Multiresolution Modeling of High-latitude Ionospheric Electric Field Variability and Impact on Joule Heating Using SuperDARN Data

Tomoko Matsuo[1], Minjie Fan[2], Xueling Shi[3], Caleb Miller[1], J. Michael Ruohoniemi[3], Debashis Paul[2], and Thomas C M Lee[2]

[1]University of Colorado Boulder
[2]University of California Davis
[3]Virginia Polytechnic Institute and State University

November 22, 2022

Abstract

The most dynamic electromagnetic coupling between the magnetosphere and ionosphere occurs in the polar upper atmosphere. It is critical to quantify the electromagnetic energy and momentum input associated with this coupling as its impacts on the ionosphere and thermosphere system are global and major, often leading to considerable disturbances in near-Earth space environments. The current general circulation models of the upper atmosphere exhibit systematic biases that can be attributed to an inadequate representation of the Joule heating rate resulting from unaccounted stochastic fluctuations of electric fields associated with the magnetosphere-ionosphere coupling. These biases exist regardless of geomagnetic activity levels. To overcome this limitation, a new multiresolution random field modeling approach is developed, and the efficacy of the approach is demonstrated using SuperDARN data carefully curated for the study during a largely quiet 4 hours period on February 29, 2012. Regional small-scale electrostatic fields sampled at different resolutions from a probabilistic distribution of electric field variability conditioned on actual SuperDARN LOS observations exhibit considerably more localized fine-scale features in comparison to global large-scale fields modeled using the SuperDARN Assimilative Mapping procedure. The overall hemispherically integrated Joule heating rate is increased by a factor of about 1.5 due to the effect of random regional small-scale electric fields, which is close to the lower end of arbitrarily adjusted Joule heating multiplicative factor of 1.5 and 2.5 typically used in upper atmosphere general circulation models. The study represents an important step towards a data-driven ensemble modeling of magnetosphere-ionosphere-atmosphere coupling processes.

Multiresolution Modeling of High-latitude Ionospheric Electric Field Variability and Impact on Joule Heating Using SuperDARN Data

Tomoko Matsuo[1,4], Minjie Fan[2], Xueling Shi[3], Caleb Miller[4], J. Michael Ruohoniemi[3], Debashis Paul[2], Thomas C. M. Lee[3]

[1]Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder
[2]Department of Statistics, University of California Davis
[3]Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University
[4]Department of Applied Mathematics, University of Colorado Boulder

Key Points:

- A new multiresolution modeling of high-latitude ionospheric electric field variability is developed.
- The efficacy of the approach is demonstrated using SuperDARN LOS plasma drift velocity data.
- The approach can help rectify the underestimation of the Joule heating rate in the current models.

Corresponding author: T. Matsuo, `tomoko.matsuo@colorado.edu`

Abstract

The most dynamic electromagnetic coupling between the magnetosphere and ionosphere occurs in the polar upper atmosphere. It is critical to quantify the electromagnetic energy and momentum input associated with this coupling as its impacts on the ionosphere and thermosphere system are global and major, often leading to considerable disturbances in near-Earth space environments. The current general circulation models of the upper atmosphere exhibit systematic biases that can be attributed to an inadequate representation of the Joule heating rate resulting from unaccounted stochastic fluctuations of electric fields associated with the magnetosphere-ionosphere coupling. These biases exist regardless of geomagnetic activity levels. To overcome this limitation, a new multiresolution random field modeling approach is developed, and the efficacy of the approach is demonstrated using SuperDARN data carefully curated for the study during a largely quiet 4 hours period on February 29, 2012. Regional small-scale electrostatic fields sampled at different resolutions from a probabilistic distribution of electric field variability conditioned on actual SuperDARN LOS observations exhibit considerably more localized fine-scale features in comparison to global large-scale fields modeled using the SuperDARN Assimilative Mapping procedure. The overall hemispherically integrated Joule heating rate is increased by a factor of about 1.5 due to the effect of random regional small-scale electric fields, which is close to the lower end of arbitrarily adjusted Joule heating multiplicative factor of 1.5 and 2.5 typically used in upper atmosphere general circulation models. The study represents an important step towards a data-driven ensemble modeling of magnetosphere-ionosphere-atmosphere coupling processes.

1 Introduction

The most dynamic electromagnetic coupling between the magnetosphere and ionosphere occurs in the Earth's polar upper atmosphere. In particular, collisions between neutrals and ions drifting under the effect of the elevated high-latitude ionospheric electric field are a major source of heating and momentum transfer, making a global impact on the upper atmosphere. The resulting energy and momentum deposition leads to the acceleration of neutral winds and Joule dissipation, triggering dramatic global upper atmosphere responses, e.g., global temperature and neutral mass density enhancements, pole-to-equator general circulation, and atmospheric traveling disturbances (e.g., Schunk, 2014; Fuller-Rowell, 2014; Burns et al., 2014). Practical effects include altered drag force on low-Earth-orbit satellites and debris by sudden changes in neutral mass density, aggravating our ability to track these objects to mitigate potential collisions; radio wave propagating disruption affected by ionospheric density changes, deteriorating reliability of communication, navigation and positioning systems; and geomagnetically induced currents in the ground resulting from intensified ionospheric currents, affecting power transmission systems, oil and gas pipelines, railway systems, and any other extended ground-based conductor systems (e.g., Marcos et al., 2010; Groves & Carrano, 2016; Pulkkinen et al., 2017). Accurate knowledge of this energy and momentum source in the polar ionosphere is therefore of great scientific interest and has important economic and societal benefits.

The current general circulation models of the upper atmosphere exhibit systematic biases that can be attributed to the underestimation of the high-latitude energy sources, likely resulting from an inadequate representation of the Joule heating rate. These biases exist regardless of geomagnetic activity levels. The Joule heating rate is proportional to the square of the electric field magnitude and scales linearly with the Pedersen conductivity. Both of these ionospheric electrodynamics state variables are highly variable and heavily influenced by magnetosphere-ionosphere coupling processes that are not usually self-consistently solved in the upper atmosphere general circulation models and thus have to be empirically parameterized and/or specified as boundary conditions. The empirical models of high-latitude ionospheric plasma convection designed to characterize the climatological behavior of the global large-scale electric fields are not suited to representing highly variable localized multi-scale electric fields and result in residual fields with a magnitude as large as the modeled global fields themselves (e.g., Codrescu et al., 2000; Matsuo et al., 2002, 2003; Cousins & Shepherd, 2012). Even with data assimilative procedures, the instantaneous states of the localized electric fields on scales smaller than 500 km and 5 minutes associated with highly transient and regional magnetosphere-ionosphere coupling processes are difficult to capture (Matsuo et al., 2005; Matsuo & Richmond, 2008). As pointed out originally by Codrescu et al. (1995) and elaborated in later work (Codrescu et al., 2000; Matsuo & Richmond, 2008; Deng et al., 2009; Zhu et al., 2018), the underrepresented electric field variability in the upper atmosphere general circulation models is considered as one of the primary causes of the underestimation of Joule heating rate.

The volume integrated Joule heating rate is given as

$$Q_J = \iiint_V \sigma_p (\mathbf{E} + \mathbf{U} \times \mathbf{B})^2 \, dV$$

where σ_p is the Pedersen conductivity which specifies the conductivity associated with ionospheric electric currents that flow perpendicular to the geomagnetic field $\mathbf{B}$ and parallel to the electric field defined in the reference frame moving at the velocity $\mathbf{U}$ (Jackson, 1999). Note that $\mathbf{E}$ is the electric field in the Earth frame of reference and essentially electrostatic on time scales longer than tens of seconds and in the bottomside ionosphere where neutral species predominate over plasma, $\mathbf{U}$ is approximately equal to the neutral wind velocity (Kelly, 2009). $\mathbf{U}\times\mathbf{B}$ thus represents the dynamo fields resulting from an electromotive force induced by the neutral wind traversing the geomagnetic field. Because of very high electrical conductivity in the direction of $\mathbf{B}$, a geomagnetic field line is effectively equipotential where it traverses the ionosphere and therefore $\mathbf{E}$ is nearly constant with altitude along the direction of the field line which is nearly radial at high latitudes. Note that for the sake of simplicity the radial component of the electric field is ignored from discussion in this paper. When the effect of $\mathbf{U}\times\mathbf{B}$ is small, Q_J can be approximated using the height integrated Pedersen conductivity $\Sigma_p (= \int \sigma_p \, dr)$ as

$$Q_J \approx \iiint_V \sigma_p \mathbf{E}^2 \, dV = \iint_A \Sigma_p(\theta, \phi) \mathbf{E}^2(\theta, \phi) \, d\theta d\phi \tag{1}$$

where θ is the polar angle (i.e., magnetic co-latitude) and ϕ is the azimuth angle (i.e., magnetic local time (MLT)). For simplicity, geomagnetic fields are here assumed strictly radial. Let us suppose that $\mathbf{E}$ can be decomposed into *global large-scale* electric fields and *regional small-scale* electric fields as

$$\mathbf{E}(\theta, \phi, t) = \overline{\mathbf{E}}(\theta, \phi, t) + \mathbf{E}'(\theta, \phi) \ , \tag{2}$$

and $\overline{\mathbf{E}}$ represents time-dependent *mean* vector fields and $\mathbf{E}'$ represents stochastic or *random* vector fields that belong to a certain probabilistic distribution. It is important to note that specific instances of $\mathbf{E}'$ are different as $\mathbf{E}'$ are random fields but its statistical characteristics of their randomness are assumed to be temporally stationary, thus in Equation (2) a dependence on t is dropped. (Note that this assumption is made due to the necessity to aggregate data over time in the current study and should be relaxed in the future as discussed later.) It is easy to see the underestimation of the Joule heating rate could result from not accounting for effects of $\mathbf{E}'$, which can be as large as $\overline{\mathbf{E}}$ at times, in the upper atmosphere general circulation models.

Additional sources of uncertainty in determining the Joule heating rate include neutral winds $\mathbf{U}$ and Pedersen conductivity σ_p. The contribution of dynamo fields $\mathbf{U}\times\mathbf{B}$ to Q_J can be as large as 30% especially when neutral winds are driven by elevated ionospheric plasma convection during geomagnetic storms (e.g., Lu et al., 1995; Ridley et al., 2003; Sangalli et al., 2009). Depending on the direction of the neutral wind, the dynamo field effect can increase or decrease the total Joule heating rate. For instance, Lu et al. (1995) found the neutral winds have approximately a 28% negative effect on Joule heating rate on average for the 2-day geomagnetically disturbed period investigated. It is also important to note that $\mathbf{U}$ is not constant with altitude, thus requiring knowledge of vertical distributions of both the neutral wind and Pedersen conductivity when computing Q_J (Thayer, 1998). In addition, ionospheric conductivity varies considerably due to ionization of the neutral species by solar extreme ultraviolet radiations and auroral energetic particle precipitations. The effects of auroral ionization can be extremely localized and transient, which are difficult to characterize with the currently available auroral models (e.g., Newell & Wing, 2009). Furthermore, Dimant & Oppenheim (2011) have pointed out that during geomagnetically disturbed conditions the Pedersen conductivity can be enhanced considerably due to strong anomalous electron heating and nonlinear electric currents resulting from the Farley-Buneman instability (Farley, 1963; Buneman, 1963). Part of the instability effect was incorporated into a recent upper atmospheric general circulation modeling study by Liu et al. (2016). In spite of recent progress in modeling, considerable uncertainty still remains in representing all physical processes responsible for Joule heating in current general circulation modeling.

Spatial and temporal coherence and other properties of randomness of the electric field variability affect the estimate of Joule heating rate as they control how effectively momentum and energy are transferred from ionospheric plasma to neutral species. Matsuo & Richmond (2008) demonstrated this effect using ensemble modeling and Gaussian random fields generated with the space-time covariance model derived from DE-2 observations in Matsuo et al. (2002, 2005). If spatiotemporal coherence is taken into account when incorporating the effects of electric field variability into an upper atmospheric general circulation model, electric field variability becomes more effective in influencing the neutral winds and thus affecting the overall Joule heating rate. The analysis of Super Dual Auroral Radar Network (SuperDARN) plasma drift measurements by Cousins & Shepherd (2012) has furthermore revealed scale-dependent non-Gaussian probabilistic behaviors of the electric field variability. The observed localized transient characteristics of electric field variability are difficult to model using currently available standard statistical inferential frameworks. In response to the need for a new framework, Fan et al. (2018) have developed a multiresolution non-Gaussian random field model by using a class of specialized needlet basis functions (Marinucci & Peccati, 2011) that has all the desired properties (including smoothness, spatial and frequency localization, frame properties), which have enabled for flexible multiresolution reconstruction of scalar electrostatic potential fields on the sphere. By using the Lyon-Fedder-Mobarry (LFM) magnetosphere-ionosphere coupled model simulation results (Wiltberger et al., 2016), Fan et al. (2018) have furthermore demonstrated a measurable impact on the Joule heating rate.

By building on the statistical inferential framework developed by Fan et al. (2018), the objective of this paper is to characterise the electric field variability as multiresolution non-Gaussian random vector fields from actual SuperDARN observations, and to evaluate its impact on the Joule heating rate. The novel elements of the data analysis method and modeling technique described in Section 3 are as follows. The work of Fan et al. (2018) is extended to vector fields in this study. This is important as existing multiresolution bases for dealing with vector fields (such as vector spherical harmonics) do not have the spatial localization property and hence are not appropriate for describing features that are spatially localized. Since needlets can be represented in terms of spherical harmonics, in particular through Legendre polynomials, the surface gradient and curl operators can be applied to them to yield vectorial needlets that inherit the spatial compactness, facilitating flexible, multiresolution representations of the curl-free multi-scale electrostatic fields. Furthermore, the adaptive Markov-Chain Monte Carlo (MCMC) estimation approach developed in Fan et al. (2018) is used to characterize non-Gaussian random electric fields from SuperDARN observations.

An additional notable element of the study is a special pre-processing of a standard SuperDARN FITACF data designed for the needlet-based approach to modeling electric field variability as described in Section 2). Although similar approaches have been used in the past (Ruohoniemi & Baker, 1998), this is the first consolidated attempt to extract randomness information from the FITACF data product, and can serve as a foundation for follow-on future studies with more data and validation with independent data. The SuperDARN data from a 4-hour period on February 29, 2012 are selected for the study with a number of considerations including data coverage and consistency in geophysical conditions, and processed as described in Section 2. As demonstrated in Section 4, and discussed and summarized in Section 5, the study is an important cross-disciplinary research and development effort that enables a more comprehensive data-driven approach to modeling of magnetosphere-ionosphere-atmosphere coupling processes.

2 SuperDARN Data

The SuperDARN is an international network consisting of more than 30 low-power HF (3-30 MHz) coherent scatter radars at middle to polar latitudes in both hemispheres that look into Earth's upper atmosphere and ionosphere (Chisham et al., 2007; Nishitani et al., 2019). The radars measure the line-of-sight (LOS) component of the F-region ionospheric plasma drift velocity when decameter-scale electron density irregularities are present and oriented favorably to produce backscatter. The irregularity motion here is due to $\mathbf{E}\times\mathbf{B}$ drift. Normally, the SuperDARN radars are scheduled for 1-minute or 2-minute azimuthal sweeps in the normal mode. The step in azimuth between adjacent beams is 3.24$^{\rm o}$ and the range resolution is 45 km. This study uses LOS plasma drift velocity ($v_{\rm LOS}$) from SuperDARN radars operating in the *normal* scan mode from the northern hemisphere over the four-hour period from 0000 to 0400 Universal Time (UT) on February 29, 2012. The location of these radars and their field-of-views (FOVs) as well as the data coverage are shown in Figure 1. This is a largely quiet period during the rising phase of Solar Cycle 24 (F107=100.1) with a minor geomagnetic activity of the Kp index of 3, the minimum Dst index of about -30 nT, and the Auroral Electrojet (AE) index ranging from 100 to 420 nT peaking at 0250 UT.

$v_{\rm LOS}(\theta,\phi)$ is related linearly to the electrostatic potential $\Phi_{\rm E}(\theta,\phi)$, where $\mathbf{E}(\theta,\phi) = -\nabla\Phi_{\rm E}(\theta,\phi)$, as described in Section 2.1. Both *global large-scale mean* electric fields $\overline{\mathbf{E}}$ and *regional small-scale random* electric fields $\mathbf{E}'$, expressed in terms of the electrostatic potential, are here estimated from SuperDARN LOS velocity data. Among multiple types of SuperDARN LOS data products made available by the SuperDARN consortium for different scientific applications, the FITACF data product is used for estimating $\mathbf{E}'$ after the pre-processing described in Section 2.2. The GRID data product that itself is a derived product of FITACF data is used for estimating $\overline{\mathbf{E}}$ using the SuperDARN Assimilative Mapping (SAM) procedure (Cousins et al., 2013b) as described in Section 2.3. The SAM analysis is conducted every two minutes using the GRID data, and global large-scale fields' contribution to the LOS plasma drift velocity is subtracted from FITACF LOS velocity data (see Section 3.2). Note that the FITACF data is aggregated over the four hours for needlet-based analysis as described in Section 3.

2.1 SuperDARN Line-of-Sight (LOS) Plasma Drift Velocity

Assuming that the geomagnetic field is strictly radial (i.e., $\mathbf{B} = -B\hat{\mathbf{r}}$), the electric field $\mathbf{E}$ is expressed as

$$\mathbf{E} = -\nabla\Phi_{\rm E} = -\frac{1}{R}\frac{\partial\Phi_{\rm E}}{\partial\theta}\hat{\boldsymbol{\theta}} - \frac{1}{R}\frac{1}{\sin\theta}\frac{\partial\Phi_{\rm E}}{\partial\phi}\hat{\boldsymbol{\phi}}, \tag{3}$$

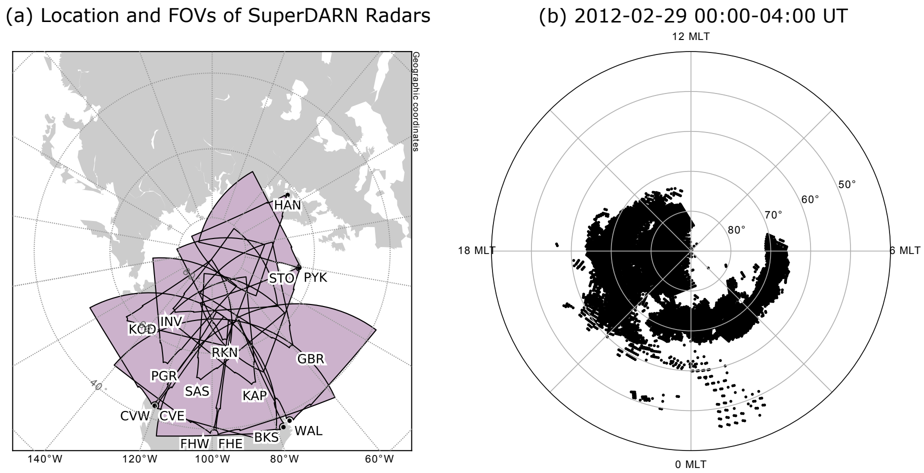

Figure 1. (a) The location of SuperDARN radars in the northern hemisphere used in this study and their FOVs; (b) scatter plot showing the SuperDARN measurement coverage during the four hour interval on February 29, 2012.

where R is the radius of the ionosphere. The plasma drift velocity $\mathbf{v} = \mathbf{E} \times \mathbf{B}/B^2$ is thus given as

$$\mathbf{v} = -\frac{1}{BR}\frac{\partial \Phi_{\mathrm{E}}}{\partial \theta}\hat{\boldsymbol{\phi}} + \frac{1}{BR}\frac{1}{\sin\theta}\frac{\partial \Phi_{\mathrm{E}}}{\partial \phi}\hat{\boldsymbol{\theta}}\ ,$$

where $B > 0$ is a magnitude of the geomagnetic field that varies over the sphere. The LOS component of the velocity $\mathbf{v}$, which is $v_{\mathrm{LOS}} = \mathbf{v} \cdot \mathbf{k}_{\mathrm{LOS}}$ where $\mathbf{k}_{\mathrm{LOS}} = k_\theta \hat{\boldsymbol{\theta}} + k_\phi \hat{\boldsymbol{\phi}}$ is a unit vector that gives the direction of the line of sight, becomes

$$v_{\mathrm{LOS}} = \frac{k_\theta}{BR}\frac{1}{\sin\theta}\frac{\partial \Phi_{\mathrm{E}}}{\partial \phi} - \frac{k_\phi}{BR}\frac{\partial \Phi_{\mathrm{E}}}{\partial \theta}. \tag{4}$$

The SuperDARN data sets include values of k_θ and k_ϕ for each v_{LOS} data points.

2.2 Pre-processing of SuperDARN FITACF Data

In order to estimate velocity and other parameters from autocorrelation functions calculated from the radar backscatter returns for each beam and range gate of a radar, the FITACF fitting routine is applied to estimate Doppler velocity, spectral width, and backscatter power (Ribeiro et al., 2013). Similar to the criteria used in Cousins & Shepherd (2012), several steps have been taken to ensure that only high-quality LOS plasma velocity measurements from the F-region ionosphere are included in the analysis. Primary selection criteria include (i) the slant range greater than 600 km, (ii) the backscatter power or signal-to-noise ratio (SNR) greater than 8 dB, and (iii) the velocity error less than 100 m/s. In addition, ground scatters are carefully excluded based on the ground scatter flag from the standard SuperDARN data processing and spectral width and velocity magnitude values. Incidental outlier data with very large velocity values can still be commonly found after the above processing, necessitating further processing on data. These outliers are excluded by keeping only data from each SuperDARN radar beam and gate cell when at least 25% of good samplings are present within a given 10-minute interval, which usually includes 5-10 scans. Note that a complete cycle through its full set of beam-azimuth settings defines a radar scan which usually takes 1-2 minutes. A good sample is specifically defined as a weight value $\mathcal{W}$ equal or greater than 1.5, which is computed using a switch function $S(bm, gt)$, where bm is the beam number and gt is the gate number, as follows:

$$\mathcal{W}(bm, gt) = \mathcal{S}(bm, gt) + 0.5 * (\mathcal{S}(bm-1, gt) + \mathcal{S}(bm+1, gt) + \mathcal{S}(bm, gt-1) + \mathcal{S}(bm, gt+1)).$$

Here $\mathcal{S}(bm, gt) = 1$ when good data points exit in the beam-gate cell, and $\mathcal{S}(bm, gt) = 0$ when no good data points exit in the beam-gate cell. The median velocity and standard deviation of velocities in each beam-gate cell are then computed with at least 25% of good samples from all the scans within each 10-minute interval. The standard deviation computed with a temporal resolution of 10 minutes provides the sense of (precision) errors. This pre-processing helps exclude outliers, poor quality data, and data with near-range meteor scatter, most E-region scatter and ground scatter from the standard FITACF data product, and only clean SuperDARN LOS plasma drift velocity data are used for the needlet-based analysis presented in Section 3.

2.3 Estimation of Global Large-Scale Electric Fields by SuperDARN Assimilative Mapping (SAM)

The distribution of global large-scale ionospheric convective electric fields $\overline{\mathbf{E}}(\theta, \phi)$ is determined at 2-minute cadence from the SuperDARN GRID data over 4-minute windows using the SAM procedure (Cousins et al., 2013b). The SAM uses a set of the spherical cap harmonics functions developed by Richmond & Kamide (1988), with the spherical harmonics of the order 12 and non-integer degrees of 72.6 for the 0^{th} order zonally symmetric harmonic functions that give the effective resolution of 15^o longitude and 2.5^o latitude in terms of the Nyquist sampling rate. The SAM solves a Bayesian spatial statistical prediction problem for ionospheric convective electric fields just as the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) (Richmond & Kamide, 1988), and computes the posterior mean given the prior mean convective electric fields specified by Cousins & Shepherd (2010). A major advantage of the SAM over the AMIE is the use of prior model error covariance developed from a large volume of SuperDARN data in Cousins et al. (2013a) for the prior model of Cousins & Shepherd (2010). The LOS plasma drift velocity due to these global large-scale electric fields is computed according to Equation (4) and subtracted from the pre-processed LOS velocity data explained in Section 2.2.

Due to the use of global spherical cap harmonics functions, with a limited resolution, in the SAM, it is sufficient to use the GRID data which provide the standardized LOS velocity values on an equal-area grid over a fixed period of time of 1 or 2 minutes, rather than the FITACF data which contain the LOS velocity measurements recorded by individual radars as a function of beam-azimuth range-gate setting. GRID data is a highly processed data product derived from FITACF data. A median filter is first applied to the individual radar scan data to remove noise to calculate the median LOS velocities of a particular scan. The LOS vectors are then mapped within the cells of an equal-area grid, which is defined in the geomagnetic coordinates system with each cell measuring 1^o in latitude, to eliminate biases that would derive from the much denser sampling over nearer radar range gates. The vectors contributed by a radar to a particular cell are averaged over a fixed period of time to obtain the GRID LOS data product. More details of GRID data processing can be found in Section 3 of Ruohoniemi & Baker (1998).

3 Needlet-based Approach to Modeling Electric Field Variability

The novel element of the statistical modeling approach presented here is the use of a multiresolution tight frame called needlets (Marinucci & Peccati, 2011) to represent stochastic fluctuations in the electric field vectors. In the same way wavelets facilitate analysis of transient and localized signals, needlets enable us to represent spatially localized features of the observed electric field variability in functions defined over a spherical domain. Needlets have been shown to be more efficient than spherical harmonics in representing spatially localized features on the sphere as linear combinations of spherical harmonics, through a construction involving Legendre polynomials (Scott, 2011). Furthermore, the surface gradient operators can be applied on them, thus facilitating multiresolution representations of the curl-free multi-scale electrostatic potential fields. In Section 3.1, spherical needlets used in this study are briefly defined, and Fan et al. (2018) should be referenced for more details.

3.1 Multiresolution Tight Frame: Spherical Needlets

Specifically, a needlet function at scale j and location k, $\psi_{jk}(\mathbf{s})$, evaluated at a point $\mathbf{s}$ on the unit sphere takes the following form:

$$\psi_{jk}(\mathbf{s}) = \sqrt{\lambda_{jk}} \sum_{l=\lceil M^{j-1}\rceil}^{\lfloor M^{j+1}\rfloor} b\left(\frac{l}{M^j}\right) \sum_{m=-l}^{l} Y_{lm}(\zeta_{jk})\overline{Y}_{lm}(\mathbf{s}) = \sqrt{\lambda_{jk}} \sum_{l=\lceil M^{j-1}\rceil}^{\lfloor M^{j+1}\rfloor} b\left(\frac{l}{M^j}\right) \frac{2l+1}{4\pi} P_l(\langle\zeta_{jk}, \mathbf{s}\rangle), \tag{5}$$

where the nonnegative function $b(\cdot)$ is bandlimited and enables a frequency tiling, $M > 1$ controls the window size of the frequency tiling, $(\zeta_{jk}, \lambda_{jk})$ are quadrature (location, weight) pairs for scale j and location k, Y_{lm}'s are the standard orthonormal, complex-valued spherical harmonics basis functions corresponding to frequency (degree) index l and phase (order) index m, and P_l is the associate Legendre polynomial of degree l. The function b is positive on the interval (M^{-1}, M) and satisfies the *resolution of identity* condition $\sum_{j=0}^{\infty} b^2(\xi/M^j) = 1$ for $\xi > 0$. From Equation (5), it is evident that needlets ψ_{jk}'s are bandlimited over spherical frequencies ranging from integer index l greater than or equal to M^{j-1} to l less than or equal to M^{j+1}. Hereafter, we choose $M = 2$ following the prior work [e.g., Fan et al., 2018]. Note that because of the linear representation of needlets in the spherical harmonic basis, needlet coefficients of a scalar function can be obtained from the spherical harmonics coefficients of the function through a linear transformation, since for any L^2 (quadratically integrable) function f on ordinary sphere $\mathbb{S}^2$,

$$\langle f, \psi_{jk}\rangle = \sqrt{\lambda_{jk}} \sum_{l=\lceil M^{j-1}\rceil}^{\lfloor M^{j+1}\rfloor} b\left(\frac{l}{M^j}\right) \sum_{m=-l}^{l} Y_{lm}(\zeta_{jk})\langle f, Y_{lm}\rangle ,$$

where $\langle f, \psi_{jk}\rangle$ and $\langle f, Y_{lm}\rangle$ denote the needlet and spherical harmonics coefficients, respectively.

3.2 Needlet-based Random Electric Fields Model

Suppose that there is a total of p_d SuperDARN LOS plasma velocity data points at locations (θ_i, ϕ_i), $i = 1, \cdots, p_d$, in the high-latitude region of the northern hemisphere. Note the data points shown in Figure 1(b) are down-sampled as explained later in Section 3.4 before being used for the needlet model estimation. Since these data contain the observational noise, they are modeled by the following statistical model

$$v_{\mathrm{LOS}}^{\mathrm{fitacf}}(\theta_i, \phi_i) = v_{\mathrm{LOS}}(\theta_i, \phi_i) + \epsilon_i,$$

where $v_{\mathrm{LOS}}^{\mathrm{fitacf}}$ represents the observational data, v_{LOS} represents the underlying true velocity value, and $\epsilon_i \sim \mathcal{N}(0, \tau_i^2)$ is the observation noise or error with standard deviation τ_i. (Note that $\mathcal{N}(\alpha, \beta)$ represents the normal distribution with a mean parameter α and a variance parameter β.) For simplicity, τ_i is henceforth assumed independent of the location (i.e., $\tau_i = \tau$, $i = 1, \cdots, p_d$). As described later in Section 3.3, τ^2 is one of the statistical model parameters to be estimated from SuperDARN LOS velocity data.

According to Equation (4), the velocity field v_{LOS} can be derived from the electrostatic potential Φ_{E} by applying the differential operators, and the electrostatic potential Φ_{E} can be decomposed into two components: *global large-scale* and spatially localized *regional small-scale* components, $\Phi_{\mathrm{E,g}}$ and $\Phi_{\mathrm{E,r}}$, which respectively correspond to

$\overline{\mathbf{E}} = -\nabla\Phi_{\mathrm{E,g}}$ and $\mathbf{E}' = -\nabla\Phi_{\mathrm{E,r}}$. Therefore, v_{LOS} can also be decomposed into two components accordingly, i.e.,

$$v_{\mathrm{LOS}} = v_{\mathrm{LOS,g}} + v_{\mathrm{LOS,r}}.$$

The SAM procedure described in Section 2.3 is well suited to estimate $v_{\mathrm{LOS,g}}$ from the SuperDARN LOS plasma velocity data $v_{\mathrm{LOS}}^{\mathrm{fitacf}}$. We subtract the fitted $v_{\mathrm{LOS,g}}$, denoted by $\hat{v}_{\mathrm{LOS,g}}$, from $v_{\mathrm{LOS}}^{\mathrm{fitacf}}$, and obtain

$$v_{\mathrm{LOS}}^{\mathrm{fitacf}}(\theta_i, \phi_i) - \hat{v}_{\mathrm{LOS,g}}(\theta_i, \phi_i) \approx v_{\mathrm{LOS,r}}(\theta_i, \phi_i) + \epsilon_i.$$

$v_{\mathrm{LOS,r}}(\theta_i, \phi_i)$ is precisely what is modeled in terms of $\Phi_{\mathrm{E,r}}$ by spherical needlets ψ_{jk}'s.

Since SuperDARN LOS plasma velocity data points are restricted to the high-latitude region, the data points are stretched to the entire sphere by mapping observation location points (θ_i, ϕ_i) to $(4\theta_i, \phi_i)$. Since the magnitude of $v_{\mathrm{LOS,r}}$ has a strong dependency on the latitude, a variance profile, i.e., the variance of the observed LOS velocity field as a function of the latitudinal location, is introduced as a function of co-latitude, and $\Phi_{\mathrm{E,r}}$ is modeled by the product of the variance profile and a linear combination of spherical needlets as follows:

$$\Phi_{\mathrm{E,r}}(\theta_i, \phi_i) = g(4\theta_i) \sum_{j=J_0}^{J} \sum_{k=1}^{p_j} c_{jk}\psi_{jk}(4\theta_i, \phi_i), \theta_i \in [0, \pi/4], \phi_i \in [0, 2\pi], \tag{6}$$

where g is the variance profile function, and c_{jk} are needlet coefficients, which are random variables. As in Fan et al. (2018), it is assumed that c_{jk}'s are distributed as scale multiples of a t-distribution, i.e. $c_{jk} \sim \sigma_j t(\nu)$, where $t(\nu)$ denotes the t-distribution with ν degrees of freedom. The t-distribution has heavier tails in comparison to the normal distribution. $\nu = 3$ is used for this study following Fan et al. (2018) wherein $\nu = 3$ was chosen among 2.5, 3, and 4 in their applications to the LFM model output as it yielded the best predictive performance for simulated data. Note that with infinite degrees of freedom, the t-distribution approaches to the standard normal distribution. The assumed distribution characterizes both scale-dependent variations and spatially localized features of the electric field variability. Moreover, c_{jk}'s are assumed to be statistically independent for simplicity. Due to the non-Gaussianity of c_{jk} and the spatial localization of ψ_{jk}, the resulting field is also non-Gaussian.

The variance profile function g is assumed to have the representation $g(\cdot) = \exp(\mathbf{h}^{\mathrm{T}}(\cdot)\boldsymbol{\eta})$, where $\mathbf{h}(\cdot)$ are the basis functions specified as cubic B-splines due to their numerical stability. To avoid the non-identifiability issue, the first B-spline is dropped in the formula. As described later in Section 3.3, the B-spline weights $\boldsymbol{\eta}$ that control the variance profile g and the t-distribution population parameters σ_j^2, $j = J_0, \cdots, J$, that determines a probabilistic distribution of needlet coefficients, given a value of $\nu = 3$, are estimated from SuperDARN LOS data.

By the chain rule, we know that

$$\frac{\partial\Phi_{\mathrm{E,r}}}{\partial\theta} = 4\left(\left.\frac{\partial g}{\partial\theta'}\right|_{\theta'=4\theta} \sum_{j,k} c_{jk}\psi_{jk}(4\theta, \phi) + g(4\theta)\sum_{j,k} c_{jk} \left.\frac{\partial\psi_{jk}}{\partial\theta'}\right|_{\theta'=4\theta} \right),$$

and

$$\frac{1}{\sin\theta}\frac{\partial\Phi_{\mathrm{E,r}}}{\partial\phi} = \frac{\sin\theta'}{\sin\theta} g(4\theta)\sum_{j,k} c_{jk}\frac{1}{\sin\theta'}\frac{\partial\psi_{jk}}{\partial\phi},$$

where $\theta' = 4\theta$. Plugging these into (4), we have

$$v_{\mathrm{LOS,s}} = \frac{1}{BR} g \sum_{j,k} c_{jk} \underbrace{\left(k_\theta \frac{\sin\theta'}{\sin\theta}\frac{1}{\sin\theta'}\frac{\partial\psi_{jk}}{\partial\phi} - 4k_\phi \frac{\partial\psi_{jk}}{\partial\theta'} \right)}_{\psi_{jk}^{(1)}} - \frac{1}{BR}\frac{\partial g}{\partial\theta'}\sum_{j,k} c_{jk} \underbrace{4k_\phi\psi_{jk}}_{\psi_{jk}^{(2)}}.$$

In this way, the spherical needlets are transformed into two new sets of basis functions $\psi_{jk}^{(n)}, n = 1, 2$ in the domain of line-of-sight velocities. Recall that

$$\psi_{jk}(\theta', \phi) = \sqrt{\lambda_{jk}} \sum_l b\left(\frac{l}{M^j}\right) \frac{2l+1}{4\pi} P_l(x_{jk} \sin\theta' \cos\phi + y_{jk} \sin\theta' \sin\phi + z_{jk} \cos\theta'),$$

where $(x_{jk}, y_{jk}, z_{jk}) \in \mathbb{S}^2$ is the centroid of the needlet ψ_{jk}. Then

$$\frac{\partial \psi_{jk}}{\partial \theta'} = \sqrt{\lambda_{jk}}(x_{jk} \cos\theta' \cos\phi + y_{jk} \cos\theta' \sin\phi - z_{jk} \sin\theta') \sum_l b\left(\frac{l}{M^j}\right) \frac{2l+1}{4\pi} \frac{dP_l(u)}{du}\bigg|_{u=u'},$$

and

$$\frac{1}{\sin\theta'} \frac{\partial \psi_{jk}}{\partial \phi} = \sqrt{\lambda_{jk}}(-x_{jk} \sin\phi + y_{jk} \cos\phi) \sum_l b\left(\frac{l}{M^j}\right) \frac{2l+1}{4\pi} \frac{dP_l(u)}{du}\bigg|_{u=u'},$$

where $u' = x_{jk} \sin\theta' \cos\phi + y_{jk} \sin\theta' \sin\phi + z_{jk} \cos\theta'$. Note that $dP_l(u)/du$ can be efficiently computed by using a recursive formula.

The statistical model for v_{LOS} can be summarized as the following matrix-vector form:

$$\mathbf{v}_{\mathrm{LOS}}^d - \hat{\mathbf{v}}_{\mathrm{LOS,g}} \approx (BR)^{-1}(\mathbf{G}_1\mathbf{A}_1 + \mathbf{G}_2\mathbf{A}_2)\mathbf{c} + \boldsymbol{\epsilon} = (BR)^{-1}\mathbf{G}\mathbf{A}\mathbf{c} + \boldsymbol{\epsilon},$$

where $\mathbf{G}_1 = \mathrm{diag}\{g(4\theta_i), i = 1, \cdots, p_d\}$, $\mathbf{G}_2 = \mathrm{diag}\left\{-\frac{\partial g}{\partial \theta'}\big|_{\theta'=4\theta_i}, i = 1, \cdots, p_d\right\}$, $\mathbf{A}_1$ and $\mathbf{A}_2$ are the design matrices constructed by the new basis functions $\psi_{jk}^{(n)}, n = 1, 2$, respectively, $\mathbf{G} = [\mathbf{G}_1; \mathbf{G}_2]$ and $\mathbf{A} = [\mathbf{A}_1; \mathbf{A}_2]$. For convenience, we hereafter use $\mathbf{D}$ to denote $(BR)^{-1}\mathbf{G}\mathbf{A}$ and $\mathbf{z}$ to stand for $\mathbf{v}_{\mathrm{LOS}}^d - \hat{\mathbf{v}}_{\mathrm{LOS,g}}$ so that

$$\mathbf{z} = \mathbf{D}\mathbf{c} + \boldsymbol{\epsilon}\ .$$

3.3 Adaptive Markov Chain Monte-Carlo Estimation

This section describes how the parameters $\sigma_{J_0}^2, \cdots, \sigma_J^2$ and $\boldsymbol{\eta}$ that determine a probabilistic distribution of needlet coefficients $\mathbf{c}$ and the observational noise parameter τ are estimated from the residual SuperDARN LOS velocity data $\mathbf{z}$ using the adaptive MCMC method. These parameters, grouped here as the vector $\boldsymbol{\omega} = (\sigma_{J_0}^2, \cdots, \sigma_J^2, \tau^2, \boldsymbol{\eta})$, are assumed *a priori* independent. The prior distributions of σ_j^2 and τ^2 are the non-informative Jeffreys' priors. The prior distribution of $\boldsymbol{\eta}$ is assumed to be $\mathcal{N}(\mathbf{0}, \tau_{\boldsymbol{\eta}}^2\mathbf{I})$, where the hyperparameter $\tau_{\boldsymbol{\eta}}$ is chosen to be sufficiently large such that the prior distribution is nearly non-informative. Under these settings, the posterior distribution of the parameters can be computed by the following MCMC algorithm.

Since a t-distribution can be expressed as scale mixture of Gaussians, the probability distribution of c_{jk} can be written in a hierarchical form by introducing an auxiliary random variable V_{jk}

$$c_{jk}|V_{jk} \sim \mathcal{N}(0, V_{jk}),$$

$$V_{jk}|\nu, \sigma_j \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu\sigma_j^2}{2}\right),$$

where $\mathcal{IG}(\alpha, \beta)$ represents the inverse gamma distribution with a shape parameter α and a scale parameter β. Denote by $\mathbf{V}$ the vector consisting of the coordinates V_{jk}, and $\boldsymbol{\sigma}^2$ to be the vector comprising of $\sigma_{J_0}^2, \cdots, \sigma_J^2$. We shall employ a Gibbs sampler to obtain samples from $[\mathbf{c}, \mathbf{V}, \boldsymbol{\omega}|\mathbf{z}]$ so that the full conditional distributions of $\mathbf{c}, \mathbf{V}, \boldsymbol{\sigma}^2$ and τ^2 have closed forms. The full conditional distribution of $\mathbf{c}$ (i.e. $[\mathbf{c}|\mathbf{z}, \mathbf{V}, \boldsymbol{\omega}]$) is multivariate Gaussian, and hence sampling from it requires $\mathcal{O}(p^3)$ operations, where p is the total number of needlets. This is computationally intractable for large p. Nonetheless, numerical experiments indicate that the subblocks $\mathbf{c}_j|\mathbf{z}, \mathbf{V}, \boldsymbol{\omega}, j = J_0, \cdots, J$ are weakly correlated, where $\mathbf{c}_j = (c_{j1}, \cdots, c_{jp_j})^{\mathrm{T}}$, where p_j is the number of needlets at level j. Therefore, the sampling step for $\mathbf{c}$ is achieved by successive draws from the conditional subblocks $[\mathbf{c}_j|\mathbf{z}, \mathbf{V}, \boldsymbol{\omega}, \mathbf{c}_{-j}]$. The full conditional distribution of $\boldsymbol{\eta}$ is not available in closed form. Therefore, we sample from $[\boldsymbol{\eta}|\mathbf{z}, \mathbf{c}, \mathbf{V}, \boldsymbol{\sigma}^2, \tau^2]$ using an adaptive Metropolis algorithm (Andrieu & Thoms, 2008, Algorithm 4) and incorporate it into the Gibbs sampler.

Suppose $\mathbf{D}_j$ denotes the subblock of $\mathbf{D}$ corresponding to the j-th level of needlets, so that $\mathbf{D} = (\mathbf{D}_{J_0}, \cdots, \mathbf{D}_J)$, and $\mathbf{V}_j = (V_{j1}, \cdots, V_{jp_j})^{\mathrm{T}}$. Then the aforementioned adaptive Metropolis-within-Gibbs sampler can be summarized as follows:

1. Sample $\mathbf{c}_j$ from $[\mathbf{c}_j|\mathbf{z}, \mathbf{V}, \boldsymbol{\omega}, \mathbf{c}_{-j}] = \mathcal{N}(\hat{\boldsymbol{\mu}}_j, \widehat{\boldsymbol{\Sigma}}_j)$, where
$$\widehat{\boldsymbol{\Sigma}}_j = \left(\frac{1}{\tau^2} \mathbf{A}_j^{\mathrm{T}} \mathbf{D}^2 \mathbf{A}_j + \mathrm{diag}(\mathbf{V}_j)^{-1} \right)^{-1},$$
and
$$\hat{\boldsymbol{\mu}}_j = \frac{1}{\tau^2} \widehat{\boldsymbol{\Sigma}}_j \mathbf{A}_j^{\mathrm{T}} \mathbf{D} (\mathbf{z} - \mathbf{D}\mathbf{A}_{-j}\mathbf{c}_{-j}).$$
2. Sample $\mathbf{V}$ from $[\mathbf{V}|\mathbf{z}, \mathbf{c}, \boldsymbol{\omega}]$, where $V_{jk}|\mathbf{z}, \mathbf{c}, \boldsymbol{\omega}$ are independent and distributed as
$$\mathcal{IG}\left(\frac{\nu+1}{2}, \frac{c_{jk}^2 + \nu\sigma_j^2}{2} \right).$$
3. Sample $\boldsymbol{\sigma}^2$ from $[\boldsymbol{\sigma}^2|\mathbf{z}, \mathbf{c}, \mathbf{V}, \tau^2, \boldsymbol{\eta}]$, where $\sigma_j^2|\mathbf{z}, \mathbf{c}, \mathbf{V}, \tau^2, \boldsymbol{\eta}$ are independent and distributed as
$$\mathcal{G}\left(\frac{\nu p_j}{2}, \frac{\nu}{2} \sum_{k=1}^{p_j} \frac{1}{V_{jk}} \right),$$
where $\mathcal{G}(\alpha, \beta)$ represents the gamma distribution with a shape parameter α and a rate parameter β.
4. Sample τ^2 from
$$[\tau^2|\mathbf{z}, \mathbf{c}, \mathbf{V}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}] = \mathcal{IG}\left(\frac{p_d}{2}, \frac{(\mathbf{z} - \mathbf{D}\mathbf{A}\mathbf{c})^T (\mathbf{z} - \mathbf{D}\mathbf{A}\mathbf{c})}{2} \right).$$
5. Sample $\boldsymbol{\eta}$ using the adaptive Metropolis algorithm from
$$[\boldsymbol{\eta}|\mathbf{z}, \mathbf{c}, \mathbf{V}, \boldsymbol{\sigma}^2, \tau^2] \propto \exp\left\{ -\frac{1}{2\tau^2} (\mathbf{z} - \mathbf{D}\mathbf{A}\mathbf{c})^T (\mathbf{z} - \mathbf{D}\mathbf{A}\mathbf{c}) \right\} \exp\left\{ -\frac{1}{2\tau_{\boldsymbol{\eta}}^2} \boldsymbol{\eta}^T \boldsymbol{\eta} \right\}.$$
The proposal distribution is chosen to be
$$Q(\boldsymbol{\eta}^*|\boldsymbol{\eta}) \sim \mathcal{N}(\boldsymbol{\eta}, \gamma\boldsymbol{\Sigma}),$$
where γ is a parameter adaptively tuned with the goal of achieving the optimal acceptance rate (Gelman et al., 1996), and $\boldsymbol{\Sigma}$ is adaptively updated to approximate the covariance matrix of the full conditional distribution of $\boldsymbol{\eta}$.

3.4 Down-sampling of SuperDARN LOS Data

The SuperDARN LOS plasma velocity measurements are down-sampled before being applied to estimation of the needlet model parameters. As shown in Figure 1(b), the LOS velocity measurements are unevenly distributed over the high-latitude region of the sphere. The aim of down-sampling is to assure that the data are evenly distributed and the computational cost is manageable. To achieve the goal, we first partition the sphere into approximately equal-area regions by applying Voronoi tessellation on the sphere. The number of partitioned regions is the same as the number of data points. We then calculate the area of each region. The sampling probability of each data point is proportional to the area of its corresponding region. Roughly speaking, the larger the surface area of one region, the further the data point in the region is from the neighboring data points. Therefore, a higher probability of retaining the data point is assigned. In this way, the data points after down-sampling are approximately uniformly distributed.

3.5 Model Performance

A summary of the performance of MCMC based estimation of the needlet-based model parameters described in Sections 3.2-3.3 is given here. We set J_0 to 2 since the *global large-scale* components have already been subtracted from the SuperDARN LOS plasma velocity data. J is set to 3 given the the computational limitation of the model with too high J as well as the SuperDARN data signal-to-noise ratio, thus needlets at two resolution levels $j = 2, 3$ are used. The logarithm of the variance profile function g is represented by a linear combination of cubic B-splines with one interior knot $\pi/2$. The parameter estimates are calculated as the average of 1000 MCMC samples.

Figure 2 shows the estimated variance profile g as a function of co-latitude. The peak is around 75$^{\text{o}}$ latitude, consistent with the locations of the high LOS residual velocity within auroral oval zone. As described in Section 3.4, the needlet model is fitted to a subset of the SuperDARN LOS plasma velocity data after down-sampling. We examine the model out-of-sample prediction performance on the remaining data, which is shown as a scattered plot of LOS velocity magnitudes in Figure 3. The predicted values generally align with the observed values as the Pearson correlation coefficient between them is approximately 0.33. In terms of the magnitude, the predicted values are mostly smaller than the observed values. This can be explained by the following reasons: (i) The observed LOS velocity residuals are quite noisy. As shown in the middle plot of Figure 4, there are clearly some extreme values, not entirely captured by the model; (ii) The needlet-based model cannot represent features with resolution higher than $J = 3$ due to the computational limitation; (iii) We have assumed a simplified structure for the underlying electrostatic field, which is longitudinally (magnetic local time) symmetric with a variance profile depending on latitudes only; (iv) Beyond these, in general, the predicted values of observations under a Bayesian paradigm (or in a random effects model) tend to shrink towards zero, even when the model represents the data perfectly. However, the observed LOS velocity residuals exhibit heterogeneous with respect to both longitudes and latitudes. These are the unique challenges of modelling the LOS velocity residuals, which we will discuss further in Section 5.

4 Results

This section summarizes the results of the needlet-based approach to modeling electric field variability using SuperDARN data described in Sections 2 and 3, and demonstrates how the approach can help better represent the impact of the high-latitude ionospheric electric field variability on Joule heating rate in the upper atmosphere general circulation models. It also illustrates how uncertainty in data-driven modeling of electromagnetic coupling between the magnetosphere and ionosphere may be represented using ensembles.

4.1 Multiresolution Non-Gaussian Electric Fields

Figure 5 shows the electrostatic potential fields at different needlet resolution levels generated using the estimated needlet model parameters (e.g., $\mathbf{c}$, $\boldsymbol{\omega}$) from SuperDARN LOS observations. Note that the electrostatic potential fields shown here correspond to *regional small-scale* electric fields $\mathbf{E}' = -\nabla\Phi_{\mathrm{E,r}}$ (see Equation (6)). The top row displays the mean prediction conditional on the SuperDARN observations, and the middle and bottom rows show two instances of random samples conditional on the observations. These mean and two random instances are shown to illustrate that the electric field variability is in fact modeled as random fields that belong to a certain probability distribution that is conditional on the SuperDARN observations in contrast to the past studies wherein the sample mean and standard deviations of observations have been often used. These two instances are part of a 1000-member ensemble set $\{\Phi_{\mathrm{E,r}}^{(1)}, \Phi_{\mathrm{E,r}}^{(2)}, \cdots, \Phi_{\mathrm{E,r}}^{(1000)}\}$ generated from 1000 independent random draws, which are being used for the Joule heat-

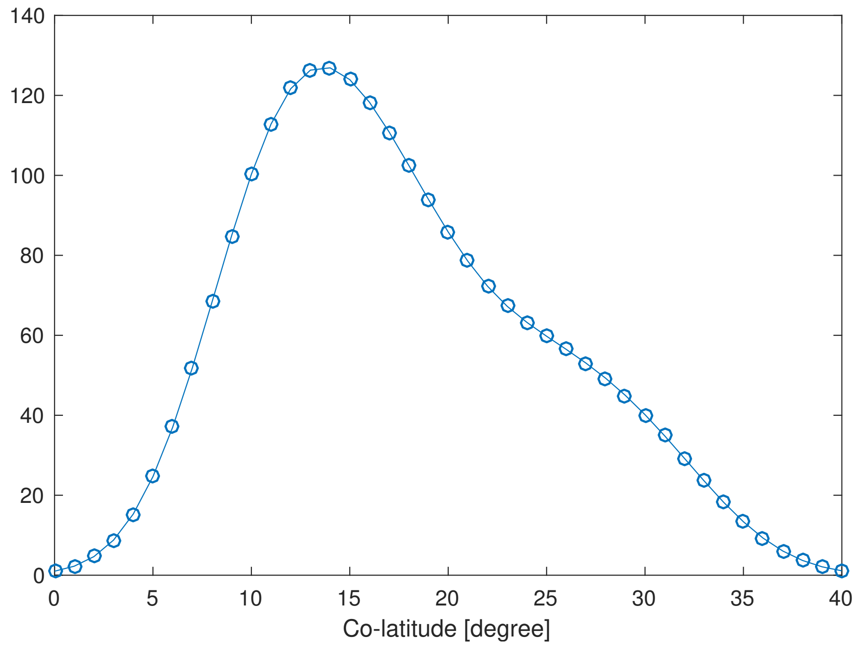

Figure 2. The fitted variance profile model g defined in Equation (6) to residual SuperDARN LOS plasma drift velocity data is shown as the latitudinal distribution of velocity standard deviation in m/s.

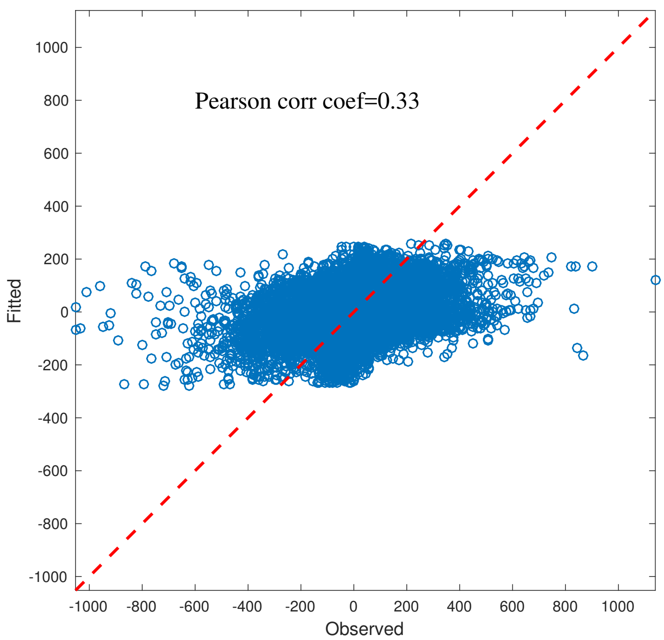

Figure 3. Out-of-sample prediction of residual SuperDARN LOS plasma drift velocities versus observed values in m/s. Pearson correlation coefficient between predicted and observed values is 0.33.

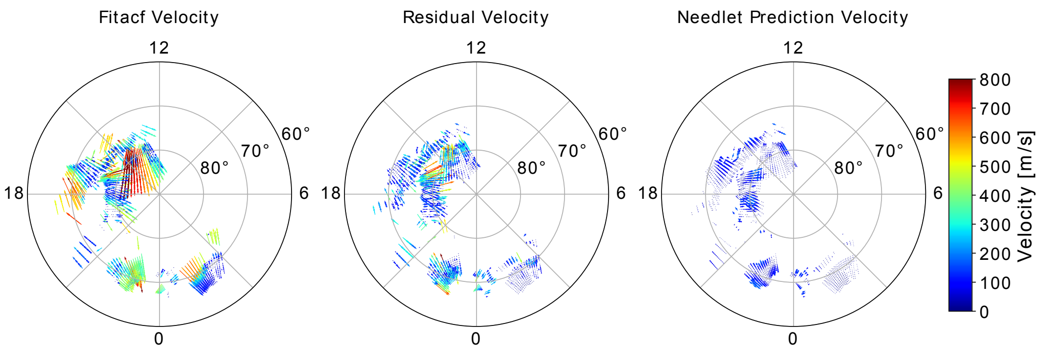

Figure 4. LOS plasma velocities from one 2-minute scan of SuperDARN radars from 0004 to 0006 UT on February 29, 2012: (Left) The FITACF LOS plasma velocity ($v_{\mathrm{LOS}}^{\mathrm{fitacf}}$), (Middle) residual velocity ($v_{\mathrm{LOS}}^{\mathrm{fitacf}} - \hat{v}_{\mathrm{LOS,g}}$), and (Right) *regional small-scale* velocity modeled by the needlet model ($\hat{v}_{\mathrm{LOS,r}}$).

ing estimation shown in the next subsection. The potential fields at two needlet resolution levels at $j = 2, 3$ are shown in the first two columns, and the total potential fields, which is a combination of all resolution levels, is shown in the right-most column.

The smallest scales resolved at these two needlet resolution levels correspond to the spherical harmonics frequency (degree) and phase (order) of $l = 8$, $m = 8$ and $l = 16$, $m = 16$, respectively, which are equivalent to the spatial scales of 5.6° in latitude and 22.5° in longitude for $j = 2$ and 2.8° in latitude and 11.25° in longitude for $j = 3$, with consideration of the factor 4 latitude coordinate stretching of the analysis domain as described in Section 3.2. (Note that an approximate spatial resolution corresponding to a certain degree and order of the spherical harmonic function is obtained using the Nyquist frequency of a half wavelength.) Even though the resolutions of needlets and spherical harmonics are comparable, due to needlets' spatial and frequency localization and frame properties that needlet-based model can better represent localized regional features that exist in the SuperDARN observations, in comparison to spherical harmonics with a global support that are designed to capture global structures.

The SAM used to model *global large-scale* electric fields from the SuperDARN observations (see Section 2.3) can resolve scales up to the spherical harmonics degree and order of about $l = 72$ and $m = 12$, corresponding to the resolution of 2.5° in latitude and 15° in longitude. As shown in Figure 6 for 0300 UT, the SAM field in fact exhibits *global large-scale* features in comparison to *regional small-scale* features that are present in muti-resolution random electrostatic fields estimated from the needlet-based modeling approach (Figure 5). Due to the variance profile, shown in Figure 2, which peaks around 75° latitude, more distinct features appear between 70° and 80° in *regional small-scale* fields.

4.2 Impact on Joule Heating Rate

Figure 7 shows the ensemble mean of the hemispherically integrated Joule heating rate computed with the effect of random *regional small-scale* electric fields $\mathbf{E}'$ (blue) as well as the hemispherically-integrated Joule heating without the effect of $\mathbf{E}'$ (black). The hemispherically-integrated Joule heating rate with the effect of $\mathbf{E}'$ is integrated over the northern hemisphere high-latitude ionosphere from 45° to 90° in latitude, following

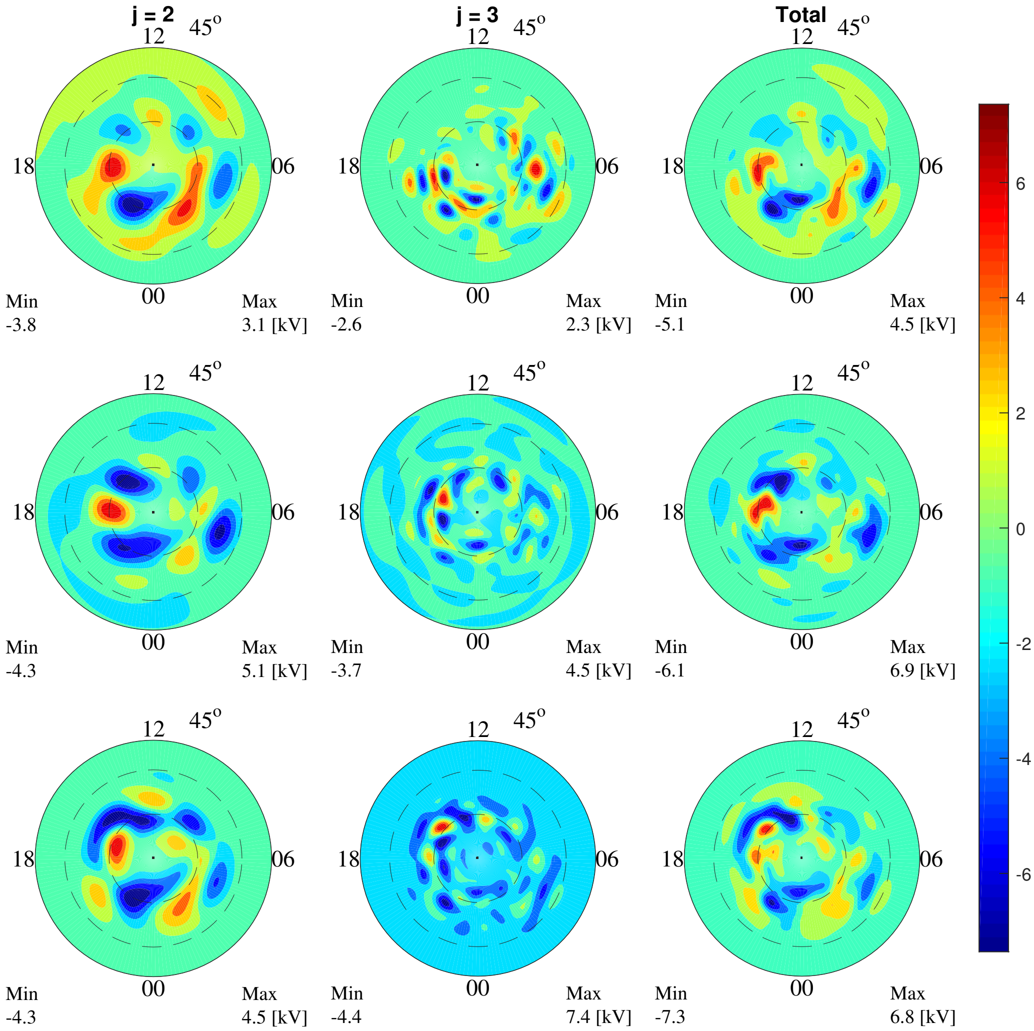

Figure 5. Electrostatic potential fields at different resolutions generated from the needlet model. Top row: mean prediction conditional on the observations; Middle row: a random sample conditional on the observations; Bottom row: a random sample conditional on the observations. Left: field at level $j = 2$; Middle: field at level $j = 3$; Right: total field summed at levels $j = 2, 3$.

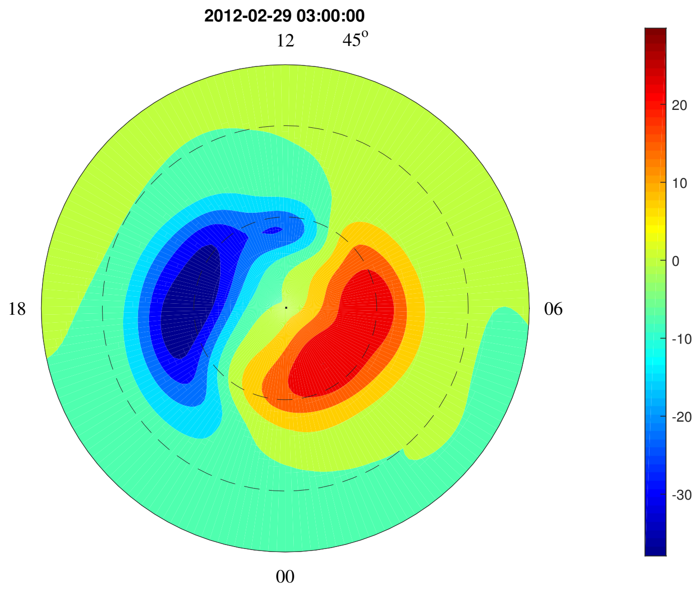

Figure 6. Global large-scale Electrostatic potential field, estimated from the SuperDARN GRID data over 4-minute windows using the SAM procedure, at 3:00 UT on February 29, 2012, around the peak of AE index.

the definition given in Equation (1), as

$$Q_J^{(n)}(t) \approx \iint_{\theta,\phi} \Sigma_p(\theta,\phi,t)(\overline{\mathbf{E}}(\theta,\phi,t) + \mathbf{E}'^{(n)}(\theta,\phi))^2 \, d\theta \, d\phi \; , \tag{7}$$

where n is an ensemble member index, $n = 1, \cdots, 1000$, the *global large-scale* electric field $\overline{\mathbf{E}}$ is specified by the SAM as described in Section 2.3. Note that a 1000-member ensemble set of random *regional small-scale* electric field $\{\mathbf{E}'^{(1)}, \mathbf{E}'^{(2)}, \cdots, \mathbf{E}'^{(1000)}\}$ is computed from $\{\Phi_{\mathrm{E,r}}^{(1)}, \Phi_{\mathrm{E,r}}^{(2)}, \cdots, \Phi_{\mathrm{E,r}}^{(1000)}\}$ as described in Section 4.1. The hemispherically-integrated Joule heating without the effect of $\mathbf{E}'$ is given as

$$\overline{Q}_J(t) \approx \iint_{\theta,\phi} \Sigma_p(\theta,\phi,t)\overline{\mathbf{E}}(\theta,\phi,t)^2 \, d\theta \, d\phi \; . \tag{8}$$

In Equations (7) and (8), the height-integrated ionospheric conductivity $\Sigma_p(\theta,\phi)$ is specified using empirical models of the solar EUV conductance and auroral conductance. The solar EUV conductance model is parameterized by solar zenith angle and the solar F10.7 index (e.g., Moen & Brekke, 1993), and the auroral conductance is based on the Ovation Prime empirical aurora model (Newell & Wing, 2009) and the empirical relationship of Robinson et al. (1987). Note that the Ovation Prime model is parameterized with respect to the upstream solar wind and interplanetary magnetic field conditions. Except for a minor geomagnetic activity, there is no notable geomagnetic activity during the time period of 00:00 to 04:00 UT on February 29, 2012. The overall Joule heating rate is thus small. In general, the Joule heating rate tracks temporal changes of the AE index (red) shown also in Figure 7, which is due to the changes of large-scale electric fields. By taking the *regional small-scale* electric field variability into account, the Joule heating rate increases by a factor of about 1.5 which is close to the lower end of an arbitrarily adjusted factor of 1.5 and 2.5 typically used in general circulation models. As discussed in Section 1, the biases in the upper atmosphere general circulation models attributed to an

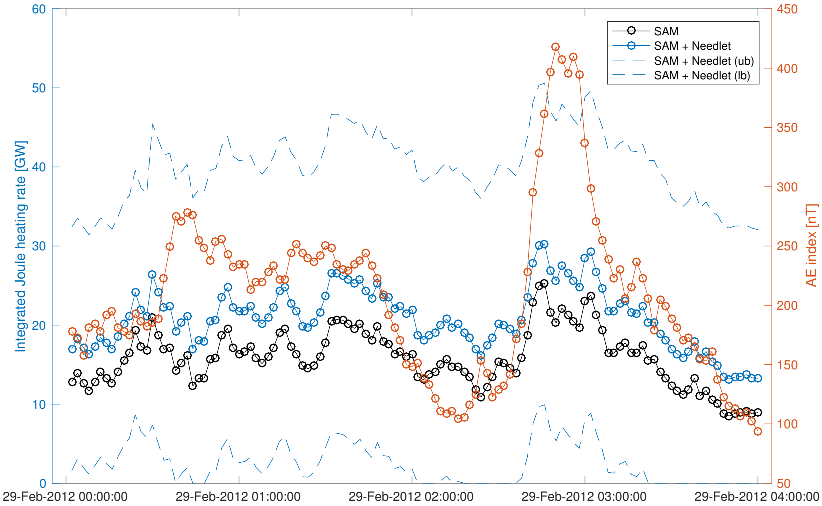

Figure 7. Hemispherically integrated Joule heating rate in GW from 00:00 to 04:00 UT on February 29, 2012. The ensemble mean of the Joule heating rate computed with the effect of random *regional small-scale* electric fields $\mathbf{E}'^{(n)}$ (Equation (7)) is shown in blue solid line, along with the upper and lower bounds given in terms of two standard deviations shown in blue dash lines. The Joule heating rate resulting only from *global large-scale* electric field $\overline{\mathbf{E}}$ without $\mathbf{E}'$ (Equation (8)) is shown in black line. As a reference, the high-latitude geomagnetic activity, AE index nT, is overlaid in red.

inadequate representation of the Joule heating rate exist regardless of geomagnetic activity levels. The future study needs to allow for more flexibility in the needlet model to account for temporal variation of the variance profile so that the changes of electric field variability at different geomagnetic activity levels can be better characterized.

5 Discussion and Conclusions

In response to the need for a new statistical inferential framework for data-driven modeling of high-latitude ionospheric electric field variability, Fan et al. (2018)'s spherical needlet-based scalar random fields model is being extended for vector random fields and applied to the carefully curated SuperDARN FITACF LOS plasma velocity data set. The modeling results for the largely quiet period from 0000 to 0400 UT on February 29, 2012 show that the approach have the potential to rectify the underestimation of the Joule heating rate in the current upper atmosphere general circulation models due to insufficient representation of the electric field variability. The study demonstrates how data-driven modeling of the magnetosphere-ionosphere-thermosphere coupling can be formulated in an ensemble modeling framework. Specific findings of the current efforts are summarized as follows.

The needlet-based approach to modeling *regional small-scale* electric field variability can help estimate a distribution of electric field variability conditioned on actual SuperDARN LOS observations. As shown in Figure 5, estimated *regional small-scale* electrostatic potential fields at different resolutions exhibit considerably more localized fine-scale features in comparison to *global large-scale* potential fields modeled using the SAM procedure (Figure 6). This is enabled by the spherical needlet frames' spatial localization and overcompleteness properties and reinforces the fact that spherical harmonic basis functions, with global support, are not suited for describing features that are spatially localized.

As shown in Figure 7, the overall hemispherically integrated Joule heating rate is increased by a factor of about 1.5 due to the effect of random *regional small-scale* electric fields $\mathbf{E}'$, which is close to the lower end of arbitrarily adjusted Joule heating multiplicative factor of 1.5 and 2.5 typically used in upper atmosphere general circulation models. The impact of the modeled electric field variability on the Joule heating rate is computed using a 1000-member ensemble set of *regional small-scale* electric fields $\{\mathbf{E}'^{(1)}, \mathbf{E}'^{(2)}, \cdots, \mathbf{E}'^{(1000)}\}$. This example demonstrates that how the uncertainty of the SuperDARN LOS data can be propagated to the estimate of Joule heating rate in general circulation models through the needlet-based modeling of the ionospheric electric variability. The approach can also be applied to the output from high-fidelity high-resolution numerical simulations that may be computationally prohibitive to perform routinely. This study is an important step towards a data-driven ensemble modeling of magnetosphere-ionosphere-atmosphere coupling processes.

Some of the methodological shortcomings identified by the current study can be addressed in future work. The needlet model can be expanded to account for non-stationarity of the electric field variability not only with respect to magnetic latitudes but also MLT. By doing so, the electric field variability associated with specific physical processes such as convection reversal, and auroral electrojet that appear in localized locations can be better represented. As suggested by the out-of-sample prediction of LOS plasma drift velocities shown in Figure 3, SuperDARN LOS residual velocities with greater magnitudes ($|v_{\mathrm{LOS}}^{\mathrm{fitacf}} - \hat{v}_{\mathrm{LOS,g}}| > 350$ m/s) are not well predicted by the needlet model. This is evident in Figure 4. This can be addressed by increasing the needlet resolution level from $j = 3$ to $j = 4$, equivalent to the spatial scales of 1.4 degrees in latitude and 5.6 degrees in longitude, ideally to $j = 5$, corresponding to the scales of 0.7 degrees in latitude and 2.8 degrees in longitude. Uncertainty resulting from inconsistent model assumptions associated with spatiotemporal stationarity of random fields should be better quantified using more data.

These methodological improvements will have to be accompanied with an improved uncertainty quantification in the determination of SuperDARN LOS velocity from radar backscatter. The availability of SuperDARN data with greater spatial coverage will alleviate the need to aggregate data over time allowing us to drop the assumption of spatiotemporal stationarity of random fields in the method. In order to increase the needlet resolution level to $j = 5$, SuperDARN data at a higher spatial resolution will be needed. Prospects for the availability of such SuperDARN LOS velocity data sets are discussed next. With more SuperDARN radars being constructed (e.g., Adak Island East and West radars in 2012, Hokkaido West radar in 2014, and Jiamusi radar in 2019 (Nishitani et al., 2019)), we could have a better spatial coverage for future work. However, there still exist a few challenges on obtaining more SuperDARN data. Firstly, lack of ionospheric backscatter in SuperDARN data during the day would cause a data gap in MLT, particularly at mid-latitudes (Figure 1b). Secondly, strong particle precipitation during geomagnetically active times could cause radar signals absorbed by the ionosphere. The Local Divergence-Free Fitting technique from Bristow et al. (2016) is able to provide SuperDARN plasma velocity with a spatial resolution (~50 km) that is comparable to the LOS velocity measurements. The LDFF technique uses all LOS velocities within a user-defined region to produce local plasma convection which can resolve finer-scale structures such as plasma flows associated with auroral arcs. This technique can be used in the future to obtain time-dependent (mean) vector fields $\overline{\mathbf{E}}$ at much finer scales than the SAM, which is expected to improve the signal-to-noise ratios of residual SuperDARN LOS velocity data for the method presented in this study.

Acknowledgments

The authors acknowledge the use of SuperDARN data. SuperDARN is a collection of radars funded by national scientific funding agencies of Australia, Canada, China, France, Italy, Japan, Norway, South Africa, United Kingdom and the United States of America. The SuperDARN data used for the study is available from https://doi.org/10.17605/OSF.IO/E7W8P. The authors are grateful for assistance and helpful guidance from Ellen Cousins and Doug Nychka at the initial stage of the study. The study is supported by NSF OPP-1443703, NSF DMS-1811279 and DMS-1811405. TM acknowledges the support of NSF AGS-1848544, and XS and JMR acknowledge the support of NSF AGS-1341918 and NSF AGS-1935110. DP acknowledges the support of NSF grants DMS-1713120 and DMS-1915894, and TCML acknowledges the support of NSF grants DMS-1811405, DMS-1811661, DMS-1916125 and CCF-1934568.

References

Andrieu, C., & Thoms, J. (2008). A tutorial on adaptive mcmc. *Statistics and Computing*, *18*(4), 343-373.

Bristow, W. A., Hampton, D. L., & Otto, A. (2016). High-spatial-resolution velocity measurements derived using Local Divergence-Free Fitting of SuperDARN observations. *J. Geophys.Res.*, *121*, 1349-1361. doi: 10.1002/2015JA021862

Buneman, O. (1963). Excitation of field aligned sound waves by electron streams. *Phys. Rev. Lett.*, *10*, 285–287.

Burns, A. G., Wang, W., Solomon, S. C., & Qian, L. (2014). In J. Huba, R. Schunk, & G. Khazanov (Eds.), (chap. Energetics and Composition in the Thermosphere). American Geophysical Union.

Chisham, G., Lester, M., Milan, S. E., Freeman, M. P., Bristow, W. A., Grocott, A., ... Walker, A. D. M. (2007). A decade of the Super Dual Auroral Radar Network (SuperDARN): scientific achievements, new techniques and future directions. *Surv. Geophys.*, *28*, 33–109. doi: 10.1007/s10712-007-9017-8

Codrescu, M. V., Fuller-Rowell, T. J., & Foster, J. C. (1995). On the importance of E-field variability for Joule heating in the high-latitude thermosphere. *Geophys.Res.Lett.*, *22*(17), 2393-2396.

Codrescu, M. V., Fuller-Rowell, T. J., Foster, J. C., Holt, J. M., & Cariglia, S. J. (2000). Electric field variability associated with the Millstone Hill electric field model. *J. Geophys.Res.*, *105*(A3), 5265-5273.

Cousins, E. D. P., Matsuo, T., & Richmond, A. D. (2013a). Mesoscale and large-scale variability in high-latitude ionospheric convection: Dominant modes and spatial/temporal coherence. *J. Geophys.Res.*, *118*, 7895-7904. doi: 10.1002/2013JA019319

Cousins, E. D. P., Matsuo, T., & Richmond, A. D. (2013b). SuperDARN assimilative mapping. *J. Geophys.Res.*, *118*, 1-9. doi: 10.1002/2013JA019321

Cousins, E. D. P., & Shepherd, S. G. (2010). A dynamical model of high-latitude convection derived from SuperDARN plasma drift measurements. *J. Geophys.Res.*, *115*. doi: 10.1029/2010JA016017

Cousins, E. D. P., & Shepherd, S. G. (2012). Statistical maps of small-scale electric field variability in the high-latitude ionosphere. *J. Geophys.Res.*, *117*. doi: 10.1029/2012JA017929

Deng, Y., Maute, A., Richmond, A. D., & Roble, R. G. (2009). Impact of electric field variability on Joule heating and thermospheric temperature and density. *Geophys.Res.Lett.*, *36*. doi: 10.1029/2008GL036916

Dimant, Y. S., & Oppenheim, M. M. (2011). Magnetosphere-ionosphere coupling through E region turbulence: 2. Anomalous conductivities and frictional heating. *J. Geophys.Res.*, *116*. doi: 10.1029/2011JA016649

Fan, M., Paul, D., Lee, T. C. M., & Matsuo, T. (2018). Multi-resolution model for non-gaussian random fields on a sphere with application to ionospheric electrostatic potentials. *The Annals of Applied Statistics*, *12*, 459-489.

Farley, D. T. (1963). A plasma instability resulting in field-aligned irregularities in the ionosphere. *J. Geophys.Res.*, *68*, 6083–6097. doi: 10.1029/JZ068i022p06083

Fuller-Rowell, T. J. (2014). In J. Huba, R. Schunk, & G. Khazanov (Eds.), (chap. Physical Characteristics and Modeling of Earth's Thermosphere). American Geophysical Union.

Gelman, A., Roberts, G., & Gilks, W. (1996). Efficient metropolis jumping rules. In J. M. Bernado et al. (Eds.), *Bayesian statistics* (Vol. 5, p. 599). OUP.

Groves, K. M., & Carrano, C. S. (2016). Space weather effects on communication and navigation. In G. Khazanov (Ed.), *Space weather fundamentals* (1st ed., p. 353-387). Boca Raton: CRC Press. doi: 10.1201/9781315368474

Jackson, J. D. (1999). *Classical electrodynamics* (3rd ed.). New York: John Wiley & Sons.

Kelly, M. (2009). *The earth's ionosphere* (2nd ed.).

Liu, J., Wang, W., Oppenheim, M., Dimant, Y., Wiltberger, M., & Merkin, S. (2016). Anomalous electron heating effects on the E region ionosphere in TIEGCM. *Geophys.Res.Lett.*, *43*, 2351–2358. doi: 10.1002/2016GL068010

Lu, G., Richmond, A., Emery, B. A., & Roble, R. G. (1995). Magnetosphere-ionosphere-thermosphere coupling: Effect of neutral winds on energy transfer and field-aligned current. *J. Geophys.Res.*, *100*(A10), 19643-19659.

Marcos, F. A., Lai, S. T., Huang, C. Y., Lin, C. S., Retterer, J. M., Delay, S. H., & Sutton, E. K. (2010, August). Towards next level satellite drag modeling. In *AIAA atmospheric and space environments conference.* Toronto, Ontario Canada.

Marinucci, D., & Peccati, G. (2011). *Random fields on the sphere: Representation, limit theorems and cosmological applications.* Cambridge: Cambridge University Press.

Matsuo, T., & Richmond, A. D. (2008). Effects of high-latitude ionospheric electric field variability on global thermospheric joule heating and mechanical energy transfer rate. *J. Geophys.Res.*, *113*. doi: 10.1029/2007JA012993

Matsuo, T., Richmond, A. D., & Hensel, K. (2003). High-latitude ionospheric electric field variability and electric potential derived from DE-2 plasma drift measurements: Dependence on IMF and dipole tilt. *Journal of Geophysical Research: Space Physics*, *108*(A1). doi: 10.1029/2002JA009429

Matsuo, T., Richmond, A. D., & Lu, G. (2005). Optimal interpolation analysis of high-latitude ionospheric electrodynamics using empirical orthogonal functions: Estimation of dominant modes of variability and temporal scales of large-scale electric fields. *J. Geophys.Res.*, *110*. doi: 10.1029/2004JA010531

Matsuo, T., Richmond, A. D., & Nychka, D. W. (2002). Modes of high-latitude electric field variability derived from DE-2 measurements: Empirical Orthogonal Function (EOF) analysis. *Geophys.Res.Lett.*, *29*(7). doi: 10.1029/2001GL014077

Moen, J., & Brekke, A. (1993). The solar flux influence on quiet time conductances in the auroral ionosphere. *Geophys.Res.Lett.*, *20*(10), 971-974. doi: 10.1029/92GL02109

Newell, T. S., P. T., & Wing, S. (2009). Diffuse, monoenergetic, and broadband aurora: The global precipitation budget. *J. Geophys.Res.*, *114*. doi: 10.1029/2009JA014326

Nishitani, N., Ruohoniemi, J. M., Lester, M., Baker, J. B. H., Koustov, A. V., Shepherd, S. G., ... Kikuchi, T. (2019). Review of the accomplishments of mid-latitude Super Dual Auroral Radar Network (SuperDARN) HF radars. *Prog. Earth Planet Sci.*, *6*. doi: 10.1186/s40645-019-0270-5

Pulkkinen, A., Bernabeu, E., Thomson, A., Viljanen, A., Pirjola, R., Boteler, D., ... MacAlester, M. (2017). Geomagnetically induced currents: Science, engineering, and applications readiness. *Space Weather*, *15*(7), 828-856. doi: 10.1002/2016SW001501

Ribeiro, A. J., Ruohoniemi, J. M., Ponomarenko, P. V., Clausen, L. B. N., Baker, J. B. H., Greenwald, R. A., . . . de Larquier, S. (2013). A comparison of SuperDARN ACF fitting methods. *Radio Sci.*, *48*, 274–282. doi: 10.1002/rds.20031

Richmond, A. D., & Kamide, Y. (1988). Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique. *J. Geophys.Res.*, *93*(A6), 5741-5759.

Ridley, A. J., Richmond, A. D., Gombosi, T. I., De Zeeuw, D. L., & Clauer, C. R. (2003). Ionospheric control of the magnetospheric configuration: Thermospheric neutral winds. *J. Geophys.Res.*, *108*(A8). doi: 10.1029/2002JA009464

Robinson, R. M., Vondrak, R. R., Miller, K., Dabbs, T., & Hardy, D. (1987). On calculating ionospheric conductances from the flux and energy of precipitating electrons. *J. Geophys.Res.*, *92*, 2565–2569.

Ruohoniemi, J. M., & Baker, K. B. (1998). Large-scale imaging of high-latitude convection with Super Dual Auroral Radar Network HF radar observations. *J. Geophys.Res.*, *103*(A9), 20797-20811. doi: 10.1029/98JA01288

Sangalli, L., Knudsen, D. J., Larsen, M. F., Zhan, T., Pfaff, R. F., & Rowland, D. (2009). Rocket-based measurements of ion velocity, neutral wind, and electric field in the collisional transition region of the auroral ionosphere. *J. Geophys.Res.*, *114*. doi: 10.1029/2008JA013757

Schunk, R. (2014).
In J. Huba, R. Schunk, & G. Khazanov (Eds.), (chap. Modeling the Ionosphere–Thermosphere System). Washington, DC, USA: American Geophysical Union. doi: 10.1002/9781118704417

Scott, J. G. (2011, 06). Bayesian estimation of intensity surfaces on the sphere via needlet shrinkage and selection. *Bayesian Analysis*, *6*(2), 307-327. doi: 10.1214/11-BA611

Thayer, J. P. (1998). Height-resolved joule heating rates in the high-latitude e region and the influence of neutral winds. *J. Geophys.Res.*, *103*(A1), 471-487. doi: 10.1029/97JA02536

Wiltberger, M., Rigler, E. J., Merkin, V., & Lyon, J. G. (2016). Structure of high latitude currents in magnetosphere-ionosphere models. *Space Sci. Rev.*, *206*, 575–598. doi: 10.1007/s11214-016-0271-2

Zhu, Q., Deng, Y., Richmond, A., & Maute, A. (2018). Small-scale and mesoscale variabilities in the electric field and particle precipitation and their impacts on Joule heating. *J. Geophys.Res.*, *123*, 9862–9872. doi: 10.1029/2018JA025771