SuperDARN have

155 been interchangeably termed *predicted observations* or simply observations, to distinguish from results
156 from the global models.

157 RMSE is a popular fit metric used to quantify the difference between predictions and observations, with a
158 value of 0 indicating perfect performance. RMSE is defined as

$$RMSE = \sqrt{< (x_{i,mod} - x_{i,obs})^2 >} \quad \text{where } i = 1, 2, 3, ..., N \tag{1}$$

159 where x_{obs} and x_{mod} are the observed and the modeled results, respectively, $< ... >$ indicates the arithmetic
160 mean taken over i ranging over N time steps. Throughout this work i corresponds to the time series over
161 individual events, with N indicating the total number of time steps in a given event(s). Because RMSE
162 takes the square of the numbers involved, the values cannot be negative.

163 The second metric, MAR is defined as the ratio of the maximum amplitudes:

$$MAR = \frac{\max(|x_{i,mod}|)}{\max(|x_{i,obs}|)} \quad \text{where } i = 1, 2, 3, ..., N \tag{2}$$

164 where i, x_{obs} and x_{mod} stand for the same variables as in Equation 1. Clearly, $MAR = 1$ indicates
165 perfect model performance, while $MAR > 1$ and $MAR < 1$ indicates over- and underestimation. This is
166 especially useful in analyzing quantities like MPSD, where it is critical to understand whether the peak
167 value of globally-modeled MPSD is overpredicted or underpredicted when compared against empirically-
168 modeled MPSD which provides useful insight regarding the physical morphology of the magnetosphere,
169 especially during storm-time magnetospheric compression.

170 The third metric EP has been used to quantify times when simulated results lay outside the range
171 of observationally-derived estimates (including their standard deviations), and if during such times the
172 simulated results overestimated or underestimated the values. This is an important aspect to study as this
173 investigation is comparing modeled results against multiple observational and modeled datasets, and it
174 is highly unlikely that the observationally-derived estimates will match with each other. Any and every
175 prediction of the MHD-modeled data that is "excluded" from the observational range (outside the range of
176 observed values) has been characterized as an incorrect prediction, and therefore counted as an exclusion.
177 Mathematically, this could be defined as

$$EP_{i,total} = \begin{cases} 1, & \text{if} \quad x_{i,mod} \nsubseteq x_{i,obs}(\max, \min) \pm \sigma_{i,obs} \\ 0, & \text{if} \quad \text{otherwise.} \end{cases} \tag{3}$$

178 Here, i, x_{obs} and x_{mod} are the same as the previous equations, while σ_{obs} is the standard deviation of the
179 observed data, and (max, min) signify the maxima and minima of observed values at timestep i. Using the
180 above relation, EP identifies the number of times when the model is outside the set limits of the observed
181 values, and measures if the exclusivity is due to underprediction or overprediction of values at each time
182 step using the following relation:

$$\begin{aligned} \textbf{Underprediction} : EP_{i,under} &= \begin{cases} 1, & \text{if} \quad x_{i,mod} < x_{i,obs}(\min) - \sigma_{i,obs} \\ 0, & \text{if} \quad \text{otherwise.} \end{cases} \\ \textbf{Overprediction} : EP_{i,over} &= \begin{cases} 1, & \text{if} \quad x_{i,mod} > x_{i,obs}(\max) + \sigma_{i,obs} \\ 0, & \text{if} \quad \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

At the end of calculations, the total number of "excluded" time steps as a fraction of the total number of time steps defines the total EP underprediction and overprediction as a percentage value, such that the addition of the total underpredicted and overpredicted fractions results in the EP:

$$EP_{X,event} = \frac{\sum_{i=1}^{N} EP_{i,X}}{N}, \text{where} \quad \begin{matrix} X = total, underpredicted, overpredicted \\ i = 1, 2, 3, ..., N \end{matrix} \tag{5}$$

where EP_{event} is the total EP as a fraction of the total number of time steps, N. Note that the under- and over-prediction percentages are as a fraction of the total event time and not of the total wrongly predicted times. For example, a model with an EP value of 50% with a total under-prediction percentage of 10% for a given event indicates that the model results lie outside the observation thresholds 50% of all times during the event but under-predict only 10% of the total time, further signifying that 40% of the total time the model results are over-predicted. This parameter was specifically introduced to understand variations in the both the MP standoff distance and CPCP values, as the observations/empirical-derived quantities themselves vary at a given time step. Further discussion about this parameter's usage is described in Sections 3 and 4.

3 RESULTS

Figure 1 displays a composite image of the performance quantification of model-predicted magnetopause standoff distances against predicted observations using the empirical models. In part (a), a time series comparison of the magnetopause distance for the August 31, 2001 event has been shown. Results from the global models displayed using the solid lines are plotted against a grey band of values encompassing the individual time-series of all 6 empirical magnetopause models. The black solid line passing through the middle of the grey band is the median value of the empirically modeled results. In part (b), the aggregate RMSE (top subplot i), MAR (middle subplot ii) and EP (bottom subplot iii) have been computed for each event. In order to compute each metric, the time series data simulated by the global models were compared against the median value of the observationally-derived estimates. LFM magnetopause distances exhibit the lowest RMSE for each event, with 6 out of 7 events having a RMSE value 1 R_E. OpenGGCM has the highest RMSE values with 5 out the 7 events have RMSE values greater than 1 R_E. SWMF follows LFM results closely for all events except for the December 2006 event (Event 7) where it exhibits the highest RMSE value among all the global models. The median RMSE values across all events are plotted in light-dashed lines – LFM has the lowest aggregate RMSE at 0.5 R_E, followed closely by SWMF with a median RMSE value of 0.76 R_E and by OpenGGCM with a median RMSE value of 2.01 R_E. In part (b-ii), all models exhibit MAR values deviated from unity by less than 0.2 except during the Halloween Storm of 2003 (Event 4) and February 2004 event (Event 6). LFM performs reasonably well for all events, exhibiting a median MAR value of 1.15 R_E. While SWMF has the least median MAR value of 1.03 R_E, it underperforms during 3 of the 7 events resulting in the highest MAR values for Events 1, 2 and 4, overpredicting by a factor of >1.4 times the observed values during the Halloween Storm of 2003 (Event 4). OpenGGCM performs reasonably for 5 of the 7 events, with the model exhibiting significant deviation from unity during the Halloween Storm of 2003 (Event 4) and the December 2006 Storm (Event 7). In part (b-iii), the magnetopause standoff distance EPs for the three global models are compared. Both LFM and SWMF exhibit EP values less than 50% for almost all events resulting in an aggregate EP value of 36% and 42% respectively, with the only exception being SWMF's performance during Event 7. By comparison, OpenGGCM has a high EP value for almost all events with the model showing a median EP value of ~78%. The EP values are model-wise re-plotted in part (c) of the figure, but the area under the curve is coloured by the proportion of underprediction and overprediction. Since underprediction and

overprediction of the EP is calculated as a fraction of the total time series, the total EP for any given model could be defined as the addition of the underpredicted fraction and the overpredicted fraction. As shown in part (c-i), SWMF mostly overpredicts the magnetopause distance during all events except Event 7. It also has a significant underprediction fraction during Events 4, 5 and 6, which along with Event 7 correspond with some of the strongest events being studied in this report. In contrast to SWMF results, both LFM and OpenGGCM predominantly underpredict during almost all events when outside of empirically-predicted range of values. The only exception to this is OpenGGCM's EP values during the Halloween Storm of 2003 where the overprediction fraction are greater than the underprediction.

Figure 2 describes the comparison of CPCP values estimated by global models and compared against AMIE and SuperDARN measurements. A similar format to Figure 1 is followed for consistency. In part (a), a time series comparison of the CPCP for the December 14, 2006 event has been shown comparing MHD-modelled results against the band of values observed by SuperDARN and predicted by AMIE. In part (b-i), while the aggregate RMSE values for each model are within 100 kV, event-wise performance varies – SWMF exhibits the lowest median RMSE value of 24 kV, with the RMSE value being ¡ 50 kV for all events except Event 4. LFM follows a similar pattern as SWMF, but displays comparatively higher RMSE values for Events 6 and 7. OpenGGCM exhibits RMSE values greater than 100 kV for Events 2, 4 and 6. The simulations of the Halloween Storm of 2003 (Event 4) lead to the highest errors for CPCP. In part (b-ii), the MAR values of all models are much higher when compared to magneopause MAR values. All three models follow a similar trend for all events, except OpenGGCM during Event 2 and 6 when it exhibits a MAR value greater than 4 times the observed median values for those events. LFM exhibits a median MAR value of 2.05 while SWMF has the closest MAR value to unity of 0.995. In part (b-iii), all models exhibit an EP value >50% for all events except Event 4 and 7. OpenGGCM has the highest median EP value at 98.7%, with 4 out of 7 events being 100% out of range. LFM shows a median EP of 78.6%, while SWMF exhibits the lowest median EP value of 72%. The EP values replotted in part (c) show that LFM (part ii) and OpenGGCM (part iii) largely overpredict the CPCP when outside the range of observed values. While SWMF largely underpredicts the CPCP during Events 1, 2, 3 and 4, CPCP during the remaining events was mainly overpredicted.

4 DISCUSSION

Because modeled MPSD and CPCP were compared against multiple datasets, the lone usage of error metrics like RMSE is not enough to meaningfully rank model performance (Liemohn et al., 2020) as has often been done before (e.g. Pulkkinen et al., 2011). Because there is no single right answer, a significant aim of this study has been to develop innovative metrics to better quantify the performance of global models against multiple, divergent observationally-derived estimates. For example, CPCP values from SuperDARN and AMIE are at significant odds with each other during stronger events as evidenced by Figure 2(a). To counter this problem, MAR and EP are used which allow us to identify whether a global model overpredicts or underpredicts; this does not give us a quantitative error value, but is able to create a blanket range of values within which a modeled result could be considered reasonable. While the usage of better metrics (e.g. Haiducek et al., 2017; Morley et al., 2018) would be strongly considered for future studies involving CPCP and MPSD validation, the rudimentary metric analysis in this study has been used to understand the differences in each model's performance and discuss future directions toward improvements.

In the performance analysis of MPSD, the metrics indicate reasonable performance during weaker events. For instance, some of the lowest EP values are exhibited by all three models during Events 3 and 5, which have the lowest AE. LFM and OpenGGCM tend to underpredict the MP standoff distance, as indicated in part (c) of Figure 1. This is probably due to the absence of an inner magnetospheric module to provide

realistic ring current pressure values. SWMF, which uses RCM to provide a much stronger ring current input, tends to overpredict the MP standoff distance. This is in agreement with the study by Samsonov et al. (2016) which found that accounting for a realistic ring current in global MHD brings values closer to the empirical MP models. However, as shown in Staples et al. (2020), the validity of MP standoff distances as estimated by the empirical models during extreme events is questionable.

The CPCP metric analysis indicates that ionospheric potential predicted by the global models are greater than the expected value sometimes by more than a factor of 8. This tendency of global models to overpredict the CPCP could be driven by field aligned current generation in the global MHD domains and/or the ionospheric conductance value, as all models use a similar numerical framework to apply Ohm's Law (Goodman, 1995). Since FAC strength and pattern is an aspect of MHD grid resolution (Ridley et al., 2010; Wiltberger et al., 2017; Welling et al., 2019; Mukhopadhyay et al., 2020), incorrect estimation of the ionospheric conductance, especially in the polar (auroral) region, should play a significant role in the overprediction of CPCP. Since each global model uses different techniques to estimate this quantity (SWMF uses an empirical conductance model, while LFM and OpenGGCM use a semi-empirical physics-driven conductance system), it is challenging to suggest a universal solution; recent advancements addressing this issue through the ongoing Ionospheric Conductance Challenge was reported by Öztürk et al. (2020). In addition, significant deviations between AMIE and SuperDARN values, especially during the Halloween Storm (Event 4) and the December 2006 event (Event 7) indicate that a performance evaluation of CPCP measurement during extreme driving is necessary. Gao (2012) has discussed the disadvantages of using SuperDARN, which under-predicts, and AMIE, which over-predicts, leading to sharp deviation in the CPCP predictions. Future studies should consider using a tertiary source of data (like DMSP or PC Index) or a different quantity (e.g. hemispheric power index) to evaluate ionospheric performance.

5 CONCLUSIONS

The present study aimed at evaluating global models' prediction of MPSD and CPCP against multiple robust observationally-derived datasets. The study used well-documented space weather events simulated using three different global MHD models through the CCMC Run-on-Request feature. The MPSD from these model results were compared against empirical magnetopause models, while the predicted ionospheric polar cap potential values were compared against those obtained from SuperDARN and AMIE. Three performance metrics – RMSE, MAR and EP - were used to quantify the predictions. While the models performed reasonably well during times of relatively weak geomagnetic activity, it was found that extreme events lead to increased errors and a tendency to overpredict the ionospheric potential. While inclusion of a ring current model in a global simulation leads to lesser underprediction of the MPSD during extreme driving, the study does not find that such an approach necessarily leads to reduced errors. Furthermore, the use of empirical models to predict MPSD, and statistics-based datasets to predict CPCP, may lead to incorrect evaluations during extreme events. Future studies should consider applying improved metrics to further evaluate these parameters.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

We use the CRediT (Contributor Roles Taxonomy) categories (Brand et al., 2015) for providing the following contribution description. AM led the conceptualization, designed the methodology, conducted the

investigation, performed data visualization and formal analysis, and wrote the original draft. XJ provided resources and supervised the initial conceptualization and methodology design. DW and ML assisted in conceptualization and formal analysis, provided the resources, funding acquisition, supervision, and aided in project administration. All authors have contributed toward the revision and editing of the manuscript.

FUNDING

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in the University of Michigan's DeepBlue Repository (DOI:10.7302/ARG3-X036). All model result data, input files and observation data are available through the Community Coordinated Modeling Center website (http://ccmc.gsfc.nasa.gov/), and the Virtual Model Repository website (http://vmr.engin.umich.edu/).

REFERENCES

Anderson, B. J., Korth, H., Welling, D. T., Merkin, V. G., Wiltberger, M. J., Raeder, J., et al. (2017). Comparison of predictive estimates of high-latitude electrodynamics with observations of global-scale birkeland currents. *Space Weather* 15, 352–373. doi:10.1002/2016SW001529

Boyle, C. B., Reiff, P. H., and Hairston, M. R. (1997). Empirical polar cap potentials. *Journal of Geophysical Research: Space Physics* 102, 111–125. doi:10.1029/96JA01742

Brand, A., Allen, L., Altman, M., Hlava, M., and Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing* 28, 151–155. doi:10.1087/20150211

Burke, W. J., Weimer, D. R., and Maynard, N. C. (1999). Geoeffective interplanetary scale sizes derived from regression analysis of polar cap potentials. *Journal of Geophysical Research: Space Physics* 104, 9989–9994. doi:10.1029/1999JA900031

Burleigh, M., Mukhopadhyay, A., Welling, D., Ridley, A., and Liemohn, M. (2019). The Importance of Self-Consistent Conductivity in Coupling Magnetosphere-Ionosphere-Thermosphere Models. In *AGU Fall Meeting Abstracts*. vol. 2019, SA41B–3168

340 Cash, M., Singer, . H., Millward, G., Toth, G., Welling, D., and Balch, C. (2018). NOAA SWPC's
341   Operational Geospace Model Performance during Earth-Affecting Events. In *42nd COSPAR Scientific*
342   *Assembly*. vol. 42, D2.3–37–18

343 Collado-Vega, Y., Rastaetter, L., Khurana, S., Sibeck, D., and Anastopulos, M. (2019). Magnetospheric
344   Science - Global Simulations: Capabilities and Limitations. In *AGU Fall Meeting Abstracts*. vol. 2019,
345   SM23D–3229

346 Connor, H. K., Zesta, E., Fedrizzi, M., Shi, Y., Raeder, J., Codrescu, M. V., et al. (2016). Modeling
347   the ionosphere-thermosphere response to a geomagnetic storm using physics-based magnetospheric
348   energy input: OpenGGCM-CTIM results. *Journal of Space Weather and Space Climate* 6, A25.
349   doi:10.1051/swsc/2016019

350 Cramer, W. D., Raeder, J., Toffoletto, F. R., Gilson, M., and Hu, B. (2017). Plasma sheet injections into
351   the inner magnetosphere: Two-way coupled OpenGGCM-RCM model results. *Journal of Geophysical*
352   *Research: Space Physics* 122, 5077–5091. doi:10.1002/2017JA024104

353 De Zeeuw, D. L., Sazykin, S., Wolf, R. A., Gombosi, T. I., Ridley, A. J., and Tóth, G. (2004). Coupling of a
354   global MHD code and an inner magnetospheric model: Initial results. *Journal of Geophysical Research:*
355   *Space Physics* 109, 1–14. doi:10.1029/2003JA010366

356 Elsen, R. K. and Winglee, R. M. (1997). The average shape of the magnetopause: A comparison of
357   three-dimensional global mhd and empirical models. *Journal of Geophysical Research: Space Physics*
358   102, 4799–4819. doi:10.1029/96JA03518

359 Fairfield, D. H. (1971). Average and unusual locations of the earth's magnetopause and bow shock. *Journal*
360   *of Geophysical Research (1896-1977)* 76, 6700–6716. doi:10.1029/JA076i028p06700

361 Fedder, J. A., Slinker, S. P., Lyon, J. G., and Elphinstone, R. D. (1995). Global numerical simulation of
362   the growth phase and the expansion onset for a substorm observed by Viking. *Journal of Geophysical*
363   *Research* 100, 19083. doi:10.1029/95JA01524

364 Gao, Y. (2012). Comparing the cross polar cap potentials measured by superdarn and amie during saturation
365   intervals. *Journal of Geophysical Research: Space Physics* 117. doi:10.1029/2012JA017690

366 Glocer, A., Fok, M., Meng, X., Toth, G., Buzulukova, N., Chen, S., et al. (2013). CRCM + BATS-R-US
367   two-way coupling. *Journal of Geophysical Research: Space Physics* 118, 1635–1650. doi:10.1002/jgra.
368   50221

369 Glocer, A., Rastätter, L., Kuznetsova, M., Pulkkinen, A., Singer, H. J., Balch, C., et al. (2016). Community-
370   wide validation of geospace model local k-index predictions to support model transition to operations.
371   *Space Weather* 14, 469–480. doi:10.1002/2016SW001387

372 Gombosi, T. I. (1998). *Physics of the Space Environment*. Cambridge Atmospheric and Space Science
373   Series (Cambridge University Press). doi:10.1017/CBO9780511529474

374 Goodman, M. L. (1995). A three-dimensional, iterative mapping procedure for the implementation of an
375   ionosphere-magnetosphere anisotropic Ohm's law boundary condition in global magnetohydrodynamic
376   simulations. *Annales Geophysicae* 13, 843–853. doi:10.1007/s00585-995-0843-z

377 Haiducek, J. D., Welling, D. T., Ganushkina, N. Y., Morley, S. K., and Ozturk, D. S. (2017). SWMF Global
378   Magnetosphere Simulations of January 2005: Geomagnetic Indices and Cross-Polar Cap Potential. *Space*
379   *Weather* 15, 1567–1587. doi:10.1002/2017SW001695

380 Hairston, M. R. and Heelis, R. A. (1996). *Analysis of Ionospheric Parameters Based on DMSP SSIES Data*
381   *Using the DBASE4 and NADIA Programs*. Tech. rep., TEXAS UNIV AT DALLAS RICHARDSON
382   CENTER FOR SPACE SCIENCES

383 Honkonen, I., Rastätter, L., Grocott, A., Pulkkinen, A., Palmroth, M., Raeder, J., et al. (2013). On the
384   performance of global magnetohydrodynamic models in the Earth's magnetosphere. *Space Weather* 11,

385 313–326. doi:10.1002/swe.20055

386 Khachikjan, G. Y., Koustov, A. V., and Sofko, G. J. (2008). Dependence of superdarn cross polar cap

387 potential upon the solar wind electric field and magnetopause subsolar distance. *Journal of Geophysical*

388 *Research: Space Physics* 113. doi:10.1029/2008JA013107

389 Kihn, E. A. and Ridley, A. J. (2005). A statistical analysis of the assimilative mapping of ionospheric

390 electrodynamics auroral specification. *Journal of Geophysical Research: Space Physics* 110. doi:10.

391 1029/2003JA010371

392 Kivelson, M. and Russell, C. (1995). *Introduction to Space Physics*

393 Kuznetsov, S. N. and Suvorova, A. V. (1998). *An Empirical Model of the Magnetopause for Broad*

394 *Ranges of Solar Wind Pressure and BZ IMF* (Dordrecht: Springer Netherlands). 51–61. doi:10.1007/

395 978-94-011-5214-3_5

396 Liemohn, M. W., Ganushkina, N. Y., De Zeeuw, D. L., Rastaetter, L., Kuznetsova, M., Welling, D. T., et al.

397 (2018). Real-time swmf at ccmc: Assessing the dst output from continuous operational simulations.

398 *Space Weather* 16, 1583–1603. doi:10.1029/2018SW001953

399 Liemohn, M. W. and Ridley, A. J. (2002). Comment on "Nonlinear response of the polar ionosphere to

400 large values of the interplanetary electric field" by C. T. Russell et al. *Journal of Geophysical Research:*

401 *Space Physics* 107, SIA 13–1–SIA 13–4. doi:10.1029/2002JA009440

402 Liemohn, M. W., Shane, A. D., Azari, A. R., Petersen, A. K., Swiger, B. M., and Mukhopadhyay, A.

403 (2020). RMSE is not enough: guidelines to robust data-model comparisons for magnetospheric physics.

404 *Journal of Atmospheric and Solar-Terrestrial Physics (submitted, manuscript JASTP-D-20-00308)*

405 Lin, R. L., Zhang, X. X., Liu, S. Q., Wang, Y. L., and Gong, J. C. (2010). A three-dimensional asymmetric

406 magnetopause model. *Journal of Geophysical Research: Space Physics* 115. doi:10.1029/2009JA014235

407 Liu, Z.-Q., Lu, J. Y., Wang, C., Kabin, K., Zhao, J. S., Wang, M., et al. (2015). A three-dimensional high

408 mach number asymmetric magnetopause model from global mhd simulation. *Journal of Geophysical*

409 *Research: Space Physics* 120, 5645–5666. doi:10.1002/2014JA020961

410 Lyon, J. G., Fedder, J. A., and Mobarry, C. M. (2004). The Lyon–Fedder–Mobarry (LFM) global MHD

411 magnetospheric simulation code. *Journal of Atmospheric and Solar-Terrestrial Physics* 66, 1333–1350.

412 doi:https://doi.org/10.1016/j.jastp.2004.03.020

413 Merkin, V. G., Milikh, G., Papadopoulos, K., Lyon, J., Dimant, Y. S., Sharma, A. S., et al. (2005a). Effect

414 of anomalous electron heating on the transpolar potential in the LFM global MHD model. *Geophysical*

415 *Research Letters* 32, n/a–n/a. doi:10.1029/2005GL023315

416 Merkin, V. G., Sharma, A. S., Papadopoulos, K., Milikh, G., Lyon, J., and Goodrich, C. (2005b). Global

417 MHD simulations of the strongly driven magnetosphere: Modeling of the transpolar potential saturation.

418 *Journal of Geophysical Research: Space Physics* 110. doi:10.1029/2004JA010993

419 Merkine, V. G., Papadopoulos, K., Milikh, G., Sharma, A. S., Shao, X., Lyon, J., et al. (2003). Effects of

420 the solar wind electric field and ionospheric conductance on the cross polar cap potential: Results of

421 global MHD modeling. *Geophysical Research Letters* 30, n/a–n/a. doi:10.1029/2003GL017903

422 Miyoshi, Y. S., Jordanova, V. K., Morioka, A., Thomsen, M. F., Reeves, G. D., Evans, D. S., et al. (2006).

423 Observations and modeling of energetic electron dynamics during the October 2001 storm. *Journal of*

424 *Geophysical Research: Space Physics* 111. doi:10.1029/2005JA011351

425 Morley, S. K., Brito, T. V., and Welling, D. T. (2018). Measures of Model Performance Based On the Log

426 Accuracy Ratio. *Space Weather* 16, 69–88. doi:10.1002/2017SW001669

427 Mukhopadhyay, A., Welling, D., Burleigh, M., Ridley, A., Liemohn, M., Anderson, B., et al. (2019).

428 Conductance in the Aurora: Influence of Magnetospheric Contributors. In *AGU Fall Meeting Abstracts*.

429 vol. 2019, SA41B–3169. doi:doi.org/10.1002/essoar.10502150.1

430 Mukhopadhyay, A., Welling, D., Liemohn, M., Zou, S., and Ridley, A. (2018). Challenges in Space
431   Weather Prediction: Estimation of Auroral Conductance. In *AGU Fall Meeting Abstracts*. vol. 2018,
432   SA33B–3462

433 Mukhopadhyay, A., Welling, D. T., Liemohn, M. W., Ridley, A. J., Chakraborty, S., and Anderson, B. J.
434   (2020). Conductance Model for Extreme Events : Impact of Auroral Conductance on Space Weather
435   Forecasts. *Space Weather* , e2020SW002551doi:10.1029/2020SW002551

436 Öztürk, D. S., Garcia-Sage, K., and Connor, H. K. (2020). All Hands on Deck for Ionospheric Modeling.
437   *Eos* , 101doi:10.1029/2020EO144365

438 Pembroke, A., Toffoletto, F., Sazykin, S., Wiltberger, M., Lyon, J., Merkin, V., et al. (2012). Initial
439   results from a dynamic coupled magnetosphere-ionosphere-ring current model. *Journal of Geophysical
440   Research: Space Physics* 117. doi:10.1029/2011JA016979

441 Petrinec, S. M. and Russell, C. T. (1993). An empirical model of the size and shape of the near-Earth
442   magnetotail. *Geophysical Research Letters* 20, 2695–2698. doi:10.1029/93GL02847

443 Petrinec, S. M. and Russell, C. T. (1996). Near-Earth magnetotail shape and size as determined from
444   the magnetopause flaring angle. *Journal of Geophysical Research: Space Physics* 101, 137–152.
445   doi:10.1029/95JA02834

446 Powell, K. G., Roe, P. L., Linde, T. J., Gombosi, T. I., and Zeeuw, D. L. D. (1999). A Solution-Adaptive
447   Upwind Scheme for Ideal Magnetohydrodynamics. *Journal of Computational Physics* 154, 284–309.
448   doi:https://doi.org/10.1006/jcph.1999.6299

449 Pulkkinen, A., Kuznetsova, M., Ridley, A., Raeder, J., Vapirev, A., Weimer, D., et al. (2011). Geospace
450   Environment Modeling 2008-2009 Challenge: Ground magnetic field perturbations. *Space Weather* 9,
451   n/a–n/a. doi:10.1029/2010SW000600

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

455 Raeder, J., Larson, D., Li, W., Kepko, E. L., and Fuller-Rowell, T. (2008). OpenGGCM Simulations for
456   the THEMIS Mission. *Space Science Reviews* 141, 535–555. doi:10.1007/s11214-008-9421-5

457 Raeder, J., McPherron, R. L., Frank, L. A., Kokubun, S., Lu, G., Mukai, T., et al. (2001). Global
458   simulation of the Geospace Environment Modeling substorm challenge event. *Journal of Geophysical
459   Research-Space Physics* 106, 381–395. doi:10.1029/2000ja000605

460 Rastätter, L., Kuznetsova, M. M., Vapirev, A., Ridley, A., Wiltberger, M., Pulkkinen, A., et al. (2011).
461   Geospace environment modeling 2008–2009 challenge: Geosynchronous magnetic field. *Space Weather*
462   9. doi:10.1029/2010SW000617

463 Reiff, P. H., Spiro, R. W., and Hill, T. W. (1981). Dependence of polar cap potential drop
464   on interplanetary parameters. *Journal of Geophysical Research: Space Physics* 86, 7639–7648.
465   doi:10.1029/JA086iA09p07639

466 Richmond, A. D. and Kamide, Y. (1988). Mapping electrodynamic features of the high-latitude ionosphere
467   from localized observations - Technique. *Journal of Geophysical Research* 93, 5741–5759. doi:10.1029/
468   JA093iA06p05741

469 Ridley, A. J. (2005). A new formulation for the ionospheric cross polar cap potential including saturation
470   effects. *Annales Geophysicae* 23, 3533–3547. doi:10.5194/angeo-23-3533-2005

471 Ridley, A. J., Gombosi, T. I., and De Zeeuw, D. L. (2004). Ionospheric control of the magnetosphere:
472   conductance. *Annales Geophysicae* 22, 567–584. doi:10.5194/angeo-22-567-2004

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

475   doi:10.5194/angeo-28-1589-2010

476 Ridley, A. J. and Kihn, E. A. (2004). Polar cap index comparisons with amie cross polar cap potential,
477   electric field, and polar cap area. *Geophysical Research Letters* 31. doi:10.1029/2003GL019113

478 Ridley, A. J. and Liemohn, M. W. (2002). A model-derived storm time asymmetric ring current driven
479   electric field description. *Journal of Geophysical Research: Space Physics* 107, SMP 2–1–SMP 2–12.
480   doi:10.1029/2001JA000051

481 Ruohoniemi, J. M. and Baker, K. B. (1998). Large-scale imaging of high-latitude convection with Super
482   Dual Auroral Radar Network HF radar observations. *Journal of Geophysical Research: Space Physics*
483   103, 20797–20811. doi:10.1029/98JA01288

484 Russell, C. T., Luhmann, J. G., and Lu, G. (2001). Nonlinear response of the polar ionosphere to
485   large values of the interplanetary electric field. *Journal of Geophysical Research: Space Physics* 106,
486   18495–18504. doi:10.1029/2001JA900053

487 Samsonov, A. A., Gordeev, E., Tsyganenko, N. A., Šafránková, J., Němeček, Z., Šimůnek, J., et al. (2016).
488   Do we know the actual magnetopause position for typical solar wind conditions? *Journal of Geophysical*
489   *Research: Space Physics* 121, 6493–6508. doi:10.1002/2016JA022471

490 Shue, J.-H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., et al. (1997). A new functional
491   form to study the solar wind control of the magnetopause size and shape. *Journal of Geophysical*
492   *Research: Space Physics* 102, 9497–9511. doi:10.1029/97JA00196

493 Shue, J.-H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., et al. (1998). Magnetopause
494   location under extreme solar wind conditions. *Journal of Geophysical Research: Space Physics* 103,
495   17691–17700. doi:10.1029/98JA01103

496 Siscoe, G. L., Crooker, N. U., and Siebert, K. D. (2002a). Transpolar potential saturation: Roles of
497   region 1 current system and solar wind ram pressure. *Journal of Geophysical Research* 107, 1321.
498   doi:10.1029/2001JA009176

499 Siscoe, G. L., Erickson, G. M., Sonnerup, B. U. Ö., Maynard, N. C., Schoendorf, J. A., Siebert, K. D., et al.
500   (2002b). Hill model of transpolar potential saturation: Comparisons with MHD simulations. *Journal of*
501   *Geophysical Research* 107, 1075. doi:10.1029/2001JA000109

502 Staples, F. A., Rae, I. J., Forsyth, C., Smith, A. R. A., Murphy, K. R., Raymer, K. M., et al. (2020).
503   Do Statistical Models Capture the Dynamics of the Magnetopause During Sudden Magnetospheric
504   Compressions? *Journal of Geophysical Research: Space Physics* 125, e2019JA027289. doi:10.1029/
505   2019JA027289

506 Tóth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C. R., De Zeeuw, D. L., et al. (2005). Space
507   Weather Modeling Framework: A new tool for the space science community. *Journal of Geophysical*
508   *Research* 110, A12226. doi:10.1029/2005JA011126

509 Tóth, G., van der Holst, B., Sokolov, I. V., De Zeeuw, D. L., Gombosi, T. I., Fang, F., et al. (2012).
510   Adaptive numerical algorithms in space weather modeling. *Journal of Computational Physics* 231,
511   870–903. doi:10.1016/j.jcp.2011.02.006

512 Troshichev, O., Hayakawa, H., Matsuoka, A., Mukai, T., and Tsuruda, K. (1996). Cross polar cap diameter
513   and voltage as a function of PC index and interplanetary quantities. *Journal of Geophysical Research:*
514   *Space Physics* 101, 13429–13435. doi:https://doi.org/10.1029/95JA03672

515 Welling, D., Dimmock, A., Rosenqvist, L., Morley, S., and Yordanova, E. (2019). Resolving small scale
516   GIC effects: What is our capability? In *AGU Fall Meeting Abstracts*. vol. 2019, SH32B–05

517 Welling, D., Love, J., Rigler, E. J., Oliveira, D., and Komar, C. (2020). Numerical simulations of the
518   geospace response to the arrival of a perfect interplanetary coronal mass ejection. *accepted in Space*
519   *Weather (Preprint Available)* , 16doi:10.1002/essoar.10502106.1

520   Welling, D. T., Anderson, B. J., Crowley, G., Pulkkinen, A. A., and Rastätter, L. (2017). Exploring
521   predictive performance: A reanalysis of the geospace model transition challenge. *Space Weather* 15,
522   192–203. doi:10.1002/2016SW001505

523   Welling, D. T., Ngwira, C. M., Opgenoorth, H., Haiducek, J. D., Savani, N. P., Morley, S. K., et al. (2018).
524   Recommendations for next-generation ground magnetic perturbation validation. *Space Weather* 16,
525   1912–1920. doi:10.1029/2018SW002064

526   Wiltberger, M., Rigler, E. J., Merkin, V., and Lyon, J. G. (2017). Structure of High Latitude
527   Currents in Magnetosphere-Ionosphere Models. *Space Science Reviews* 206, 575–598. doi:10.1007/
528   s11214-016-0271-2

529   Wiltberger, M., Weigel, R. S., Lotko, W., and Fedder, J. A. (2001). Modeling seasonal variations of auroral
530   particle precipitation in a global-scale magnetosphere-ionosphere simulation. *Journal of Geophysical*
531   *Research-Space Physics* 114, 381–395. doi:10.1029/2008JA013108

532   Wolf, R. A., Harel, M., Spiro, R. W., Voigt, G.-H., Reiff, P. H., and Chen, C.-K. (1982). Computer
533   simulation of inner magnetospheric dynamics for the magnetic storm of July 29, 1977. *Journal of*
534   *Geophysical Research* 87, 5949. doi:10.1029/JA087iA08p05949

535   Yermolaev, Y., Zelenyi, L. M., Kuznetsov, V. D., Chertok, I. M., Panasyuk, M. I., Myagkova, I. N., et al.
536   (2008). Magnetic storm of November, 2004: Solar, interplanetary, and magnetospheric disturbances.
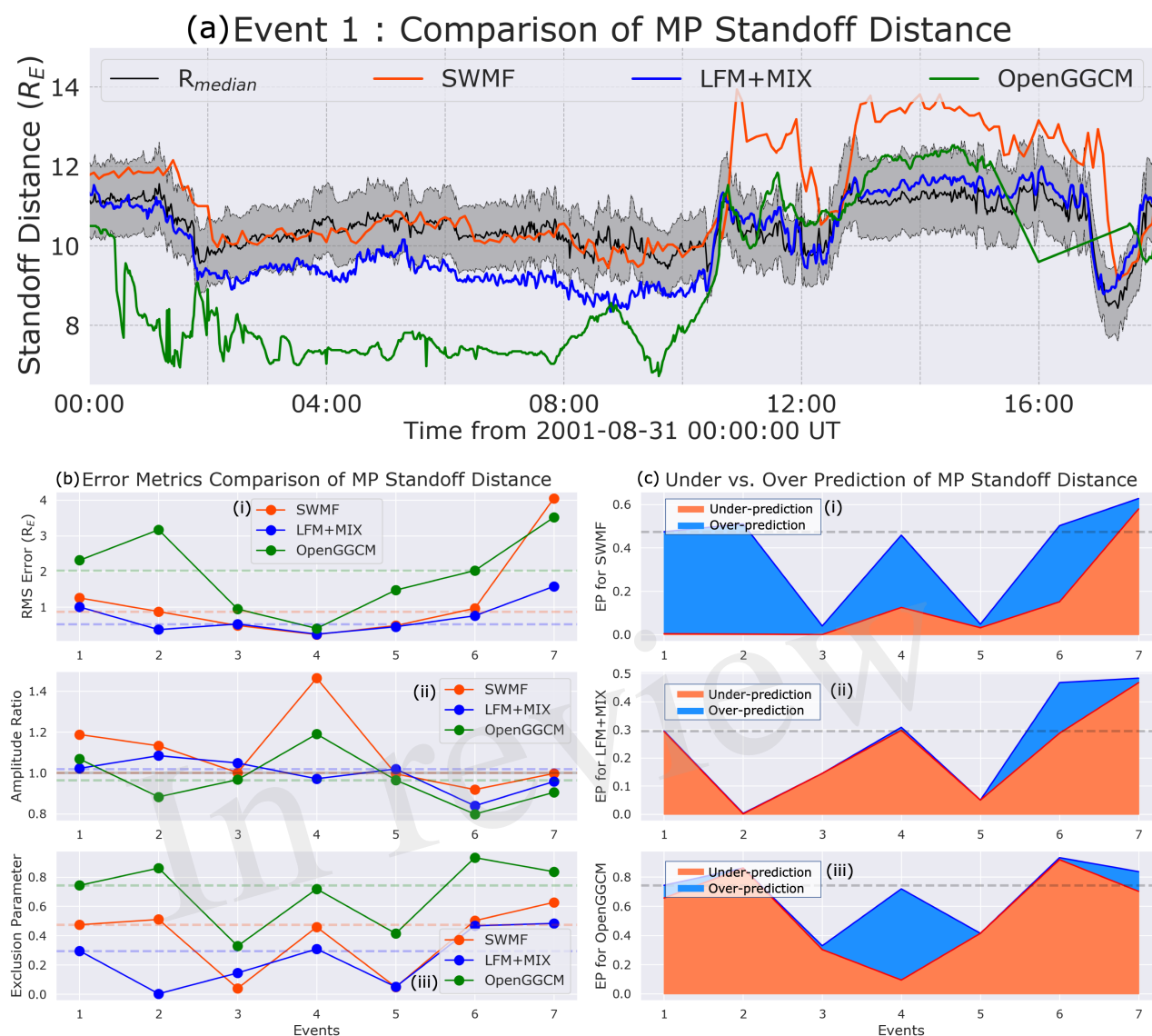537   *Journal of Atmospheric and Solar-Terrestrial Physics* 70, 334–341. doi:https://doi.org/10.1016/j.jastp.
538   2007.08.020

TABLES & FIGURES

Event #	Date and Time	Dst	AE
1	Aug 31, 2001 00:00 UT - Sept 1, 2001 00:00 UT	-40 nT	959 nT
2	Oct 5, 2001 00:00 UT - Oct 6, 2001 06:00 UT	-43 nT	837 nT
3	Sept 8, 2002 18:00 UT to 23:00 UT	-79 nT	428 nT
4	Oct 29, 2003 0600 UT to Oct 30, 2003 0600 UT	-353 nT	4056 nT
5	Nov 19, 2004 0000 UT to Nov 20, 2004 1200 UT	-40 nT	1146 nT
6	Feb 18, 2004 1400 UT to Feb 19, 2004 0000 UT	-23 nT	825 nT
7	Dec 14, 2006 1200 UT to Dec 16, 2006 0000 UT	-162 nT	2284 nT

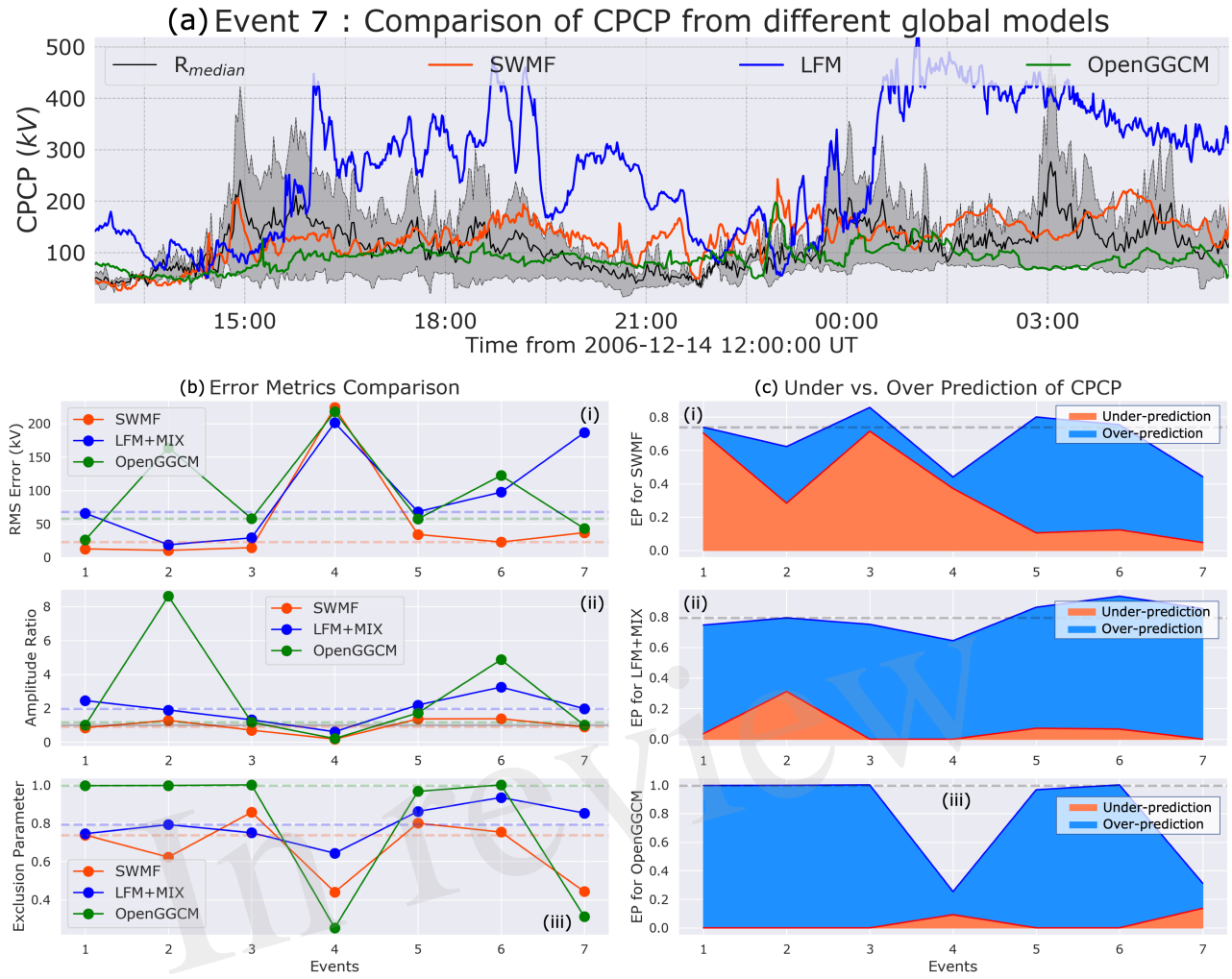**Table 1.** List of all geospace events studied in the present work.

Model	Specific Dependence	$\sigma(d)$
Petrinec and Russell (1996)	z-component of B-field, dynamic pressure	0.703
Shue et al. (1997)	z-component of B-field, dynamic pressure	0.791
Shue et al. (1998)	z-component of B-field, dynamic pressure	0.791
Kuznetsov and Suvorova (1998)	z-component of B-field, dynamic pressure, L-shell value	0.651
Lin et al. (2010)	z-component of B-field, dynamic pressure, magnetic pressure, L-shell value, polar angles	0.539
Liu et al. (2015)	z-component & y-component of B-field, dynamic pressure, magnetic pressure, L-shell value, polar angles	N/A

Table 2. Summary of the Empirical Models with a list of solar-wind dependencies required for their execution. The table also shows standard deviation value for a subsolar standoff distance lesser than 30 degrees that have been derived from Lin et al. (2010).

Figure 1. Analysis of the MHD-predicted magnetopause standoff distances against the empirical models – (a) An example time-transient comparison of the magnetopause distances during Event 1 from SWMF (in red), LFM+MIX (in blue) and OpenGGCM (in green) compared against the six empirical magnetopause models which have been shown here as a range of values demarcated by the grey band, with the black line in the center of the band being the median value. (b) Comparison of (i) RMS Error, (ii) Amplitude Ratio and (iii) Exclusion Parameter for the three models across all 7 events (in same colour scheme as Top plot). The dashed lines in the background signify the median value of these metrics across all events. (c) Comparison of the Underprediction and Overprediction score from the Exclusion parameter. The results have been separately presented for (i) SWMF, (ii) LFM+MIX and (iii) OpenGGCM, with orange signifying underprediction of values, and light-blue indicating overprediction. The overpredicted values are plotted atop the underpredicted values.
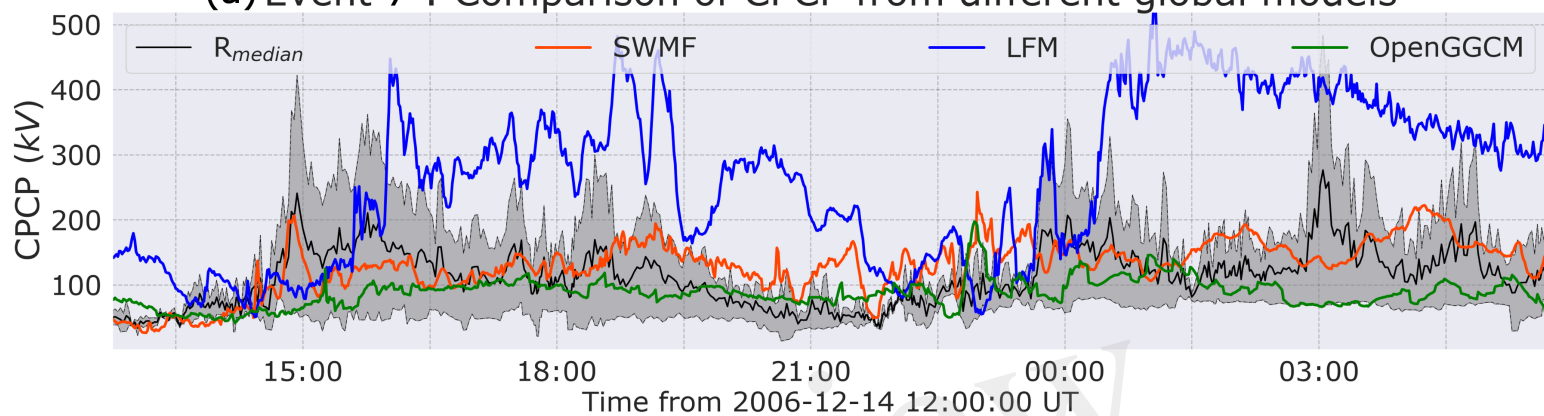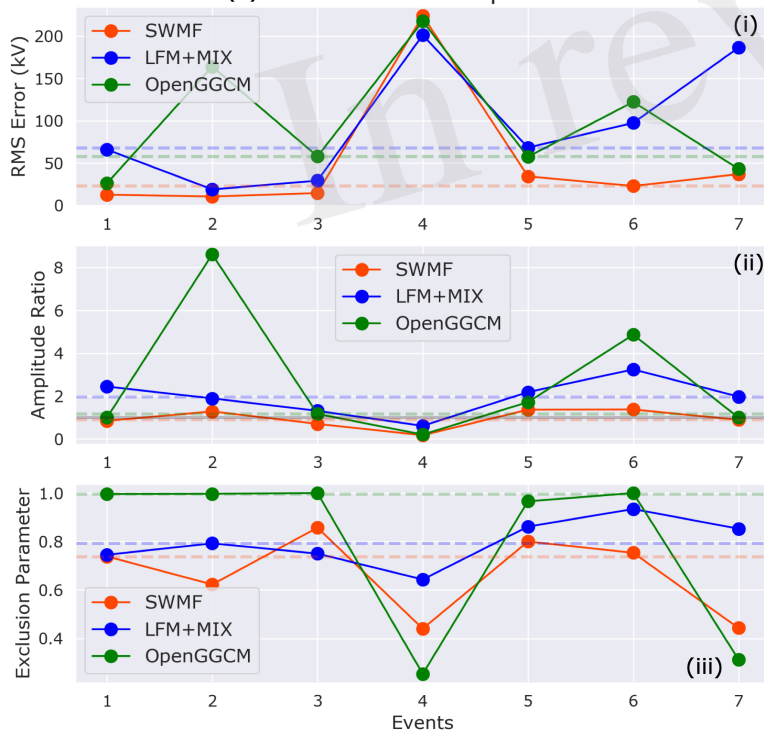
Figure 2. Analysis of the MHD-predicted cross polar cap potential (CPCP) against AMIE and SuperDARN estimates – (a) An example time-transient comparison of the CPCP during Event 8 from SWMF (in red), LFM+MIX (in blue) and OpenGGCM (in green) compared against the combined range of values between AMIE and SuperDARN estimates demarcated here by the grey band, with the black line in the center of the band being the median value. Formats for subplots (b) and (c) are similar to Figure 1(b) and (c).
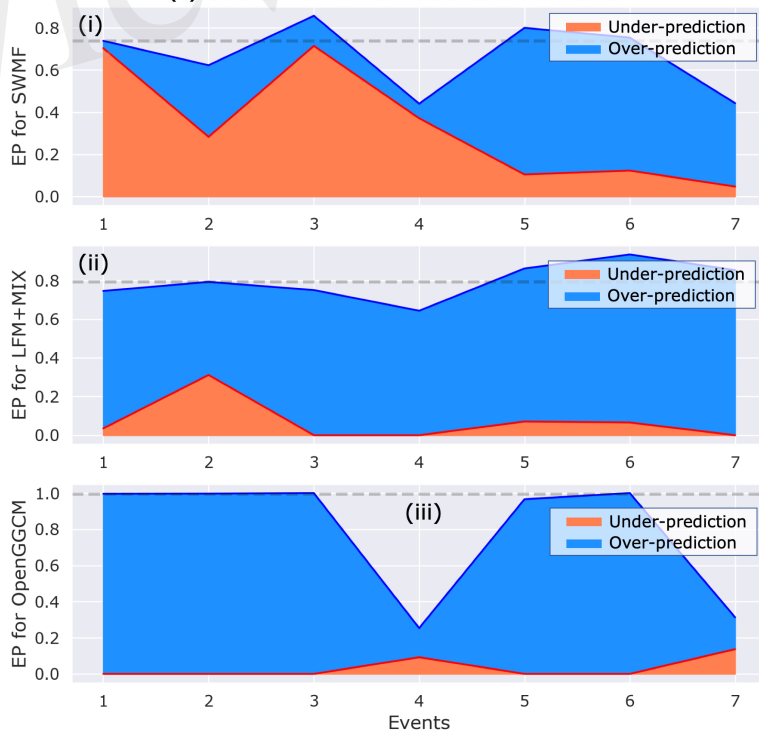
Figure 1.JPEG

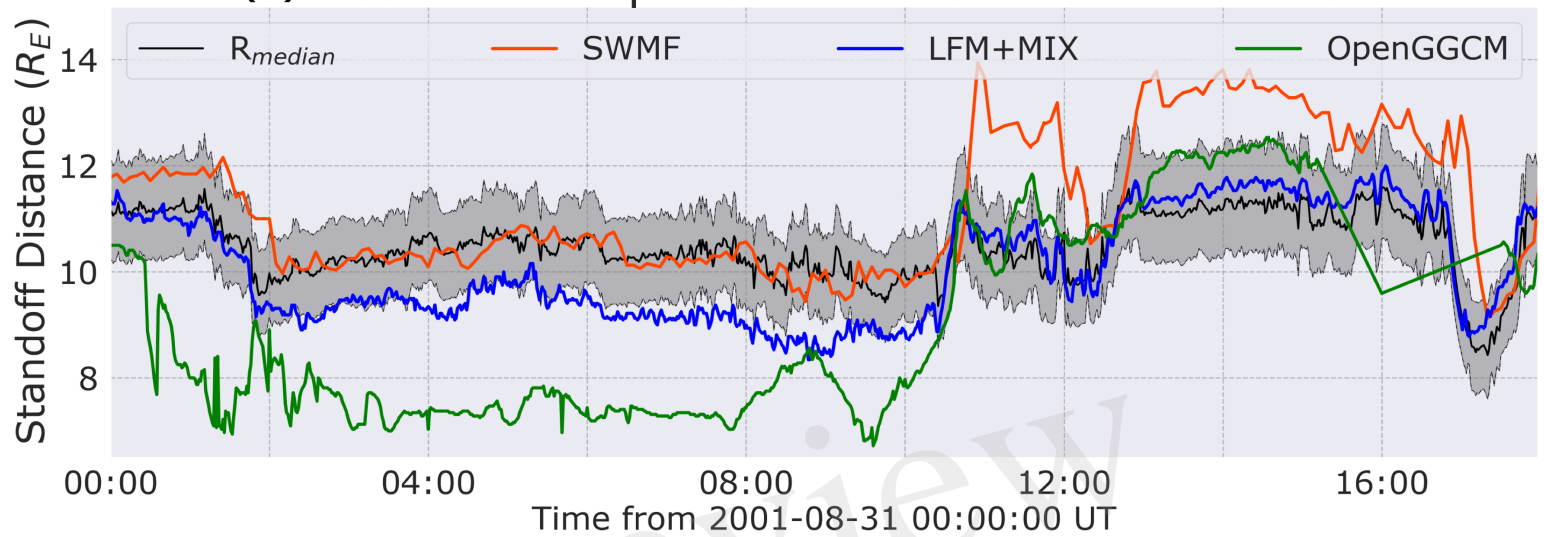(a) Event 7 : Comparison of CPCP from different global models
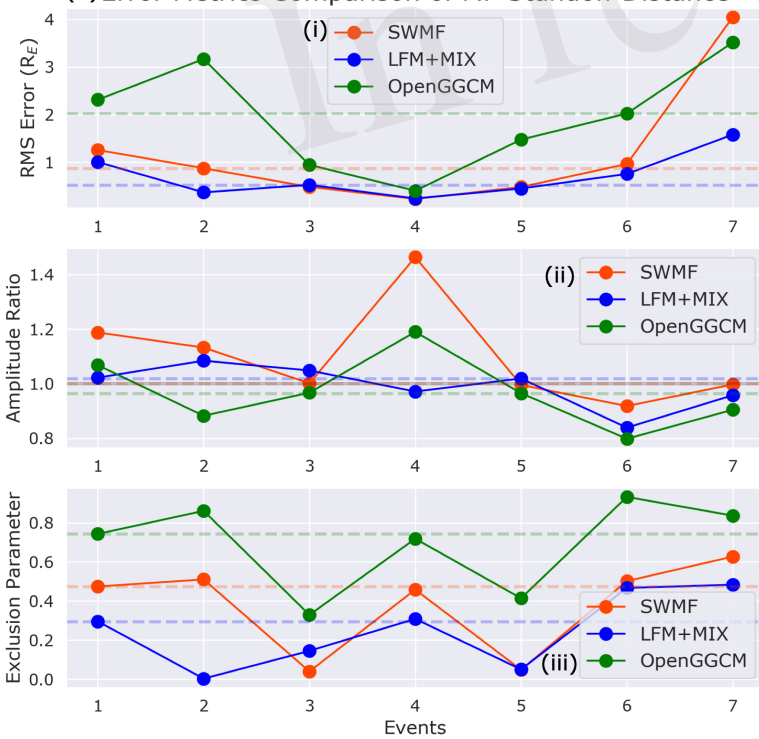
(b) Error Metrics Comparison

(c) Under vs. Over Prediction of CPCP

Figure 2.JPEG

(a)Event 1 : Comparison of MP Standoff Distance

(b)Error Metrics Comparison of MP Standoff Distance

(c) Under vs. Over Prediction of MP Standoff Distance