



## RESEARCH ARTICLE

10.1029/2023SW003428

### Key Points:

- We present a regression model of high-latitude ionospheric plasma convection driven by a small number of space weather variables
- The model could be a valuable tool for operational forecasting due to its simplicity
- The model reproduces its parent data set best around solar maximum, with a mean percentage of explained variance of 0.7 over 12 years

### Correspondence to:

M. M. Lam,  
mml@bas.ac.uk

### Citation:

Lam, M. M., Shore, R. M., Chisham, G., Freeman, M. P., Grocott, A., Walach, M.-T., & Orr, L. (2023). A model of high latitude ionospheric convection derived from SuperDARN EOF model data. *Space Weather*, 21, e2023SW003428. <https://doi.org/10.1029/2023SW003428>

Received 10 JAN 2023

Accepted 5 JUL 2023

### Author Contributions:

**Conceptualization:** M. P. Freeman  
**Data curation:** M. M. Lam  
**Formal analysis:** M. M. Lam, G. Chisham  
**Funding acquisition:** G. Chisham, M. P. Freeman, A. Grocott  
**Investigation:** M. M. Lam, R. M. Shore, G. Chisham, M. P. Freeman, A. Grocott, M.-T. Walach, L. Orr  
**Methodology:** R. M. Shore, G. Chisham, M. P. Freeman  
**Project Administration:** G. Chisham  
**Software:** M. M. Lam, R. M. Shore  
**Supervision:** G. Chisham  
**Validation:** M. M. Lam  
**Visualization:** M. M. Lam  
**Writing – original draft:** M. M. Lam  
**Writing – review & editing:** M. M. Lam, R. M. Shore, G. Chisham, M. P. Freeman, A. Grocott, M.-T. Walach, L. Orr

© 2023. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

# A Model of High Latitude Ionospheric Convection Derived From SuperDARN EOF Model Data

M. M. Lam<sup>1</sup> , R. M. Shore<sup>1</sup> , G. Chisham<sup>1</sup> , M. P. Freeman<sup>1</sup> , A. Grocott<sup>2</sup> , M.-T. Walach<sup>2</sup> , and L. Orr<sup>2</sup>

<sup>1</sup>British Antarctic Survey, Cambridge, UK, <sup>2</sup>Lancaster University, Lancaster, UK

**Abstract** Forecasting of the effects of thermospheric drag on satellites will be improved significantly with better modeling of space weather effects on the high-latitude ionosphere, in particular the Joule heating arising from electric field variability. We use a regression analysis to build a model of the ionospheric convection drift velocity which is driven by relatively few solar and solar wind variables. The model is developed using a solar cycle's worth (1997–2008 inclusive) of 5-min resolution Empirical Orthogonal Function (EOF) patterns derived from Super Dual Auroral Radar Network (SuperDARN) line-of-sight observations of the convection velocity across the high-latitude northern hemisphere ionosphere. At key stages of development of the model, we use the percentage of explained variance  $P$  to see how well the model reproduces the EOF data. The final model is driven by four variables: (a) the interplanetary magnetic field component  $B_y$ , (b) the solar wind coupling parameter  $\epsilon$ , (c) a trigonometric function of day-of-year, and (d) the monthly  $F_{10.7}$  index. The model can reproduce the EOF velocities with a characteristic  $P = 0.7$ . The model and EOF data compare best around the solar maximum of 2001.  $P$  is lower around solar minimum, due to occasional limitations in the geographical and temporal coverage of the SuperDARN measurements. This may indicate the need to modify our model around the minimum of the solar cycle. Our model has the potential to be used to forecast the ionospheric electric field using the real-time solar wind data available from spacecraft located upstream of the Earth.

**Plain Language Summary** Variations in space weather in the ionized region of the Earth's atmosphere (the ionosphere) can result in expansion of the atmosphere, increasing the atmospheric drag on objects, such as satellites, in the thermosphere. We aim to significantly improve the forecasting of the effects of atmospheric drag on satellites by more accurate modeling of space weather effects on the motion of ionized particles (plasma) in the ionosphere. We have developed a model of the variation in plasma motion using a few solar wind variables which are all now available in real time from satellites upstream of the Earth. The model was built using 5-min resolution observations of the ionospheric plasma motion from a 12-year interval, to capture effects on the solar cycle timescale. Our model is good at reproducing the original data set—if 0 indicates that there is no reproduction and 1 indicates exact reproduction, then our model scores 0.7. Data set reproduction is best around the maximum in the solar cycle and worst at solar minimum. This is mainly due to differences in the spatiotemporal data coverage between these times but possibly also the model's specification of the physical processes coupling the Sun to the Earth's ionosphere.

## 1. Introduction

Space debris is a major problem for low-Earth orbit (LEO) spacecraft, with the potential to cause significant socio-economic impacts (Lloyds, 2018; Undseth et al., 2020). There are over 36,000 objects larger than 10 cm, a million between 1 and 10 cm, and likely over a hundred million smaller objects (European Space Agency, 2022). Such debris present a significant risk of damage or destruction to operational satellites due to the chance of collision (e.g., Lewis, 2020). Moreover, as the population of LEO satellites and debris grows, the probability of collisions between them increases. Hypervelocity collisions create clouds of additional debris, further increasing the risk of subsequent collisions (Lewis et al., 2011). The debris density at some LEO altitudes has now passed the tipping point where the increase in debris due to collisions exceeds losses to the atmosphere, leading to a net growth of the space debris population in LEO even without new satellite launches (Lewis et al., 2011). Therefore, there is an increasingly important need to more accurately track spacecraft and space debris to enable timely and efficient planning of collision avoidance maneuvers (e.g., Lewis et al., 2012). Current orbit modeling and prediction are insufficiently accurate because of the poorly specified time-varying drag effect of the upper

atmosphere. The dominant unknown variable in orbital trajectory predictions of LEO objects is the density of the thermosphere, which exerts a time and location dependent drag.

Coupling of the solar wind to the magnetosphere results in energy being injected into the magnetosphere-ionosphere system. In the ionosphere, this is measurable as an electric field that is generated by plasma convection. This energy can be dissipated via Joule heating in the ionosphere-thermosphere, which has been estimated to dissipate over 50% of the total solar wind energy input to the Earth system (Østgaard et al., 2002). Joule heating is a key contributor to the power input into the thermosphere and thereby to the drag on satellites (e.g., Knipp et al., 2005), via upward expansion of the polar atmosphere and resultant significant density changes in the thermosphere. Accurate models and prediction of these high latitude processes are therefore vital to safeguard space assets. However, given that the thermospheric density can vary by ~80% diurnally and by ~250% during a geomagnetic storm, this is a major modeling challenge (Sutton et al., 2005).

Joule heating is specified by  $\mathbf{J} \cdot \mathbf{E}$  in the rest frame of the ionosphere, where  $\mathbf{J}$  and  $\mathbf{E}$  are the current density and electric field vectors (e.g., Vasyliūnas & Song, 2005). The model presented here will provide the high-latitude ionospheric convection ( $\mathbf{E} \times \mathbf{B}$  drift) velocity field and consequently the associated convection electric field  $\mathbf{E}$ , having specified the magnetic field  $\mathbf{B}$ . These can be used in fully coupled (neutral and ionized) models of the lower and upper atmosphere to specify the global thermospheric density at the lower end of LEO orbits. Such models include the Thermosphere-Ionosphere Electrodynamics General Circulation Model (GCM) (Qian et al., 2014) and the Advanced Ensemble electron density (Ne) Assimilation System (AENeAS), which is a data assimilation model derived from TIEGCM (Elvidge & Angling, 2019). The motivation for this study, and the production of the new model, was given by the UK's Space Weather Instrumentation, Measurement, Modeling and Risk—Thermosphere (SWIMMR-T) project, a multi-million-pound project during 2020–2024 to operationalize UK space weather modeling and forecasting.

Expertise in empirical modeling of the polar ionospheric electric field has been acquired over a period of more than four decades using a variety of data sets. Models have exploited both incoherent scatter radars (e.g., Foster, 1983; Zhang et al., 2007) and coherent scatter radars (e.g., Pettigrew et al., 2010; Ruohoniemi & Greenwald, 1996, 2005; Thomas & Shepherd, 2018), both low-altitude spacecraft (e.g., Hairston & Heelis, 1990; Heppner, 1977; Heppner & Maynard, 1987; V. O. Papitashvili & Rich, 2002; Weimer, 1995, 1996, 2001, 2005) and the high-altitude Cluster spacecraft (e.g., Förster et al., 2007), and ground-based magnetometer arrays (e.g., Ridley et al., 2000). To clarify, the high-latitude ionospheric plasma drift velocity field is typically referred to in this scientific literature as “convection.”

The basis for our model is 12 years of data from the Super Dual Auroral Radar Network (SuperDARN), which has been used to monitor mid to high-latitude ionospheric plasma velocities over the last few decades (Chisham et al., 2007; Greenwald et al., 1995; Nishitani et al., 2019). The interval of data, from 1997 to 2008 inclusive, starts just after the solar minimum of August 1996, and includes the solar maximum of November 2001 and the solar minimum of December 2008. In building our model, we first note the mostly two-cell morphology of the climatology of the ionospheric convection at high latitudes driven by magnetic reconnection, and its strong dependence on the interplanetary magnetic field (IMF) magnitude and the Sun-Earth component of the solar wind velocity  $v_x$ . It is also dependent on the IMF clock angle (e.g., Grocott & Milan, 2014). If we consider the IMF components  $B_x$ ,  $B_y$ , and  $B_z$  in Geocentric Solar Magnetospheric (GSM) co-ordinates (Hapgood, 1992), then the IMF clock angle,  $\theta_{\text{clock}}$ , is the angle between the projection of the IMF vector onto the GSM y-z plane and the GSM z axis,

$$\theta_{\text{clock}} = \tan^{-1} \left( \frac{B_y}{B_z} \right) \quad (1)$$

Indeed, the solar wind electric field magnitude ( $|v_x| \sqrt{B_y^2 + B_z^2}$ ), the IMF clock angle, and the dipole tilt of the Earth, drive the TS18 model (Thomas & Shepherd, 2018), which is also based on SuperDARN data. Second, we note that a study into the data-driven basis functions of the SuperDARN velocity field that we use here (Shore et al., 2021) finds that the IMF  $B_z$  component is the dominant driver for the background mean field and, as expected,  $B_z$  also drives a series of non-leading modes that describe the variability of the two-cell motion (e.g., Cowley & Lockwood, 1992) and that of substorms (e.g., Akasofu, 1964). However, it is IMF  $B_y$  that dominates the variability of the convection velocity (Shore et al., 2021). This observation, and the strong azimuthal asymmetries imposed by IMF  $B_y$  (e.g., Friis-Christensen & Wilhjelm, 1975; Friis-Christensen et al., 1985; Tenfjord et al., 2015), are motivations for the inclusion of IMF  $B_y$  as a driver in our model. Third, we consider the time lag between changes in the solar wind impinging the magnetosphere and the response in the ionosphere. Shore et al. (2019) performed a regression of surface external and induced magnetic field (SEIMF) variations onto solar wind data. Their paper examined the correlation between the terrestrial magnetic field and solar wind coupling parameters, as a function of time lag between the

two quantities. It was found that the epsilon coupling parameter  $\epsilon$  (Akasofu, 1979) resulted in a greater percentage of explained variance (which Shore et al. term the “prediction efficiency”) than the Milan parameter (Milan et al., 2012).

In addition to the cross-polar-cap electric potential driven by magnetic reconnection, the spatial variation of the high-latitude ionospheric convection velocity is also influenced indirectly by the plasma conductivity. Conductivity is a strong function of magnetic local time (MLT), the seasonal change in the tilt of the Earth toward/away from the Sun, and on changes in the output irradiance of the Sun. Our final additions to the model will allow for these conductivity variations and will also accommodate solar cycle variations via a dependence on both the day-of-year and the phase of the solar cycle. We are developing a model from purely northern hemisphere data, and so this can be straightforwardly applied to the northern hemisphere. It can also be used for the southern hemisphere by applying the model with all the same input coefficient values, but for an IMF  $B_y$  of opposite sign, with the caveat that any hemispheric differences other than the large-scale IMF  $B_y$ -related ones will be missing. Due to the lower amount of SuperDARN data covering the southern hemisphere, this is the best that we can currently achieve.

We initially create a hindcast model (the Version 1 or v1 model) using  $\epsilon$ , and then develop a second hindcasting version (v2) of the model that also includes IMF  $B_y$  as a driver. Finally, a version (v3) of the model is produced that is dependent on the day-of-year and a monthly value of  $F_{10.7}$ , the solar radio flux at 10.7 cm (Covington, 1948), to represent annual and solar cycle variations, respectively, including those in the ionospheric conductivity. We give details of the data sets used to develop the models in Section 2. In Section 3 we present our methodology for producing the three versions of the model. The results are presented in Section 4, including values for the percentage of explained variance, here given as a number between 0 and 1, which gives us a measure of how well the regression model is reproducing the parent data set. In Section 5, we focus on potential causes of variability in the percentage of explained variance, possible future improvements to the model, and our next steps. In Section 6, we present our summary and conclusions.

## 2. Data

In this section we discuss the data used to build the models. We build our model using 144 months of data for the interval 1997 to 2008 inclusive. A previous study used data from February 2001 to demonstrate the derivation of Empirical Orthogonal Function (EOF) model patterns from SuperDARN data (Shore et al., 2021), as this month has particularly good data coverage. We also use February 2001 (Figure 1) to illustrate our method. Please refer to the Data Availability Statement at the end of this paper for details regarding access to the data.

### 2.1. Solar Wind Data

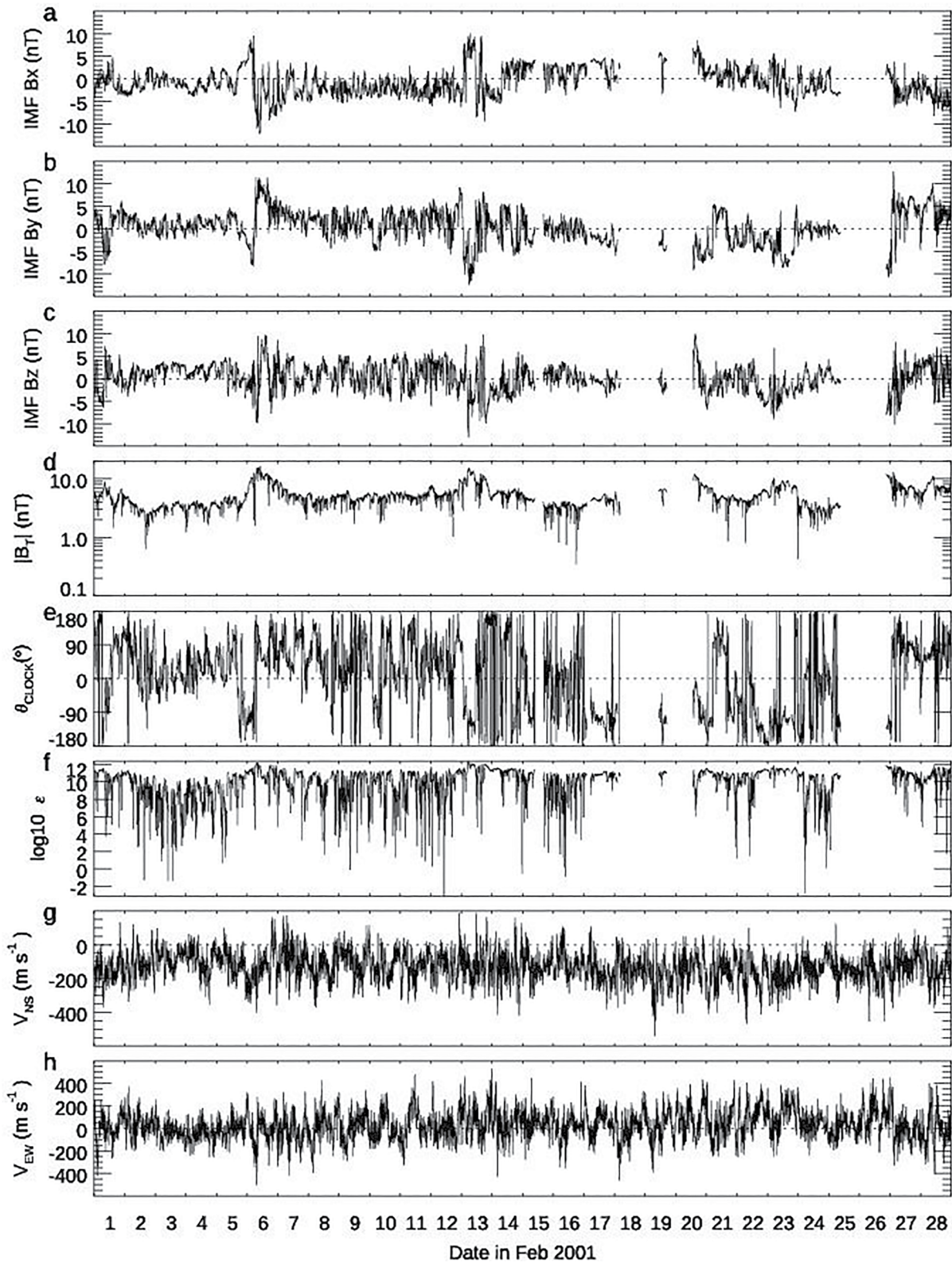
We make use of the IMF and the solar wind velocity data as extracted through OMNIWeb, specifically the OMNI 1-min data set (King & Papitashvili, 2005). The variables are in GSM co-ordinates (Hapgood, 1992). The OMNI solar wind data are provided already lagged from near the L1 Lagrangian point to the arrival time at the bow shock nose using the bow shock model of Farris and Russell (1994). This model requires the magnetopause nose distance which has been calculated for the data in terms of the solar wind pressure and IMF  $B_z$  using the magnetopause model of Shue et al. (1997). We create 5-min means of the IMF components  $B_x$ ,  $B_y$ , and  $B_z$ , and the solar wind velocity component  $v_x$ , time-stamped at the center of the 5-min epoch. In subsequent calculations, we omit any 5-min epoch for which there is no OMNI coverage. The remaining epochs are used to calculate the coupling parameter  $\epsilon$  at 5-min resolution,

$$\epsilon = \frac{4\pi}{\mu_0} \cdot |v_x| |B_T|^2 \sin^4\left(\frac{\theta_{\text{CLOCK}}}{2}\right) l_0^2 \quad (2)$$

where the permeability of free space  $\mu_0 = 4\pi \times 10^{-7}$ ,  $|B_T|$  is the solar wind magnetic field magnitude,  $l_0^2$  is a scale factor intended to represent the cross-sectional area over which dayside reconnection takes place with  $l_0 = 7 R_E$ , and  $R_E = 6,371.2$  km is the mean Earth radius. Here,  $\epsilon$  is measured in Watts (e.g., Akasofu, 1979).

### 2.2. The Shore EOF Analysis of SuperDARN Velocity Data

The convection velocities that we use here to develop our models have been published (Shore et al., 2022). These data will be referred to in this paper as the Shore EOF model analysis patterns or values. The Shore EOF model analysis used SuperDARN plasma velocity observations of the  $F$ -region ionosphere measured using the northern hemisphere radars of the SuperDARN global array. The fitted Doppler velocities were derived from the



**Figure 1.** Solar wind and ionospheric convection velocity data for February 2001. Interplanetary magnetic field (IMF) values (a)  $B_x$ , (b)  $B_y$ , (c)  $B_z$ , and (d)  $|B_T|$ , (e) IMF clock angle  $\theta_{\text{CLOCK}}$ , (f) logged epsilon parameter  $\log_{10} \epsilon$ , (g) north-south, and (h) east-west components of the ionospheric convection velocity as given by the Shore Super Dual Auroral Radar Network Empirical Orthogonal Function model analysis at 73.0°N magnetic latitude and 00:09 magnetic local time.

original autocorrelation functions using version 4.5 of the radar software toolkit, RSTv4.5 (SuperDARN Data Analysis Working Group, 2021) and, within that toolkit, fitting routine “FitACF v2.5.” In the production of the Shore EOF model patterns, the geolocation is dealt with using the Chisham virtual height model (Chisham et al., 2008), which is the state-of-the-art methodology. Further details about the selection of good quality data from the F-region are available in the data section of Shore et al. (2021).

The Shore EOF model analysis patterns were determined by applying the EOF method (Shore et al., 2021) to 12 years of SuperDARN plasma velocity measurements (1997–2008 inclusive), which is similar to the length of a typical solar cycle. In contrast to other SuperDARN methods of representing convection, the Shore method includes a self-consistent SuperDARN infill solution, which does not rely on climatological averages or external information, and which achieves complete coverage of the plasma velocity field variability in time and space.

The Shore EOF model analysis patterns (Shore et al., 2022) provide a complete representation of the northern hemisphere convection velocity field at ~250–400 km altitude (within the F-region ionosphere), at 5-min resolution. The velocity data are presented separately as the north-south (NS) and east-west (EW) components of the flow at 559 spatial locations. These locations are defined as the central co-ordinates of equal-area spatial bins extending from 60°N latitude to the pole in the Quasi-Dipolar co-ordinate system (Richmond, 1995). The latitude step for the equal-area grid is ~3.0°. The north and east directions are defined as positive for these two velocity components. Figures 1g and 1h show example NS and EW velocity components from the Shore EOF model analysis for February 2001 for a single location (73°N magnetic latitude, 00:09 MLT).

### 2.3. Solar Irradiance Proxy: $F_{10.7}$ Index

We make use of observations of the solar radio flux at 10.7 cm/2800 MHz (often called the  $F_{10.7}$  index) from the Low-Resolution OMNI (LRO) data set. This index correlates well with several ultraviolet (UV) and visible solar irradiance records and is easily measured at the Earth's surface. The daily values of  $F_{10.7}$  were used to create monthly mean values for the interval from 1997 to 2008 inclusive. Three of the months had missing data (missing 5 daily values or less), so those days were not included in the calculation of the monthly mean value.

## 3. Methodology

The OMNI data set timestamp takes into account the time lag between the satellite location and the Earth's bow shock nose. In order to best estimate the further time lag  $\tau$  of the ionospheric response, for use in their solar-wind-driven model of the Earth's magnetic field, Shore et al. (2019) calculated the peak correlation between the terrestrial magnetic field and  $\epsilon$ . The peak correlation commonly occurred at a time lag of ~20 min, but it varied with MLT and latitude. It was longer than 20 min in some parts of the nightside, reflecting the dominance there of the substorm response to the solar wind. In a similar fashion, we calculated the Pearson correlation coefficient between the  $\epsilon$  parameter and each velocity component of the ionospheric plasma  $\mathbf{E} \times \mathbf{B}$  drift velocity at all locations, for the February 2001 data set. We calculated the correlation for a series of time lags ranging from –10 to +500 min in 5-min steps up to  $\tau = 150$  min, then 10-min steps thereafter, totaling 68 separate lags. The results (not shown) strongly resemble the published correlations between  $\epsilon$  and the Earth's SEIMF (Shore et al., 2019) in that the correlation in the polar region tends to peak at a time lag of ~20 min, with a secondary peak corresponding to a longer time lag for some parts of the nightside ionosphere. The time lag value of ~20 min is supported by much of the literature (e.g., see Grocott and Milan (2014)) and therefore we deem 20 min to be a representative time lag for much of the ionosphere and use that value to develop our models in this paper.

The Shore EOF model data set comprises 144 monthly individual analyses of SuperDARN plasma velocity data, extending from January 1997 through to December 2008. For each month of EOF model data, we produce two versions of a hindcast model. Version 1 (v1) is produced via a regression of the data onto  $\epsilon$  alone, Version 2 (v2) is produced from a regression of the data onto both  $\epsilon$  and IMF  $B_y$ . We then produce a Version 3 (v3) model by performing a regression analysis of the 144 sets of monthly regression coefficients from the v2 model, with respect to trigonometric functions of day-of-year and to the monthly  $F_{10.7}$ . The coefficients resulting from this analysis can be used to form a model that could potentially be used for forecasting, which we will refer to as the Lam 2023 (v3) model. We make use of version 8.7 of the Interactive Data Language (IDL) REGRESS function, which performs a multiple linear regression fit (Harris Geospatial Solutions, 2022).

We will use a simple measure to assess how well each model that we build reproduces the Shore EOF model values. Let  $n$  denote the index for the 559 spatial bins and  $i$  denote the index for each 5-min interval in a given

month. For a given velocity component, let  $y_n^i$  denote the set of Shore EOF model velocity values,  $\bar{y}^i$  denote the spatial mean of  $y_n^i$ , and  $\hat{V}_n^i$  denote the set of our model velocity values. For each of our model versions, we calculate:

$$P(t_i) = 1 - \frac{\sum_{n=1}^{559} (y_n^i - \hat{V}_n^i)^2}{\sum_{n=1}^{559} (y_n^i - \bar{y}^i)^2} = 1 - e/A \quad (3)$$

We can see that Equation 3 can be expressed in terms of the mean squared “error”  $e$  between one of our regression models (v1, v2, or v3) and the Shore EOF model values, and the mean squared deviation  $A$  of the Shore values from their spatial average value. This is similar to the formula for the prediction efficiency which is a measure of skill (e.g., Equation 19 of Liemohn et al. (2021)). We shall use  $P(t_i)$  as a way of evaluating the ability that each of our models has at reproducing the parent data set, namely the Shore EOF model values. Hence, we refer to  $P(t_i)$  as the “percentage of explained variance” rather than the “prediction efficiency.” For instance, we can calculate the time development of the percentage  $P(t_i)$  for the NS velocity component in February 2001,  $P_{NS}(t_i)$ , for  $i = 1$  to 8,064 where 8,064 is the number of 5-min intervals in that month. We can calculate  $P_{EW}(t_i)$  in a similar fashion. This measure applies to the velocity component across the whole spatial region for any given 5-min interval. Obviously, it does not provide a validation of the model against other independent data sets provided, for instance, by satellite or incoherent scatter radar. It only provides a measure of how well the regression model performs at reproducing the data set from which it was derived.

### 3.1. Version 1 Model

We first create a simple linear hindcast model. The model relates the NS and EW components of the Shore Super-DARN EOF model velocity to  $\epsilon$  via:

$$V_{INS} = m_{1\epsilon NS} \epsilon + C_{1NS} \quad (4a)$$

$$V_{IEW} = m_{1\epsilon EW} \epsilon + C_{1EW} \quad (4b)$$

In all versions of our models, we use  $m$  to denote the slopes and  $C$  to denote the constant terms. The subscripts in these coefficients are as follows: 1 denotes model version 1,  $\epsilon$  denotes that the coefficient relates to the solar wind coupling parameter  $\epsilon$ , and NS (or EW) relates to that specific velocity component. For a given month, for each velocity component and for each of the 559 EOF model analysis location bins, we perform a linear regression analysis of the 5-min velocity data with respect to  $\epsilon$ , as indicated by Equations 4a and 4b.

### 3.2. Version 2 Model

The second version of the hindcast model relates the NS and EW components of the Shore EOF model velocity to  $\epsilon$  and IMF  $B_y$  via:

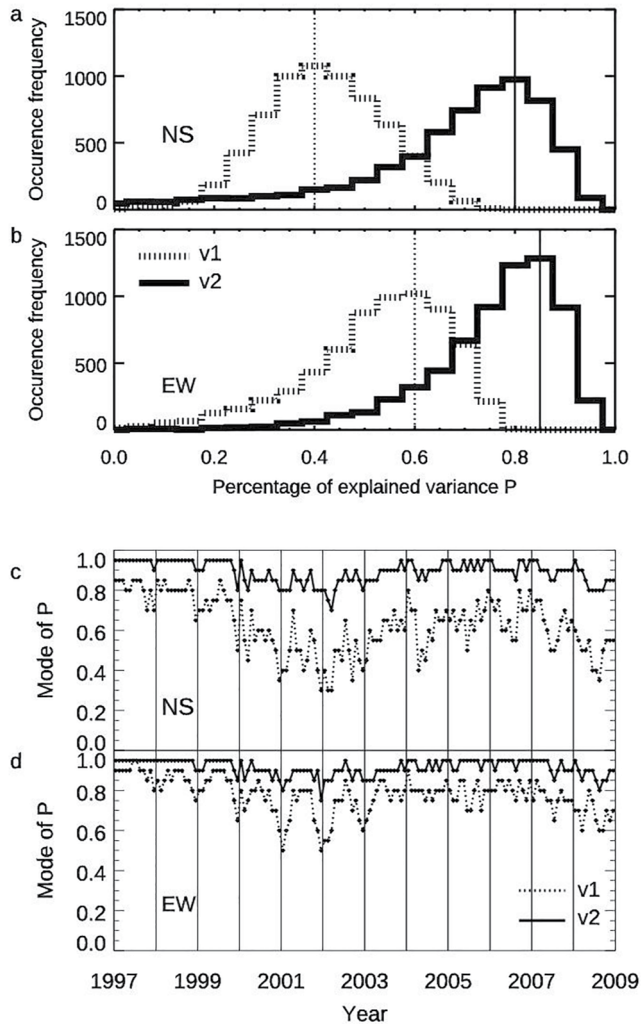
$$V_{2NS} = m_{2\epsilon NS} \epsilon + m_{2BNS} B_y + C_{2NS} \quad (5a)$$

$$V_{2EW} = m_{2\epsilon EW} \epsilon + m_{2BEW} B_y + C_{2EW} \quad (5b)$$

where the subscript  $B$  in the coefficients denotes the IMF component  $B_y$ . We again perform the regression analysis for each velocity component, each location bin and for each of the 144 months.

### 3.3. Version 3 (Lam 2023) Model

The third and final version of the model could potentially be used for forecasting, dependent on the availability of forecasts of the input variables, and allows for the indirect effects of conductivity variations and solar cycle variations. It is built from the six v2 model coefficients in Equations 5a and 5b, for which there are values for each of 144 months and each of the 559 spatial locations. Conductivity depends on the seasonal change in the tilt of the Earth toward or away from the Sun, and on changes in the output irradiance of the Sun. Therefore, for each velocity component and each location bin, we perform a regression analysis using the 144 monthly values of each



**Figure 2.** Comparison of the percentage of explained variance  $P(t)$  for the version 1 and 2 hindcast models. The occurrence frequency of  $P(t)$  across the whole model region during February 2001, with the monthly mode marked by a vertical line for (a) the north-south (NS) and (b) the east-west (EW) components of convection velocity. The monthly mode of the percentage of explained variance  $P(t)$  for the v1 and v2 models for (c) the NS, and (d) the EW velocity components of ionospheric velocity over 12 years.

regression coefficient, in terms of a trigonometric function of the day-of-year and the monthly value of  $F_{10.7}$ . For instance, for the NS velocity component, the regression equations are:

$$m_{2\epsilon NS} = s_{3\epsilon NS} \sin x + c_{3\epsilon NS} \cos x + F_{3\epsilon NS} F_{10.7} + C_{3\epsilon NS} \quad (6a)$$

$$m_{2BNS} = s_{3BNS} \sin x + c_{3BNS} \cos x + F_{3BNS} F_{10.7} + C_{3BNS} \quad (6b)$$

$$C_{2NS} = s_{3CNS} \sin x + c_{3CNS} \cos x + F_{3CNS} F_{10.7} + C_{3CNS} \quad (6c)$$

where  $x = 2\pi(t_j - 79)/365.25$  and  $t_j$  is the day-of-year in the middle of each of the 12 months of the year ( $j = 1$  to 12). Following Shore et al. (2019), we assume that the length of a year is 365.25 days and place the zero of the sine function at Spring equinox and we use day-of-year 79 (20 March) to represent the vernal equinox (Coxon et al., 2016). The regression analysis on the coefficient relating to the slope of  $\epsilon$  in the v2 model, that is, Equation 6a, yields four coefficients:  $s_{3\epsilon NS}$ ,  $c_{3\epsilon NS}$ , and  $F_{3\epsilon NS}$  are the slopes and  $C_{3\epsilon NS}$  is the intercept term. This is also the case for the slope in  $B_y$  (Equation 6b) and the intercept term in the v2 model (Equation 6c), resulting in 12 coefficients for the NS regression analysis. There is an equivalent set of equations for the EW velocity component (not given here), which means that there are 24 regression coefficients that define the v3 model. These model coefficients (Lam et al., 2023) could be used to forecast by specifying the day-of-year in the middle of the current month (or the actual DOY) and the current monthly mean of  $F_{10.7}$ . Equations 6a–6c then define values of the three coefficients of the v2 NS model for each of the 559 spatial bins. The same process is used to find the values of the three coefficients for the EW component of the v2 model for each of the 559 spatial bins. Equations 5a and 5b require values of  $\epsilon$  and IMF  $B_y$  to produce the associated plasma velocities. Real-Time Solar Wind (RTSW) data from spacecraft located upstream of the Earth, typically orbiting the L1 Lagrange point, can provide the values needed to forecast Earth's ionospheric electric field, as will be discussed in Section 5.

## 4. Results

### 4.1. Version 1 Model

We use the regression coefficients found from Equations 4a and 4b to determine the v1 model convection velocities. We assess how well the v1 model recreates the Shore EOF model analysis of the SuperDARN data by calculating the regional values of percentage  $P(t)$  using Equation 3. Figures 2a and 2b show the occurrence distribution of  $P(t)$  for the two velocity components for February 2001 (dashed line). The shape of the distribution has a clear peak value (the monthly mode value), which is typical of the distributions observed for any of the 144 months. The mode values for this month are 0.40 (NS velocity component) and 0.60 (EW velocity component). These mode values are always positive which indicates that the v1 model has skill in recreating the Shore EOF model values.

There is a cyclical nature to the mode in  $P(t)$  over the course of the 12-year interval (Figures 2c and 2d, dashed line) which is about the length of a solar cycle. There also appears to be a seasonal cycle within the variation of percentage  $P(t)$ . For instance, in (northern hemisphere) summertime  $P(t)$  is higher for 1999–2003 (especially in the EW component) and lower for 2004–2008 in the NS component. We characterize percentage  $P(t)$  for this version of the model by finding the average value,  $\chi_{v1}$ , of the monthly mode values over the whole 12-year period and over both velocity components. The characteristic percentage of explained variance for the v1 model  $\chi_{v1} = 0.70$ .

#### 4.2. Version 2 Model

As discussed in the introduction, there are reasons why the model would be improved by adding an explicit IMF  $B_y$  dependence to the convection velocity. Having obtained the regression coefficients  $m_{2eNS}$ ,  $m_{2BNS}$ ,  $C_{2NS}$ ,  $m_{2eEW}$ ,  $m_{2BEW}$ , and  $C_{2EW}$ , we use Equations 5a and 5b to construct the velocities for the v2 model. We derive the values of percentage  $P(t)$  for the v2 model using Equation 3 for February 2001 (Figures 2a and 2b, solid line). A comparison with the value for the v1 model shows that the monthly mode of percentage  $P(t)$  for the polar region has increased from 0.40 (v1) to 0.80 (v2) for the NS plasma velocity component, and from 0.60 (v1) to 0.85 (v2) for the EW component. The v2 formulation has markedly improved the percentage of explained variance for both velocity components compared with those of the v1 model. This improvement, resulting from the addition of IMF  $B_y$ , is seen for the whole 12-year period (Figures 2c and 2d). The seasonal variation in the monthly mode value of percentage  $P(t)$  is no longer apparent in the v2 model results, and the sinusoidal-like solar-cycle-related dependence is much reduced in magnitude. The characteristic value of percentage  $P(t)$  for the v2 model  $\chi_{v2} = 0.90$  is also an improvement over that of the v1 model ( $\chi_{v1} = 0.70$ ). This indicates that the v2 model is a high-quality hindcast model that is good enough to base a forecast model on.

For February 2001, the maps of the regression coefficients of the v2 model (Figures 3a–3f) possess a high degree of coherent spatial structure, which is further evidence that the solar wind variables in the v2 model are good choices. The spatial pattern formed by the IMF  $B_y$  slope coefficients that relates to the EW plasma velocity values,  $m_{2BEW}$  (Figure 3d), is similar to the dominant spatial pattern of the EOF analysis of the SuperDARN data from February 2001 (Figure 4, Shore et al., 2021). This dominant mode (Mode 1, labeled Velocity Perturbation Y in Shore et al. (2021)) relates to the flow driven by variability in IMF  $B_y$  after reconnection and is very similar to the Disturbance-Polar Type Y in the magnetic field (Friis-Christensen & Wilhjelm, 1975). Using the v2 model coefficients to build a model that could be used for operational purposes has the advantage that the v2 model velocities can be estimated from only two solar wind quantities, namely the IMF and the solar wind velocity.

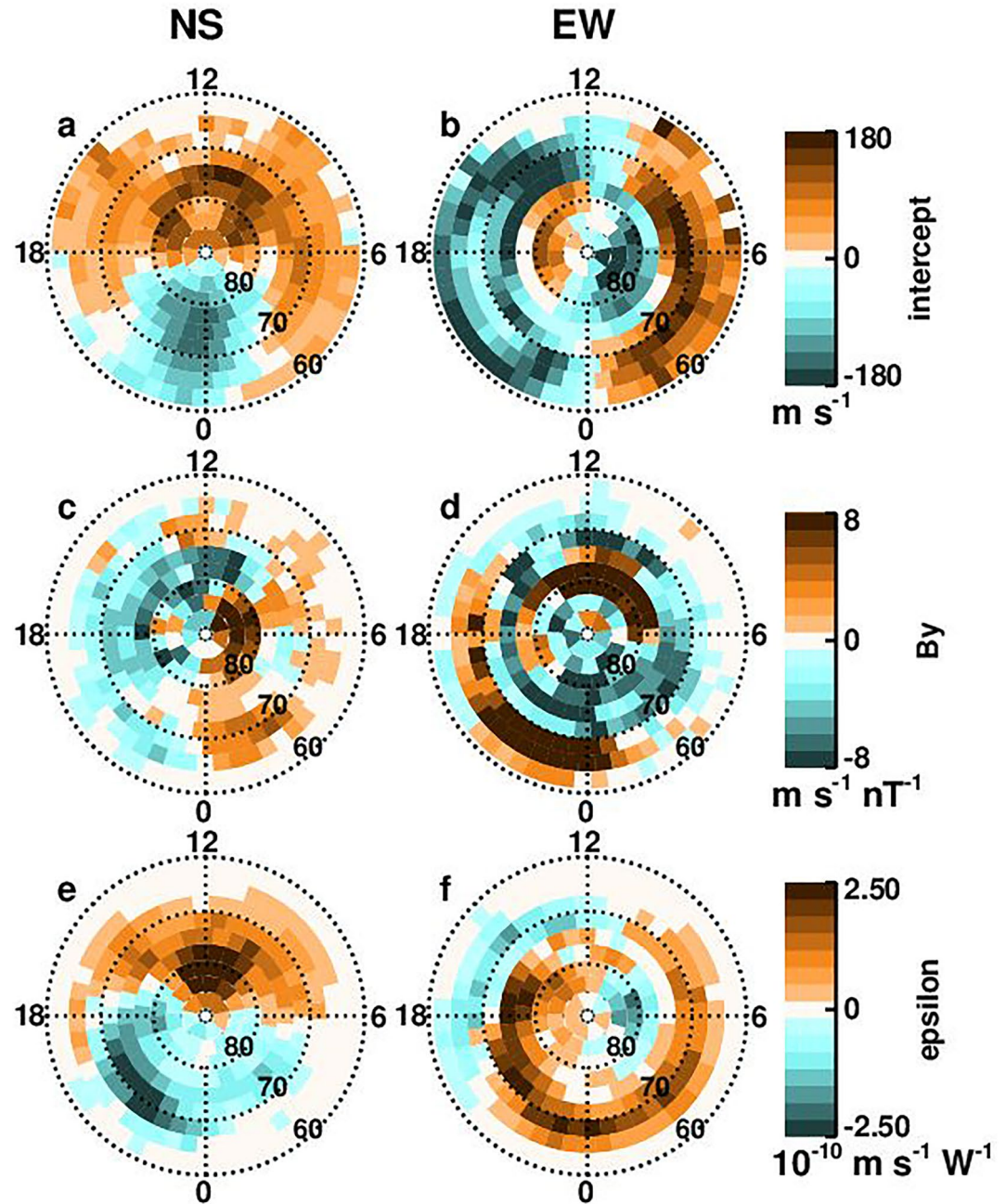
#### 4.3. Version 3 (Lam 2023) Model

The high-latitude ionospheric convection velocity is influenced by solar cycle and seasonal variations, such as those in the ionospheric plasma conductivity. Hence, we formulate the final version of the model using a regression analysis of v2 model regression coefficients with respect to a trigonometric function of the day-of-year, the monthly value of  $F_{10.7}$ , and an intercept (Equations 6a–6c).  $F_{10.7}$  is more than a conductivity marker—it tracks the solar cycle and so its inclusion may allow the representation of other solar-cycle-related variations. The maps of the regression coefficient values in the polar region (Figures 4 and 5) show a high degree of coherent spatial structure. If the maps lacked coherent spatial structure, then we might question whether the variables in the v3 model are reasonable choices, or whether there are sufficient data in the analysis. The Lam 2023 (v3) model is visibly less skilled (Figures 6a and 6b) than the v2 hindcast model at reproducing the Shore EOF model patterns, but this is understandable given that the v2 model is based on EOF patterns for the month in question, whereas the Lam (v3) model is not. The Lam model has an overall characteristic value for percentage  $P(t)$  of  $\chi_{v3} = 0.70$  and is particularly skilled for the years 2000–2004 inclusive (for which  $\chi_{v3} = 0.75$ ), as will be discussed further in Section 5.

We represent the convection velocity visually by taking the velocity value for a particular spatial bin and assuming a packet of plasma travels on the velocity trajectory for a characteristic time related to the resolution of the measurement. The end position of the packet is found using spherical trigonometry and scaled for ease of viewing. We present an example of how well the Lam 2023 (v3) model reproduces the Shore EOF model velocity field at a time when percentage  $P(t)$  is very high, during February 2001. At 05:42 UT (13 February 2001),  $P(t) \approx 0.85$  for both the NS and EW velocity components. The morphology of the Shore EOF model velocity field (Figure 7a) is reproduced very well by the Lam (v3) model (Figure 7b), including the shape of the two-cell motion and much of the plasma flow across the magnetic pole.

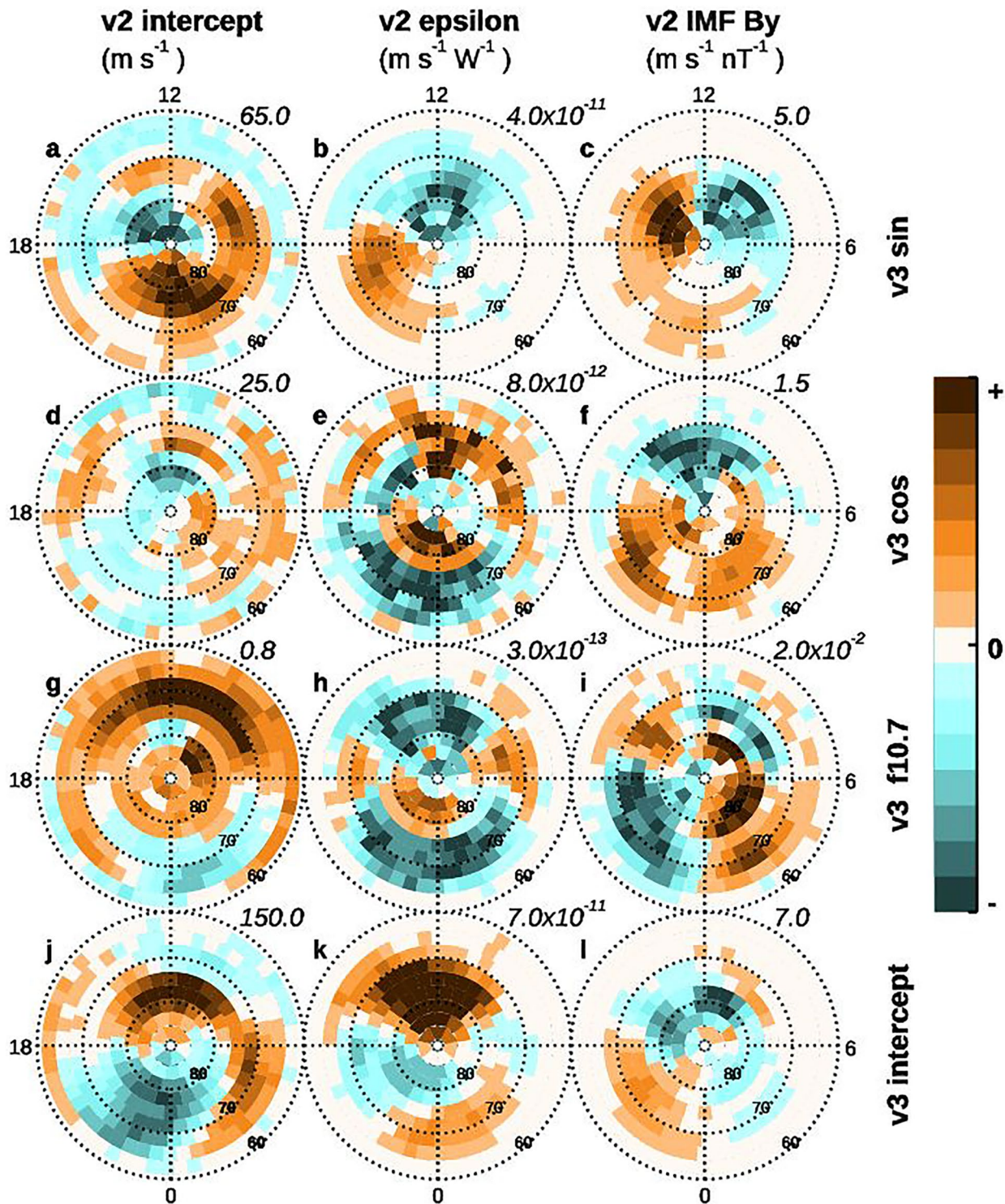
The two intervals either side of solar maximum, 1997 to 1999 inclusive and 2005 to 2008 inclusive, are characterized by lower values for percentage  $P(t)$  ( $\chi_{v3} = 0.68$  and  $0.67$ , respectively) and have excursions in the monthly mode values that fall below 0.5 (Figures 6a and 6b). One of the lowest values in the monthly mode of percentage  $P(t)$  occurs for June 1999. We again present an example of how well the Lam 2023 (v3) model reproduces the Shore EOF model velocity field, but for a time when percentage  $P(t)$  has a low value. On the 18 June at 07:17 UT, percentage  $P(t) = 0.40$  in both the NS and EW components. The Lam (v3) model (Figure 7d) velocity field possesses a clearer two-cell structure than the Shore EOF model data (Figure 7c), especially around





**Figure 3.** Version 2 model regression coefficients for February 2001. Maps of the northern hemisphere polar region in magnetic latitude and magnetic local time. The values of the intercept: (a)  $C_{2NS}$  (for the NS convection velocity component) and (b)  $C_{2EW}$  (for EW); the interplanetary magnetic field  $B_y$  slope values: (c)  $m_{2BNS}$  (north-south [NS]) and (d)  $m_{2BEW}$  (east-west [EW]); the  $\epsilon$  slope values: (e)  $m_{2eNS}$  (NS) and (f)  $m_{2eEW}$  (EW), in Equations 5a and 5b.

noon. The Lam model has also replaced the anomalous vectors (those not of similar size and/or direction to the surrounding flow) at lower latitudes in the Shore SuperDARN EOF model velocity field. Since these values are likely to be poor data (Shore et al., 2021), the Lam model values may be more accurate than these specific Shore EOF model values. This means that the v2 model has some usefulness in its own right as a data set (also see Section 5).



**Figure 4.** Regression coefficients for the Lam 2023 ( $v_3$ ) model. These values relate to Equations 6a–6c (north-south velocity component). The coordinate system is the same as in Figure 3. The coefficients for the sine terms: (a)  $s_{3\text{CNS}}$ , (b)  $s_{3\text{eNS}}$ , and (c)  $s_{3\text{BNS}}$ ; the cosine terms: (d)  $c_{3\text{CNS}}$ , (e)  $c_{3\text{eNS}}$ , and (f)  $c_{3\text{BNS}}$ ; the  $F_{10.7}$  terms: (g)  $F_{3\text{CNS}}$ , (h)  $F_{3\text{eNS}}$ , and (i)  $F_{3\text{BNS}}$ ; and for the intercept terms: (j)  $C_{3\text{CNS}}$ , (k)  $C_{3\text{eNS}}$ , and (l)  $C_{3\text{BNS}}$ . The maximum value (and minus the minimum value) of the color scale are at the top right of each dial in italics. The units for each column of coefficients are written at the top of that column, for example,  $v_2$ -intercept-related values are in  $\text{m s}^{-1}$ .

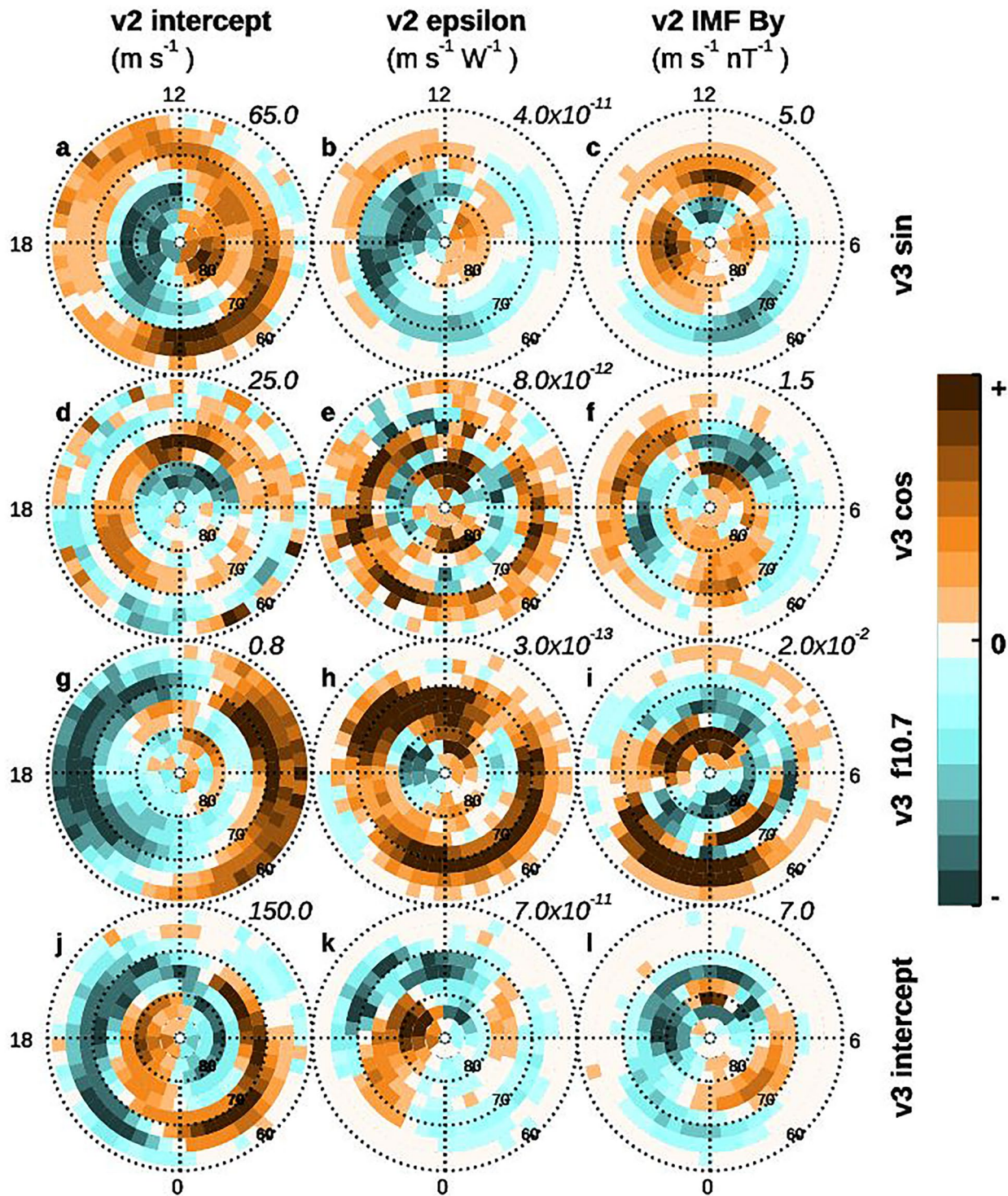
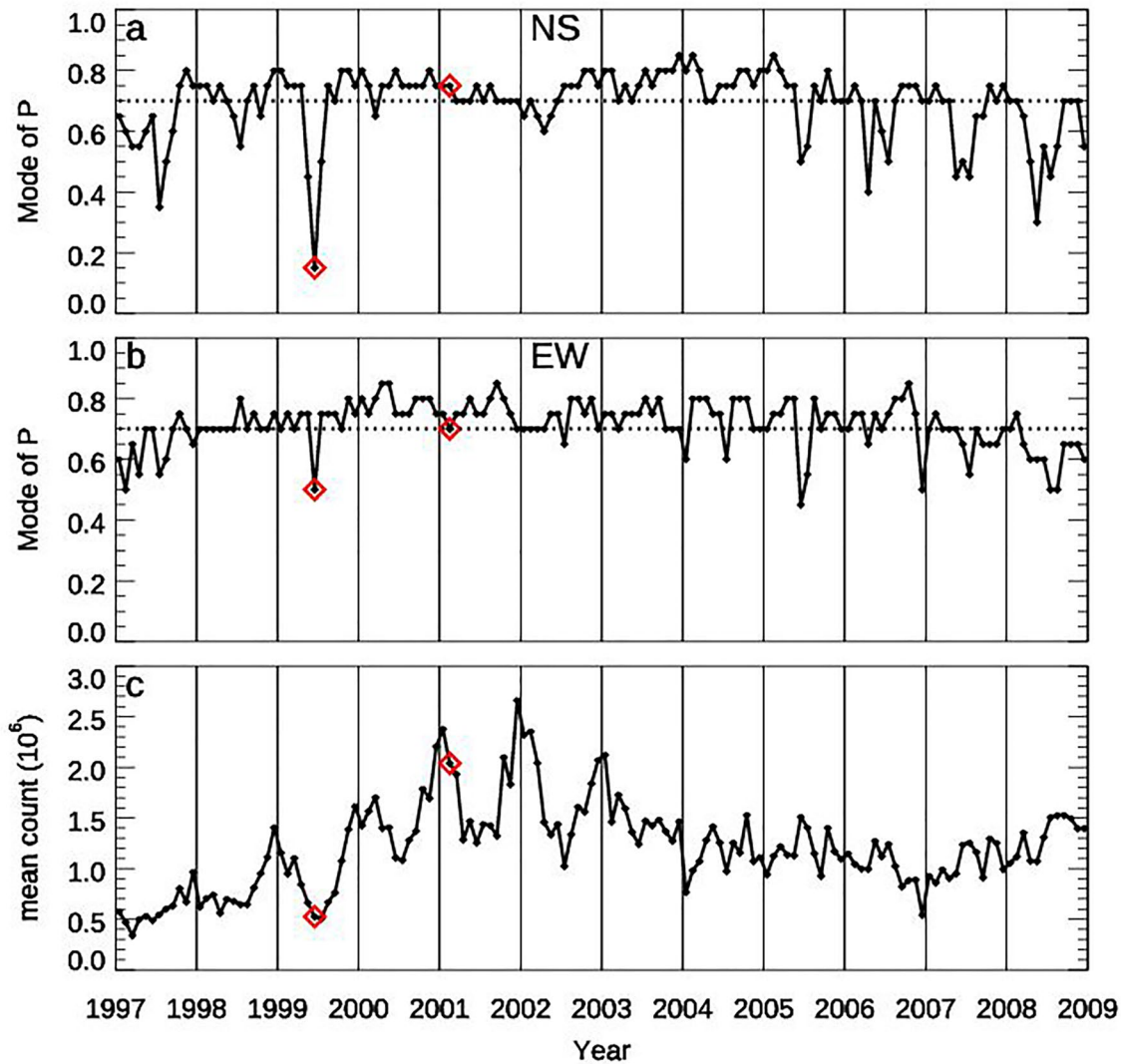


Figure 5. Regression coefficients ( $v_3$  model). As for Figure 4, but for the east-west ionospheric velocity component.

#### 4.4. Summary of Model Results

The  $v_1$  regression model, based on the epsilon parameter (Akasofu, 1979) has some ability to reproduce the original Shore EOF model patterns, and possesses both a seasonal and a solar-cycle-related sinusoidal-like structure in the monthly mode of the percentage of explained variance  $P(t)$ . We defined a “characteristic value” for  $P(t)$  as the mean value,  $\chi_{v_1}$ , of the monthly mode value of  $P(t)$  over the whole 12-year period and

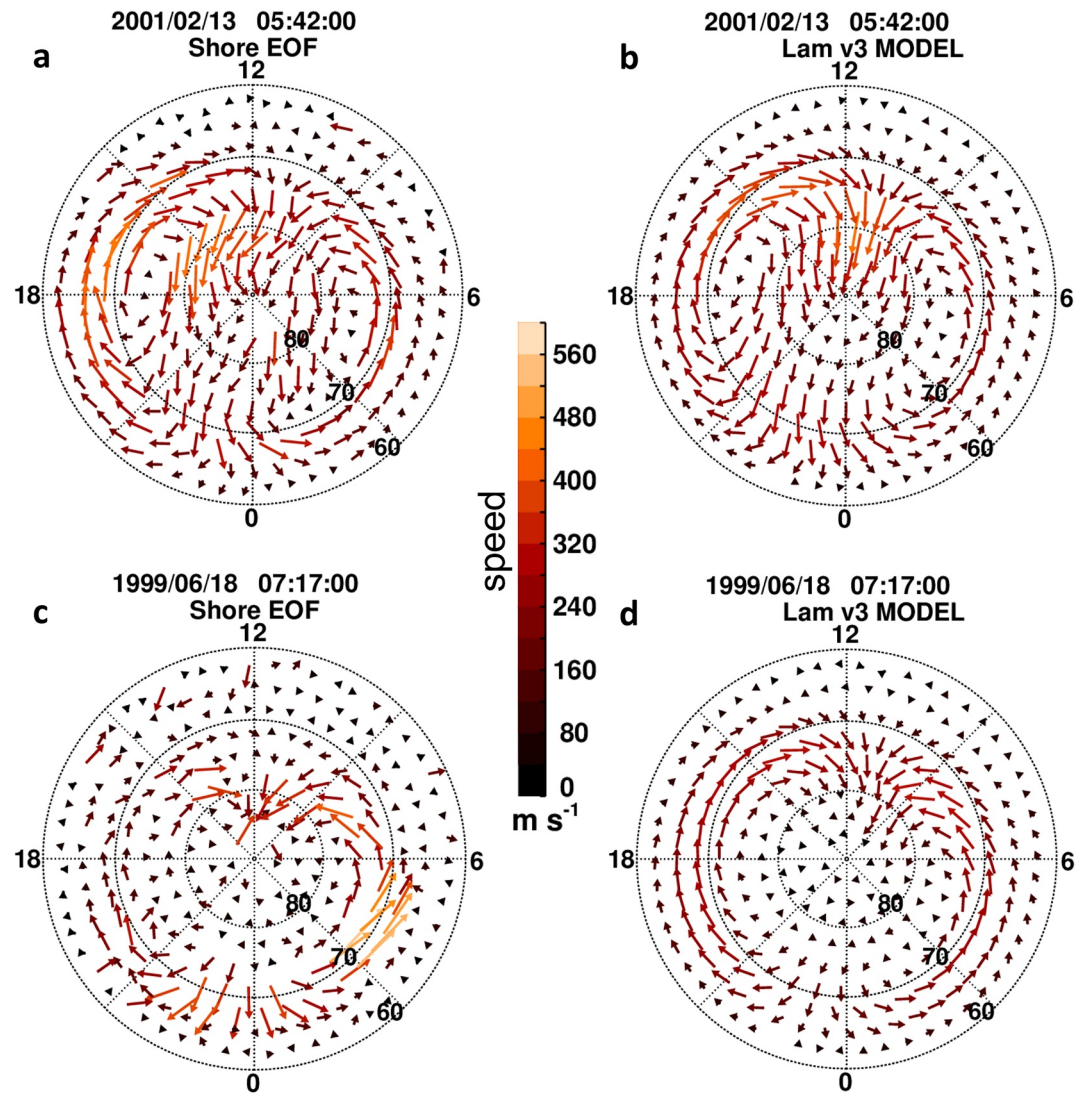


**Figure 6.** The percentage of the explained variance  $P(t)$  for the Lam 2023 model (v3) and the Super Dual Auroral Radar Network (SuperDARN) radar data coverage. The monthly mode of the percentage of explained variance  $P(t)$  for the v3 model for the: (a) north-south and (b) east-west ionospheric velocity components. The dotted lines in panels (a, b) show the characteristic value of the percentage of explained variance for the v3 model; (c) the data coverage level of the SuperDARN instruments as given by the total monthly data count using all location bins. For all panels, the data points for June 1999 and February 2001 are marked by red diamonds.

over both velocity components. The characteristic value for  $P(t)$  for the v1 model  $\chi_{v1} = 0.70$ . This value is improved on by the v2 model for the whole 12-year period (Figures 2c and 2d), due to the addition of IMF  $B_y$ . Compared to the v1 model, the seasonal structure in the monthly v2 mode of  $P(t)$  is no longer apparent, the solar-cycle variation has been much reduced, and the v2 model characteristic percentage  $P(t)$  is significantly larger ( $\chi_{v2} = 0.90$ ). The Lam 2023 (v3) model has a lower value for  $P(t)$  than the v2 hindcast model and the same value as the v1 hindcast model ( $\chi_{v3} = 0.70$ ) but, importantly, it has the potential to be used to forecast ionospheric convection or to nowcast outside of the epoch of the Shore EOF model analysis, as outlined in the next section.

### 5. Discussion

We have developed a model of high latitude ionospheric convection that is both skilled at reproducing its parent data set and suitable for operational purposes, as part of the SWIMMR-T project. It is worth briefly speculating here on the causes of the drops in the percentage of explained variance  $P(t)$  that we see in the 12-year interval examined, both for scientific interest and to aid the development of future improved versions of the model.



**Figure 7.** Comparison of Lam 2023 (v3) model with the original Shore Empirical Orthogonal Function (EOF) model velocity field. Five-min snapshots selected from the two months marked by red diamonds in Figure 6. The top row shows velocities at a time when the percentage of explained variance  $P(t)$  is high (February 2001): (a) the Shore Super Dual Auroral Radar Network (SuperDARN) EOF model analysis patterns, (b) the Lam 2023 (v3) model. The bottom row is for a time of low  $P(t)$  (June 1999), (c) the Shore SuperDARN EOF model analysis patterns, and (d) the Lam 2023 (v3) model. Color is used to indicate speed as there is no single arrow length for a reference speed.

First, one potential cause of reductions in percentage  $P(t)$  is a decrease in the amount of SuperDARN data used to produce the Shore EOF model velocity patterns for a particular month. The data coverage within the SuperDARN archive has gaps, both spatially and temporally. The SuperDARN data coverage varies with time of day, year, and solar cycle. This is due to the variation in the high-frequency (HF) propagation conditions, the level of irregularity occurrence, and the dependence of solar wind-magnetosphere coupling on solar cycle phase. In addition, since the radars only measure signals along the line-of-sight direction, the accuracy of the velocity vectors will vary with location and time, causing variations in the degree to which the EOF analysis of the data set is able to represent the plasma motion in any given region. At low latitudes, data gaps have been filled by a sinusoidal fitting procedure (Shore et al., 2021). When the data coverage is very low, the fitted north and east directions can be unrepresentative of the true values, resulting in high-error vectors at lower latitudes in the Shore EOF model values (e.g., Figure 7c), that do not resemble the surrounding flow.

We examine the binned monthly SuperDARN data coverage for the Shore EOF model patterns, where full coverage for a given location bin is defined as having a valid data point in all the  $6^\circ$ -wide look direction bins and each of the 5-min epochs for the whole month. The monthly Shore EOF model analysis radar count is the number of data points in that month when summed over the direction bins, 5-min epochs and location bins. We investigate here whether there is any relationship between the amount of SuperDARN data in a particular month and the value of the percentage of explained variance  $P(t)$  for the Lam 2023 (v3) model. The monthly Shore EOF model data count (Figure 6c) is mostly below  $\sim 1.5 \times 10^6$  from the start of the data set (January 1997) until September 2000. Between September 2000 and April 2003, the data count oscillates and is often above  $\sim 1.5 \times 10^6$  but is mostly below this value for the remainder of the 12-year interval examined. Therefore, there is a gross correspondence between the SuperDARN data coverage (Figure 6c) and the value of percentage  $P(t)$  for the Lam (v3) model (Figures 6a and 6b). However, there is no indication that there is a one-to-one correspondence on a monthly timescale. This is consistent with values of the correlation (Pearson coefficient) between the mean data count and the Lam 2023 (v3) model value for percentage  $P(t)$ , which are 0.24 for both the NS and EW velocity components.

It is undisputed that the reduced SuperDARN data coverage will affect the quality of the Shore EOF model patterns and therefore the percentage of explained variance  $P(t)$  of the model developed, but it does not seem to be the only factor. If the quality of the Shore SuperDARN EOF model values is generally high for the 12-year interval, then the quality of the Lam (v3) model has the potential to be good, and indeed appears to be good, according to the characteristic value for percentage  $P(t)$ . Therefore, it is possible that at times when percentage  $P(t)$  is low, the Lam (v3) model may give a better idea of the plasma motion when the SuperDARN coverage is low and the Shore EOF model analysis values are unreliable, especially around solar maximum. Future comparisons with an independent data set should resolve this.

A second potential cause for a fall in the percentage of variance explained by the model is a decrease in how well the physical processes driving the convection velocity are being represented. The Lam (v3) model exhibits two intervals of lower and variable values for percentage  $P(t)$ . The first spans 1997 to 1999 inclusive, which are the three years just after the solar minimum of August 1996. The second (2005–2008 inclusive) comprises the three years leading up to the solar minimum of December 2008. In contrast, the Lam (v3) model reproduces the parent data set consistently well in the months at the end of 1999 and for the interval 2000 to 2004 inclusive, that is, the two years before and the three years after the solar maximum of November 2001. It is possible that our model of magnetosphere coupling is more accurate around solar maximum and less accurate around solar minimum, due to the differing nature of the solar wind at different phases of the solar cycle (e.g., McComas et al., 2003) and the resultant differing geomagnetic effects (e.g., Tsurutani et al., 2006). We do not aim to prove that this is the case here but propose this as a hypothesis for further study. Although coupling functions should be used with care (Lockwood, 2022) and other more recent coupling functions are available (Lockwood & McWilliams, 2021), the Lam (v3) model has a respectable level of skill and should prove to be a valuable tool for operational purposes, due to its simplicity.

We anticipate that the Lam (v3) model may be improved by a change of the formulation, including additions to the parameters currently used. First, a study of the spatial dependence of the appropriate time lag, as has been done for the terrestrial magnetic field (Shore et al., 2019), would allow the model to be adjusted to have a location-dependent time lag. Second, an increase in the percentage of explained variance of the model may result from developing the v2 model from the preceding month's data and using it to forecast the coming month. Such a model would require continuous determination of monthly EOF model values, which in turn would require all SuperDARN data to be available on a near-real-time basis. This may be possible in the future. Finally, the movement of plasma in the ionosphere is ordered by the open-closed magnetic field line boundary (OCB) (e.g., Milan & Grocott, 2021). The inclusion of the OCB location and motion in the model should decrease errors that would otherwise arise when the OCB motion is fast and/or significant. We would welcome the production of a new version of the SuperDARN EOF model analysis values at higher spatial and temporal resolution, as we believe that this may prove to be a useful improvement to the data.

Our next challenges are to drive the model using RTSW data rather than the quality-checked OMNI 1-min data set used here, and to implement our model within a GCM such as TIEGCM (Qian et al., 2014). In practice, the latter will involve generating electric potential values from our model velocity values, since this is how the solar wind influence on the high-latitude ionospheric plasma is input into TIEGCM. The U.S. National Oceanic and

Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) data service provides real-time measurements from L1 (presently DSCOVR is the active satellite, with the advanced composition explorer (ACE) also contributing). These will be provided in real time by a combination of satellites monitoring near-Earth space, and forecast models, and so may have quality and data gap issues. This means that we have an opportunity to forecast the ionospheric potential that will be imposed on the Earth's ionosphere in advance, by a time interval  $T_{adv}$ . Observations of interplanetary shocks by the ACE spacecraft at L1 and associated sudden commencements in the magnetosphere can give an estimate of the propagation delay between L1 and the Earth. For instance, Baumann and McCloskey (2021) estimate that the delay varies between  $\sim 20$  min for extremely fast ICMEs (1,000 km/s) and  $\sim 90$  min for slow shocks (300 km/s). If we assume that there is also  $\sim 20$  min delay for the ionosphere to respond, then we can estimate the ionospheric potential in advance by a time interval  $T_{adv} \geq 40$  min, using the solar wind variables observed at L1 and currently provided by the NOAA SWPC (<https://www.swpc.noaa.gov/products/real-time-solar-wind>). The term RTSW refers to data from any spacecraft located upwind of Earth (typically orbiting the L1 Lagrange point) that is being tracked by the RTSW Network of tracking stations. Since we can trivially specify the day-of-year and only need the rolling monthly value of the  $F_{10.7}$  index to use our model, all the variables needed are already available to use the Lam (v3) model to forecast the ionospheric electric field. Our ultimate goal is to quantify the model's effectiveness in estimating satellite drag due to Joule heating in comparison to the existing climatological electric field models such as Weimer (1995, 1996, 2001, 2005) and Heelis et al. (1982).

## 6. Conclusions

We present the Lam (2023) model of the northern hemisphere high-latitude ionospheric convection velocity (and by extension the convection electric field), suitable for future development as an operational forecast model. It has been developed from 12 years (approximately a solar cycle's worth) of SuperDARN HF radar EOF model analysis patterns. The model has been developed in three key stages. First, we created a linear hindcast model of plasma velocity driven by the epsilon parameter  $\epsilon$ , with a universal time lag between the driver and ionospheric plasma at all polar region locations of 20 min. Second, we expanded the hindcast model to include IMF  $B_y$ , which generally resulted in a significant improvement in the monthly value of the percentage of explained variance during the solar cycle examined. Finally, a regression analysis of the seasonal and solar cycle dependence of the hindcast model coefficients was used to build a final version of the model. This final model (the Lam 2023 model) was driven by day-of-year, the monthly value of  $F_{10.7}$ , IMF  $B_y$ , and  $\epsilon$ . Therefore, forecasts of the ionospheric plasma convection velocity could be obtained from the observed rolling monthly average of  $F_{10.7}$  and currently available real-time values of IMF  $B_y$  and  $\epsilon$  at the L1 point upstream of Earth.

When the Lam 2023 model was assessed to see how well it could reproduce the Shore EOF model analysis patterns (the parent data set), it exhibited a respectable level of skill, and hence it could be a valuable tool for operational purposes due to its simplicity. In this paper, however, we have only verified our model by ensuring that the formulation is consistent with the data set upon which it was constructed. The validation of the model against other independent data sets provided, for instance, by satellite or incoherent scatter radar remains a necessary outstanding task. Although the data coverage of the parent data set from which the model is developed (the Shore et al. (2021) SuperDARN velocity EOF model analysis) will influence the ability of the Lam 2023 model to reproduce the parent data set, it is also possible that the assumed form of the regression equations is more accurate around solar maximum and less so around solar minimum. In which case we can improve on the regression equations used in future versions of the model.

## Data Availability Statement

We used NASA/GSFC's Space Physics Data Facility's OMNIWeb service <https://omniweb.gsfc.nasa.gov> and OMNI data, provided by the Goddard Space Flight Center (GSFC) at the National Aeronautics and Space Administration (NASA) to access the solar wind data such as IMF and velocity (N. E. Papitashvili & King, 2020a). It was also used to access the solar radio flux at 10.7 cm/2800 MHz which was obtained from the Low Resolution OMNI (LRO) data set (<https://omniweb.gsfc.nasa.gov/ow.html>) (N. E. Papitashvili & King, 2020b). The British Antarctic Survey (BAS) EOF model analysis of the SuperDARN plasma velocity data and supporting software (<https://doi.org/10.5285/2b9f0e9f-34ec-4467-9e02-abc771070cd9>), and also the peer-reviewed description of its derivation (Shore et al., 2021) have all been published. The regression coefficients for the models presented in

this study and accompanying software have also been published (<https://doi.org/10.5285/22272b8e-1aa3-483b-9867-224fe02db4e8>). We used version 8.7 of IDL to produce our figures (Harris Geospatial Solutions, 2022).

### Acknowledgments

The authors gratefully acknowledge the use of SuperDARN data. SuperDARN is a collection of radars funded by the national scientific funding agencies of Australia, Canada, China, France, Italy, Japan, Norway, South Africa, United Kingdom, and the United States. MML thanks S. Bronsdon (University of Leeds) for useful discussions. MML, RMS, GC, and MPF gratefully acknowledge funding from NERC through the standard Grant NE/V002732/1. RMS and MPF also gratefully acknowledge funding from NERC through the standard Grant NE/N01099X/1. AG and LO gratefully acknowledge funding from the National Environment Research Council (NERC) through the standard Grant NE/V00283X/1 and NE/V002686/1. AG and M-TW gratefully acknowledge funding from the National Environment Research Council (NERC) through the standard Grant NE/T000937/1.

### References

- Akasofu, S.-I. (1964). The development of the auroral substorm. *Planetary and Space Science*, 12(4), 273–282. [https://doi.org/10.1016/0032-0633\(64\)90151-5](https://doi.org/10.1016/0032-0633(64)90151-5)
- Akasofu, S.-I. (1979). Interplanetary energy flux associated with magnetospheric substorms. *Planetary and Space Science*, 27(4), 425–431. [https://doi.org/10.1016/0032-0633\(79\)90119-3](https://doi.org/10.1016/0032-0633(79)90119-3)
- Baumann, C., & McCloskey, A. E. (2021). Timing of the solar wind propagation delay between L1 and Earth based on machine learning. *Journal of Space Weather and Space Climate*, 11, 41. <https://doi.org/10.1051/swsc/2021026>
- Chisham, G., Lester, M., Milan, S. E., Freeman, M. P., Bristow, W. A., Grocott, A., et al. (2007). A decade of the super dual auroral radar Network (SuperDARN): Scientific achievements, new techniques and future directions. *Surveys in Geophysics*, 28(1), 33–109. <https://doi.org/10.1007/s10712-007-9017-8>
- Chisham, G., Yeoman, T. K., & Sofko, G. J. (2008). Mapping ionospheric backscatter measured by the SuperDARN HF radars – Part 1: A new empirical virtual height model. *Annales Geophysicae*, 26(4), 823–841. <https://doi.org/10.5194/angeo-26-823-2008>
- Covington, A. E. (1948). Solar noise observations on 10.7 centimeters. *Proceedings of the IRE*, 36(4), 454–457. <https://doi.org/10.1109/JRPROC.1948.234598>
- Cowley, S. W. H., & Lockwood, M. (1992). Excitation and decay of solar wind-driven flows in the magnetosphere-ionosphere system. *Annales Geophysicae*, 10(1–2), 103–115.
- Coxon, J. C., Milan, S. E., Carter, J. A., Clausen, L. B. N., Anderson, B. J., & Korth, H. (2016). Seasonal and diurnal variations in AMPERE observations of the Birkeland currents compared to modeled results. *Journal of Geophysical Research: Space Physics*, 121(5), 4027–4040. <https://doi.org/10.1002/2015JA022050>
- Elvidge, S., & Angling, M. J. (2019). Using the local ensemble transform Kalman filter for upper atmospheric modelling. *Journal of Space Weather and Space Climate*, 9, A30. <https://doi.org/10.1051/swsc/2019018>
- European Space Agency. (2022). Space debris by the numbers. Retrieved from [https://www.esa.int/Space\\_Safety/Space\\_Debris/Space\\_debris\\_by\\_the\\_numbers](https://www.esa.int/Space_Safety/Space_Debris/Space_debris_by_the_numbers)
- Farris, M. H., & Russell, C. T. (1994). Determining the standoff distance of the bow shock: Mach number dependence and use of models. *Journal of Geophysical Research*, 99(A9), 17681–17689. <https://doi.org/10.1029/94JA01020>
- Förster, M., Paschmann, G., Haaland, S. E., Quinn, J. M., Torbert, R. B., Vaith, H., & Kletzing, C. A. (2007). High-latitude plasma convection from Cluster EDI measurements: Variances and solar wind correlations. *Annales Geophysicae*, 25(7), 1691–1707. <https://doi.org/10.5194/angeo-25-1691-2007>
- Foster, J. C. (1983). An empirical electric field model derived from Chatanika radar data. *Journal of Geophysical Research*, 88(A2), 981–987. <https://doi.org/10.1029/JA088iA02p00981>
- Friis-Christensen, E., Kamide, Y., Richmond, A. D., & Matsushita, S. (1985). Interplanetary magnetic field control of high-latitude electric fields and currents determined from Greenland magnetometer data. *Journal of Geophysical Research*, 90(A2), 1325–1338. <https://doi.org/10.1029/JA090iA02p01325>
- Friis-Christensen, E., & Wilhelm, J. (1975). Polar cap currents for different directions of the interplanetary magnetic field in the Y-Z plane. *Journal of Geophysical Research*, 80(10), 1248–1260. <https://doi.org/10.1029/JA080i010p01248>
- Greenwald, R. A., Baker, K. B., Dudeney, J. R., Pinnock, M., Jones, T. B., Thomas, E. C., et al. (1995). DARN/SuperDARN: A global view of the dynamics of high-latitude convection. *Space Science Reviews*, 71(1), 761–796. <https://doi.org/10.1007/BF00751350>
- Grocott, A., & Milan, S. E. (2014). The influence of IMF clock angle timescales on the morphology of ionospheric convection. *Journal of Geophysical Research: Space Physics*, 119(7), 5861–5876. <https://doi.org/10.1002/2014JA020136>
- Hairston, M. R., & Heelis, R. A. (1990). Model of the high-latitude ionospheric convection pattern during southward interplanetary magnetic field using DE 2 data. *Journal of Geophysical Research*, 95(A3), 2333–2343. <https://doi.org/10.1029/JA095iA03p02333>
- Hapgood, M. A. (1992). Space physics coordinate transformations – A user guide. *Planetary and Space Science*, 40(5), 711–717. [https://doi.org/10.1016/0032-0633\(92\)90012-D](https://doi.org/10.1016/0032-0633(92)90012-D)
- Harris Geospatial Solutions. (2022). Regress. Retrieved from <https://www.13harrisgeospatial.com/docs/regress.html>
- Heelis, R., Lowell, J. K., & Spiro, R. W. (1982). A model of the high-latitude ionospheric convection pattern. *Journal of Geophysical Research*, 87(A8), 6339–6345. <https://doi.org/10.1029/JA087iA08p06339>
- Heppner, J. P. (1977). Empirical models of high-latitude electric field. *Journal of Geophysical Research*, 82(7), 1115–1125. <https://doi.org/10.1029/JA082i007p01115>
- Heppner, J. P., & Maynard, N. C. (1987). Empirical high-latitude electric field models. *Journal of Geophysical Research*, 92(A5), 4467–4489. <https://doi.org/10.1029/JA092iA05p04467>
- King, J. H., & Papitashvili, N. E. (2005). Solar wind spatial scales in and comparisons of hourly wind and ACE plasma and magnetic field data. *Journal of Geophysical Research*, 110(A2), A02104. <https://doi.org/10.1029/2004JA010649>
- Knipp, D. J., Welliver, T., McHarg, M. G., Chun, F. K., Tobiska, W. K., & Evans, D. (2005). Climatology of extreme upper atmospheric heating events. *Advances in Space Research*, 36(12), 2506–2510. <https://doi.org/10.1016/j.asr.2004.02.019>
- Lam, M. M., Shore, R., Chisham, G., & Freeman, M. (2023). Forecast regression model of the northern Super Dual Auroral Radar Network (SuperDARN) high-latitude ionospheric plasma motion built from data interval 1997–2008 inclusive (Version 1.0) [Dataset]. NERC EDS UK Polar Data Centre. <https://doi.org/10.5285/22272b8e-1aa3-483b-9867-224fe02db4e8>
- Lewis, H. G. (2020). Understanding long-term orbital debris population dynamics. *Journal of Space Safety Engineering*, 7(3), 164–170. <https://doi.org/10.1016/j.jsse.2020.06.006>
- Lewis, H. G., Saunders, A., Swinerd, G., & Newland, R. J. (2011). Effect of thermospheric contraction on remediation of the near-Earth space debris environment. *Journal of Geophysical Research*, 116(A2), A00H08. <https://doi.org/10.1029/2011JA016482>
- Lewis, H. G., White, A. E., Crowther, R., & Stokes, H. (2012). Synergy of debris mitigation and removal. *Acta Astronautica*, 81(1), 62–68. <https://doi.org/10.1016/j.actaastro.2012.06.012>
- Liemohn, M. W., Shane, A. D., Azari, A. R., Petersen, A. K., Swiger, B. M., & Mukhopadhyay, A. (2021). RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, 218, 105624. <https://doi.org/10.1016/j.jastp.2021.105624>
- Lloyds. (2018). Realistic disaster scenarios, EM 218, v1.0.



- Lockwood, M. (2022). Solar wind–Magnetosphere coupling functions: Pitfalls, limitations, and applications. *Space Weather*, 20(2), e2021SW002989. <https://doi.org/10.1029/2021SW002989>
- Lockwood, M., & McWilliams, K. A. (2021). On optimum solar wind-magnetosphere coupling functions for transpolar voltage and planetary geomagnetic activity. *Journal of Geophysical Research: Space Physics*, 126(12), e2021JA029946. <https://doi.org/10.1029/2021JA029946>
- McComas, D. J., Elliott, H. A., Schwadron, N. A., Gosling, J. T., Skoug, R. M., & Goldstein, B. E. (2003). The three-dimensional solar wind around solar maximum. *Geophysical Research Letters*, 30(10), 1517. <https://doi.org/10.1029/2003GL017136>
- Milan, S. E., Gosling, J. S., & Hubert, B. (2012). Relationship between interplanetary parameters and the magnetopause reconnection rate quantified from observations of the expanding polar cap. *Journal of Geophysical Research*, 117(A3), A03226. <https://doi.org/10.1029/2011JA017082>
- Milan, S. E., & Grocott, A. (2021). High latitude ionospheric convection. In C. Huang, G. Lu, Y. Zhang, & L. J. Paxton (Eds.), *Ionosphere dynamics and applications*. <https://doi.org/10.1002/9781119815617.ch2>
- Nishitani, N., Ruohoniemi, J. M., Lester, M., Baker, J. B. H., Koustov, A. V., Shepherd, S. G., et al. (2019). Review of the accomplishments of mid-latitude super dual auroral radar Network (SuperDARN) HF radars. *Progress in Earth and Planetary Science*, 6(1), 27. <https://doi.org/10.1186/s40645-019-0270-5>
- Østgaard, N., Germany, G., Stadsnes, J., & Vondrak, R. R. (2002). Energy analysis of substorms based on remote sensing techniques, solar wind measurements, and geomagnetic indices. *Journal of Geophysical Research*, 107(A9), 1233. <https://doi.org/10.1029/2001JA002002>
- Papitashvili, N. E., & King, J. H. (2020a). “OMNI 1-min data” [Dataset]. NASA Space Physics Data Facility. <https://doi.org/10.48322/45bb-8792>
- Papitashvili, N. E., & King, J. H. (2020b). “OMNI daily data” [Dataset]. NASA Space Physics Data Facility. <https://doi.org/10.48322/5fmx-hv56>
- Papitashvili, V. O., & Rich, F. J. (2002). High-latitude ionospheric convection models derived from Defense Meteorological Satellite Program ion drift observations and parameterized by the interplanetary magnetic field strength and direction. *Journal of Geophysical Research*, 107(A8), S1A 17-1–S1A 17-13. <https://doi.org/10.1029/2001JA000264>
- Pettigrew, E. D., Shepherd, S. G., & Ruohoniemi, J. M. (2010). Climatological patterns of high-latitude convection in the Northern and Southern hemispheres: Dipole tilt dependencies and interhemispheric comparisons. *Journal of Geophysical Research*, 115(A7), A07305. <https://doi.org/10.1029/2009JA014956>
- Qian, L., Burns, A. G., Emery, B. A., Foster, B., Lu, G., Maute, A., et al. (2014). The NCAR TIE-GCM. In J. Huba, R. Schunk, & G. Khazanov (Eds.), *Modeling the ionosphere–thermosphere system*. <https://doi.org/10.1002/9781118704417.ch7>
- Richmond, A. D. (1995). Ionospheric electrodynamics using magnetic apex coordinates. *Journal of Geomagnetism and Geoelectricity*, 47(2), 191–212. <https://doi.org/10.5636/jgg.47.191>
- Ridley, A. J., Crowley, G., & Freitas, C. (2000). An empirical model of the ionospheric electric potential. *Geophysical Research Letters*, 27(22), 3675–3678. <https://doi.org/10.1029/1999GL011161>
- Ruohoniemi, J. M., & Greenwald, R. A. (1996). Statistical patterns of high-latitude convection obtained from Goose bay HF radar observations. *Journal of Geophysical Research*, 101(A10), 21743–21763. <https://doi.org/10.1029/96JA01584>
- Ruohoniemi, J. M., & Greenwald, R. A. (2005). Dependencies of high-latitude plasma convection: Consideration of interplanetary magnetic field, seasonal, and universal time factors in statistical patterns. *Journal of Geophysical Research*, 110(A9), A09204. <https://doi.org/10.1029/2004JA010815>
- Shore, R. M., Freeman, M., Chisham, G., Lam, M. M., & Breen, P. (2022). Dominant spatial and temporal patterns of horizontal ionospheric plasma velocity variation covering the northern polar region, from 1997.0 to 2009.0 - VERSION 2.0 (Version 2.0) [Dataset]. NERC EDS UK Polar Data Centre. <https://doi.org/10.5285/2b9f0e9f-34ec-4467-9e02-abc771070cd9>
- Shore, R. M., Freeman, M. P., & Chisham, G. (2021). Data-driven basis functions for SuperDARN ionospheric plasma flow characterization and prediction. *Journal of Geophysical Research: Space Physics*, 126(7), e2021JA029272. <https://doi.org/10.1029/2021JA029272>
- Shore, R. M., Freeman, M. P., Coxon, J. C., Thomas, E. G., Gjerloev, J. W., & Olsen, N. (2019). Spatial variation in the responses of the surface external and induced magnetic field to the solar wind. *Journal of Geophysical Research: Space Physics*, 124(7), 6195–6211. <https://doi.org/10.1029/2019JA026543>
- Shue, J.-H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., & Singer, H. J. (1997). A new functional form to study the solar wind control of the magnetopause size and shape. *Journal of Geophysical Research*, 102(A5), 9497–9511. <https://doi.org/10.1029/97JA00196>
- SuperDARN Data Analysis Working Group, Thomas, E. G., Schmidt, M. T., Bland, E. C., Burrell, A. G., Ponomarenko, P. V., et al. (2021). SuperDARN radar software toolkit (RST) 4.5 (v4.5). *Zenodo*. <https://doi.org/10.5281/zenodo.4435297>
- Sutton, E. K., Forbes, J. M., & Nerem, R. S. (2005). Global thermospheric neutral density and wind response to the severe 2003 geomagnetic storms from CHAMP accelerometer data. *Journal of Geophysical Research*, 110(A9), A09S40. <https://doi.org/10.1029/2004JA010985>
- Tenfjord, P., Østgaard, N., Snekvik, K., Laundal, K. M., Reistad, J. P., Haaland, S., & Milan, S. E. (2015). How the IMF by induces a by component in the closed magnetosphere and how it leads to asymmetric currents and convection patterns in the two hemispheres. *Journal of Geophysical Research: Space Physics*, 120(11), 9368–9384. <https://doi.org/10.1002/2015JA021579>
- Thomas, E. G., & Shepherd, S. G. (2018). Statistical patterns of ionospheric convection derived from mid-latitude, high-latitude, and polar SuperDARN HF radar observations. *Journal of Geophysical Research: Space Physics*, 123(4), 3196–3216. <https://doi.org/10.1002/2018JA025280>
- Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L. C., Guarnieri, F. L., Gopalswamy, N., Grande, M., et al. (2006). Corotating solar wind streams and recurrent geomagnetic activity: A review. *Journal of Geophysical Research*, 111(A7), A07S01. <https://doi.org/10.1029/2005JA011273>
- Undseth, M., Jolly, C., & Olivari, M. (2020). Space sustainability: The economics of space debris in perspective. In *OECD science, technology and industry policy papers* (Vol. 87). OECD Publishing. <https://doi.org/10.1787/a339de43-en>
- Vasyliūnas, V. M., & Song, P. (2005). Meaning of ionospheric Joule heating. *Journal of Geophysical Research*, 110(A2), A02301. <https://doi.org/10.1029/2004JA010615>
- Weimer, D. R. (1995). Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients. *Journal of Geophysical Research*, 100(A10), 19595–19607. <https://doi.org/10.1029/95JA01755>
- Weimer, D. R. (1996). A flexible, IMF dependent model of high-latitude electric potentials having “Space Weather” applications. *Geophysical Research Letters*, 23(18), 2549–2552. <https://doi.org/10.1029/96GL02255>
- Weimer, D. R. (2001). An improved model of ionospheric electric potentials including substorm perturbations and application to the geospace environment modeling November 24 1996 event. *Journal of Geophysical Research*, 106(A1), 407–416. <https://doi.org/10.1029/2000JA000604>
- Weimer, D. R. (2005). Improved ionospheric electrodynamic models and application to calculating Joule heating rates. *Journal of Geophysical Research*, 110(A5), A05306. <https://doi.org/10.1029/2004JA010884>
- Zhang, S.-R., Holt, J. M., & McCready, M. (2007). High latitude convection based on long-term incoherent scatter radar observations in North America. *Journal of Atmospheric and Solar-Terrestrial Physics*, 69(10–11), 1273–1291. <https://doi.org/10.1016/j.jastp.2006.08.017>