ADVANCING EARTH AND SPACE SCIENCE

Space Weather

RESEARCH ARTICLE
10.1029/2021SW002880

Key Points:
- The multi-resolution Gaussian process model has been successfully adapted to assimilate electric fields from line-of-sight measurements
- This method captures finer scale variations in regional data than background models and decreases fitting errors with increasing resolutions
- The multi-level basis functions enable the regional fine-scale modeling and achieve the best fitting performance with reduced computation

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
H. Wu and X. Lu,
haonanw@g.clemson.edu;
xianl@clemson.edu

Citation:
Wu, H., & Lu, X. (2022). Data assimilation of high-latitude electric fields: Extension of a multi-resolution Gaussian process model (Lattice Kriging) to vector fields. *Space Weather*, *20*, e2021SW002880. https://doi.org/10.1029/2021SW002880

Received 13 AUG 2021
Accepted 3 DEC 2021

Author Contributions:
Conceptualization: Haonan Wu, Xian Lu
Formal analysis: Haonan Wu, Xian Lu
Funding acquisition: Xian Lu
Investigation: Haonan Wu, Xian Lu
Methodology: Haonan Wu, Xian Lu
Project Administration: Xian Lu
Resources: Xian Lu
Software: Haonan Wu
Supervision: Xian Lu
Validation: Haonan Wu
Visualization: Haonan Wu, Xian Lu

Data Assimilation of High-Latitude Electric Fields: Extension of a Multi-Resolution Gaussian Process Model (Lattice Kriging) to Vector Fields

Haonan Wu[1] and Xian Lu[1]

[1]Department of Physics and Astronomy, Clemson University, Clemson, SC, USA

Abstract We develop a new methodology for the multi-resolution assimilation of electric fields by extending a Gaussian process model (Lattice Kriging) used for scalar field originally to vector field. This method takes the background empirical model as "a priori" knowledge and fuses real observations under the Gaussian process framework. The comparison of assimilated results under two different background models and three different resolutions suggests that (a) the new method significantly reduces fitting errors compared with the global spherical harmonic fitting (SHF) because it uses range-limited basis functions ideal for the local fitting and (b) the fitting resolution, determined by the number of basis functions, is adjustable and higher resolution leads to smaller errors, indicating that more structures in the data are captured. We also test the sensitivity of the fitting results to the total amount of input data: (a) as the data amount increases, the fitting results deviate from the background model and become more determined by data and (b) the impacts of data can reach remote regions with no data available. The assimilation also better captures short-period variations in local PFISR measurements than the SHF and maintains a coherent pattern with the surrounding. The multi-resolution Lattice Kriging is examined via attributing basis functions into multiple levels with different resolutions (fine level is located in the region with observations). Such multi-resolution fitting has the smallest error and shortest computation time, making the regional high-resolution modeling efficient. Our method can be modified to achieve the multi-resolution assimilation for other vector fields from unevenly distributed observations.

Plain Language Summary Earth's upper atmosphere is vulnerable to energy injection from the Sun via the channel of solar wind-magnetosphere-ionosphere-thermosphere coupling. Geomagnetic storms induce dramatic changes in ion motions, electron densities, and composition, which further drive neutral atmosphere into a highly disturbed condition. Such disturbances in neutrals, ions, and electrons lead to an important component of space weather, affecting satellite drag, ionospheric scintillation, radio propagation, and GPS navigation. It is therefore critical to better understand space weather and underlying processes. The high-latitude electric field, largely imposed from the magnetosphere, controls electrodynamic and dynamic processes in the ionosphere-thermosphere. It is also an important driver for numerical models to simulate and predict space weather. Empirical models of electric field are unrealistically smooth and miss localized feature. Data assimilation technique helps to fuse data information and solves this issue. Previous techniques rely on the global harmonics fitting, which is not friendly in local fitting. In this paper, we introduce a Gaussian process model (Lattice Kriging) and propose a new methodology based on its extension to vector fields. This new method is robust and significantly decreases fitting errors, providing a useful tool for the research aiming to perform regional high-resolution and multi-resolution data assimilation upon unevenly distributed observations.

1. Introduction

Ionospheric plasma convection is largely driven by the electrodynamic processes in the magnetosphere, which is controlled by the interaction between magnetosphere and solar wind. Thus, ion convection is a key indicator of the ionospheric responses to geomagnetic variations. Ion motion can enhance, recede, and even reverse in reacting to different interplanetary magnetic field (IMF) and solar wind conditions. During disturbed periods, enhanced ion convection transports midlatitude plasma into the polar cap, leading to tongues of ionization and patches, which can disrupt communication and navigation in the polar region (Buchau et al., 1983; Nishimura et al., 2021; Weber et al., 1984). The radio backscatter technique has been widely used to measure ion motions. For instance, the Super Dual Auroral Radar Network (SuperDARN) scans over azimuth sectors on a regular basis

Writing – original draft: Haonan Wu, Xian Lu
Writing – review & editing: Haonan Wu, Xian Lu

(typically 2 min) and measures line-of-sight (LOS) ion drifts therein. The LOS ion drift measurements provide ion convection information over high-latitude and midlatitude regions. When different radars receive signals from different directions at the same location, the vector ion drift at that location can be directly retrieved (Bristow et al., 2016; Hanuise et al., 1993). However, due to the limited coverage of SuperDARN radars, the rate of overlapping field of view is relatively low; thus, the direct derivation of vector drifts from the LOS measurements is limited (Bristow et al., 1995; Sánchez et al., 1996).

The retrieval of the global convection pattern from LOS measurements using other techniques has long been investigated. Best-known techniques include SuperDARN spherical harmonic fitting (SHF) (Ruohoniemi & Baker, 1998) and assimilative mapping of ionospheric electrodynamics (AMIE, Richmond & Kamide, 1988). SuperDARN SHF derives vector ion drifts by minimizing the weighted squared errors between the LOS ion drifts and spherical harmonic expansions, and the fitted patterns are widely used in both quiet- and storm-time studies (e.g., Maimaiti et al., 2018; Zhang et al., 2020). AMIE uses spherical cap harmonics as the basis function to fit the LOS ion drifts and provides more realistic high-latitude electric fields in the storm time than the empirical models (Hsu et al., 2021; Lu et al., 2020; Richmond, 1992). Both methods use a known background model to provide constraints in the fitting process. The degree of the fitting precision, also referred to as "resolution," is controlled by the order of basis functions, which describes the number of harmonics along the longitudinal circle. For both methods, the background model is weighted where no observations are available. For higher orders, more model points are sampled so the patterns are more heavily weighted by the background model (Bristow et al., 2016). Thus, for those methods using global basis functions (e.g., spherical harmonics and spherical cap harmonics), the fitting is constrained by limiting the sampling of the background model; thus, the fitting resolution cannot be too high. A typical choice is on the order of 10° in longitude and 2° in latitude (Lu, 2017; Matsuo, 2020).

Nevertheless, the magnetosphere-ionosphere-thermosphere coupled system embraces a variety of important medium- to small-scale electrodynamic processes, which are below the resolution resolvable by the SuperDARN SHF or AMIE methods. The electric fields at polar cap and auroral region exhibit cross-scale power spectra all the way from planetary scales down to 0.5 km (Golovchanskaya & Kozelov, 2010a; Kozelov & Golovchanskaya, 2006), which lead to the deviation from the global large-scale two-cell ion convection pattern (Cousins & Shepherd, 2012a, 2012b). Small-scale electric fields have been often observed and found to impact the energy budget during magnetic storms (Codrescu et al., 1995; Cosgrove & Codrescu, 2009). Wu et al. (2020) found that an accurate specification of the local electric fields varying in short temporal scales and satellite-observed auroras showing small-scale spatial variations is essential to reproduce the observed local temperature enhancement ($\sim$500 K) and inversion layer in the E-region ($\sim$130 km). Sheng et al. (2020) found that using ground-based auroral imager observations characterized by mesoscale features better resolves the large-scale traveling atmospheric disturbances than using the empirical auroral maps. These studies illustrate that a better quantification of energy inputs fusing data information with regional scales significantly improves the simulation of ionospheric/thermospheric responses to geomagnetic storms.

To better use the LOS ion drift measurements, which are typically of 1° (e.g., SuperDARN), several different methods have been proposed to accommodate the high-resolution data. For example, Amm et al. (2010) used spherical elementary current system (SECS) as the basis functions and solved the coefficients of SECS using the divergence-free condition of ion drifts. This method does not rely on "a priori" information provided by background models. Bristow et al. (2016) proposed a local divergence-free fitting technique, which also relies on the divergence-free assumption of ion drifts. The relation between the vector field and its LOS component is imposed as an additional constraint of the system from which the vector ion drift is derived during the fitting process. Large-scale SHF results are imposed as boundary constraints and additionally as "a priori" knowledge to the framework. Both methods produce vector ion drifts at higher resolutions than the typical global basis function fitting technique.

Here, we propose an alternative method of retrieving vector ion drifts or equivalently electric fields, out of the LOS measurements using a multi-resolution Gaussian process model also called Lattice Kriging (Nychka et al., 2015). This methodology has been used to analyze surface temperatures and make the prediction at regions without observations (Heaton et al., 2019; Wiens et al., 2020). However, the previous applications of the Lattice Kriging method are largely limited to the assimilation of scalar fields. In this paper, we develop an extension of this methodology and apply it for the assimilation of vector fields, that is, electric fields in our case. Such extension assumes that the high-latitude electric field is curl-free (equivalent to the divergence-free constraint of ion

10.1029/2021SW002880

drifts). Using LOS information of electric fields and certain background models, the electric potential can be retrieved over the whole domain. By adjusting the sparseness/fineness of the basis functions in multiple levels, the model can be used to study the multi-resolution structures of the electric field. Our testing results show that the fitted results are in accordance with the inputs both locally and over the high-latitude region where SuperDARN observations are available. The method reduces the error of LOS electric fields, and the improvement is more significant when a set of higher resolution basis function is used.

The mathematical formulation is presented in Section 2, including a synthetic test to verify the formulation. Then, we apply this method to real observations (i.e., SuperDARN and Poker Flat Incoherent Scatter Radar (PFISR)) to explore its applicational performance (Section 3). Conclusions are given in Section 4.

2. Methodology

2.1. Fundamental Formulation of Lattice Kriging for Data Assimilation

The original formulation of Lattice Kriging was proposed in Nychka et al. (2015). Suppose there are n observations, an observation y_i ($1 \leq i \leq n$) made at location x_i is separated into a statistical mean μ_i and a deviation $g(x_i)$ plus an error term ϵ_i

$$y_i = \mu_i + g(x_i) + \epsilon_i \tag{1}$$

Both observation y_i and deviation $g(x_i)$ are taken as Gaussian processes. μ_i provides a prior knowledge of the statistical characteristics of the field and serves as a background (starting point) for data assimilation. In space physics, μ_i is usually taken as an empirical model $Z_i(\boldsymbol{a})$ such as the Weimer model (Weimer, 2005) in which $\boldsymbol{a}$ contains the inputs of the model (e.g., solar wind parameters and IMF). Since we take the results from the empirical models directly as our inputs, the dependence of Z_i on $\boldsymbol{a}$ is treated by the empirical models and not relevant to our procedures. Therefore, $Z_i(\boldsymbol{a})$ is simplified as Z_i throughout this paper. For our application, we consider the systematic bias of the empirical model compared to observations and assign a linear scaling factor d to be multiplied by the empirical model when taken as statistical mean, so

$$\mu_i = Z_i d \tag{2}$$

Then, the only input parameter to be estimated when we apply the empirical model is the scaling factor d.

The formulation of Lattice Kriging assumes that $g(x)$, the value of the deviation field at a certain location x, can be decomposed into a set of basis functions $R_j(x)$ ($1 \leq j \leq m$, the total number of basis functions is m),

$$g(x) = \sum_{j=1}^{m} c_j R_j(x) \tag{3}$$

where c_j is the coefficient of the jth basis function. Being a Gaussian process, the covariance function of $g(x)$, describing the spatial correlation of any two locations (x and x'), can be written as a second-order term of $R(x)$,

$$\mathrm{cov}(g(x), g(x')) = \sum_{\substack{1 \leq j \leq m \\ 1 \leq j' \leq m}} R_j(x) \boldsymbol{Q}^{-1}_{j,j'} R_{j'}(x') \tag{4}$$

where j and j' are indices of two basis functions and $\boldsymbol{Q}^{-1}$ is the covariance matrix in the representation of $R_j(x)$. Nychka et al. (2015) choose the basis function $R_j(x)$ as the radial basis function (RBF), and the detailed formulation of $\boldsymbol{Q}^{-1}$ can be found in Sections 2.4 and 2.5 of the paper.

Now, we write the basis functions into an $n \times$ m matrix $\boldsymbol{R}$ with its elements satisfying

$$R_{i,j} = R_j(x_i) \tag{5}$$

which is the evaluation of the jth basis function at the observation location of x_i. Note that the total number of basis functions is m and the total number of observational points is n. Then, expanding Equations 3 and 4 to all n observational points leads to their corresponding matrix forms as follows:

$$g(\boldsymbol{x}) = \boldsymbol{Rc} \tag{6}$$

$$\mathrm{cov}(g(\boldsymbol{x}), g(\boldsymbol{x'})) = \boldsymbol{RQ}^{-1}\boldsymbol{R}^{T} \tag{7}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots x_n)$ and $\boldsymbol{c} = (c_1, c_2, \ldots c_m)$. The error covariance matrix $\boldsymbol{W}^{-1}$ is assumed to be diagonal

$$\boldsymbol{W}^{-1} = \mathrm{diag}(\boldsymbol{\epsilon}) \tag{8}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots \epsilon_n)$. Stacking all observations y_i into a vector $\boldsymbol{y} = (y_1, y_2, \ldots y_n)$ and all empirical model outputs $\boldsymbol{Z} = (Z_1, Z_2, \ldots Z_n)$, then the matrix form of Equation 1 can be written as

$$\boldsymbol{y} = \boldsymbol{Z}d + \boldsymbol{Rc} + \boldsymbol{\epsilon} \tag{9}$$

$\boldsymbol{y}$ satisfies a multi-variate normal (MVN) distribution with a mean of $\boldsymbol{Z}d$ and a standard deviation of $\boldsymbol{RQ}^{-1}\boldsymbol{R}^{\mathrm{T}} + \boldsymbol{W}^{-1}$

$$\boldsymbol{y} \sim \mathrm{MVN}(\boldsymbol{Z}d, \boldsymbol{RQ}^{-1}\boldsymbol{R}^{\mathrm{T}} + \boldsymbol{W}^{-1}) \tag{10}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector, $\boldsymbol{Z}$ is an $n \times 1$ matrix, d is a scalar, $\boldsymbol{R}$ is an $n \times$ m matrix, $\boldsymbol{c}$ is an $m \times 1$ vector, $\boldsymbol{Q}$ is an $m \times m$ matrix, $\boldsymbol{\epsilon}$ is an $n \times 1$ vector, and $\boldsymbol{W}$ is an $n \times n$ matrix. According to the general Gaussian process theory, the best linear unbiased predictions (BLUPs) of the scaling factor d and the coefficients of basis functions $\boldsymbol{c}$ are (Cressie, 1993)

$$\hat{d} = (\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{RQ}^{-1}\boldsymbol{R}^{\mathrm{T}} + \boldsymbol{W}^{-1})\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{RQ}^{-1}\boldsymbol{R}^{\mathrm{T}} + \boldsymbol{W}^{-1})\boldsymbol{y} \tag{11}$$

$$\hat{\boldsymbol{c}} = \boldsymbol{Q}^{-1}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{RQ}^{-1}\boldsymbol{R}^{\mathrm{T}} + \boldsymbol{W}^{-1})(\boldsymbol{y} - \boldsymbol{Z}\hat{d}) \tag{12}$$

where $\boldsymbol{y}$ is the vector of the observed values. Then, the reconstruction of the field at all locations is symbolically written as

$$\boldsymbol{y}' = \boldsymbol{Z}'\hat{d} + \boldsymbol{R}'\hat{\boldsymbol{c}} \tag{13}$$

The primes on $\boldsymbol{y}'$, $\boldsymbol{Z}'$, and $\boldsymbol{R}'$ indicate that the reconstructed field may be taken at different locations from the observations. Note that in Nychka et al. (2015), $\boldsymbol{R}$ and $\boldsymbol{Q}$ are specified as sparse matrices; therefore, the whole calculation can be largely accelerated through sparse matrix calculation, which makes it ideal in processing large data sets. The specific form of RBF used in this study is

$$R(s) = \begin{cases} (1-s)^6(35s^2 + 18s + 3)/3, & \text{for } 0 \le s < 1 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where s is the normalized distance between observations and basis functions. Even though Nychka et al. (2015) use RBF as the basis function, the choices of basis functions are flexible as far as the function is range limited, which enables localized fitting. The basis functions can be easily modified to accommodate for applicational needs.

2.2. Extension of Lattice Kriging to Assimilate Electric Fields

A straightforward approach for vector field modeling is to perform data assimilation separately for its components. However, for electric fields, as mentioned earlier, the measurements are only LOS components, so such independent fitting is not feasible. Therefore, to derive the electric field from its LOS component, an additional equation that relates the two orthogonal components of the vector must be used as a constraint. For electric fields, we can use the curl-free condition, which is a reasonable approximation in the ionosphere-thermosphere system (Eccles, 1998; Mayr & Harris, 1978)

$$\frac{\partial E_y}{\partial x} = \frac{\partial E_x}{\partial y} \tag{15}$$

This differential equation indicates that these two variables are not independent. Using this relation, only one unknown parameter needs to be derived despite the electric field has two components. By using this relation and projecting the vector electrical fields to the LOS direction along which the LOS drifts have measurements, the retrieval of E_x and E_y is possible and described as follows.

With the curl-free condition for electric fields, a common practice is to define a scalar potential ϕ satisfying

$$\boldsymbol{E} = -\nabla\phi \tag{16}$$

Then, the curl-free condition is automatically satisfied.

Since electric fields and electric potentials are related by partial derivatives, a natural choice to obtain the basis functions of electric fields is by taking the directional derivatives of potential fields. We choose the basis function of the potential field ϕ to be the RBF $R(\boldsymbol{x})$ following Nychka et al. (2015). We further define two functions $R_x(\boldsymbol{x})$ and $R_y(\boldsymbol{x})$, which are related to $R(\boldsymbol{x})$ by

$$R_x(\boldsymbol{x}) = -\frac{\partial R(\boldsymbol{x})}{\partial x} \tag{17}$$

$$R_y(\boldsymbol{x}) = -\frac{\partial R(\boldsymbol{x})}{\partial y} \tag{18}$$

Therefore, if the potential field is decomposed onto a set of $R_j(\boldsymbol{x})$ with coefficients c_j satisfying, then

$$\phi(\boldsymbol{x}) = \sum_{j=1}^{m} c_j R_j(\boldsymbol{x}) \tag{19}$$

Then, the components of the electric fields will follow the relation

$$E_x(\boldsymbol{x}) = -\frac{\partial\phi(\boldsymbol{x})}{\partial x} = -\sum_{j=1}^{m} c_j \frac{\partial R_j(\boldsymbol{x})}{\partial x} = \sum_{j=1}^{m} c_j R_{x,j}(\boldsymbol{x}) \tag{20}$$

$$E_y(\boldsymbol{x}) = -\frac{\partial\phi(\boldsymbol{x})}{\partial y} = -\sum_{j=1}^{m} c_j \frac{\partial R_j(\boldsymbol{x})}{\partial y} = \sum_{j=1}^{m} c_j R_{y,j}(\boldsymbol{x}) \tag{21}$$

This suggests that $R_x(\boldsymbol{x})$ and $R_y(\boldsymbol{x})$ are basis functions of $E_x(\boldsymbol{x})$ and $E_y(\boldsymbol{x})$, respectively. Equations 20 and 21 can be rewritten into a vector form as

$$\boldsymbol{E}(\boldsymbol{x}) = -\nabla\phi(\boldsymbol{x}) = -\sum_{j=1}^{m} c_j \nabla R_j(\boldsymbol{x}) \tag{22}$$

To further relate the electric field $\boldsymbol{E}(\boldsymbol{x})$ with its LOS component, we project every electric field observation onto its corresponding LOS direction $\boldsymbol{k}(\boldsymbol{x}_i)$

$$E_{\mathrm{LOS}}(\boldsymbol{x}_i) = \boldsymbol{E}(\boldsymbol{x}_i)\cdot\boldsymbol{k}(\boldsymbol{x}_i) = -\sum_{j=1}^{m} c_j \nabla R_j(\boldsymbol{x}_i)\cdot\boldsymbol{k}(\boldsymbol{x}_i) = \sum_{j=1}^{m} c_j(-\nabla R_j(\boldsymbol{x}_i)\cdot\boldsymbol{k}(\boldsymbol{x}_i)) \tag{23}$$

This is equivalent to defining a new set of observation-dependent basis functions

$$R_{\mathrm{LOS},j}(\boldsymbol{x}_i) = -\nabla R_j(\boldsymbol{x}_i)\cdot\boldsymbol{k}(\boldsymbol{x}_i) \tag{24}$$

and projecting the LOS electric field onto the new basis set

$$E_{\mathrm{LOS}}(\boldsymbol{x}_i) = \sum_{j=1}^{m} c_j R_{\mathrm{LOS},j}(\boldsymbol{x}_i) \tag{25}$$

The Gaussian process model for LOS electric field then becomes

$$\boldsymbol{y}_{\mathrm{LOS}} = \boldsymbol{Z}_{\mathrm{LOS}}\boldsymbol{d} + \boldsymbol{R}_{\mathrm{LOS}}\boldsymbol{c} + \epsilon_{\mathrm{LOS}} \tag{26}$$

$\boldsymbol{Z}_{\mathrm{LOS}}$ is the projection of background model values onto the LOS direction. $\boldsymbol{\epsilon}_{\mathrm{LOS}}$ is the measurement error of the LOS electric field. After obtaining the BLUPs of $\hat{\mathrm{d}}$ and $\hat{\boldsymbol{c}}$ using the observations of LOS electric fields ($\boldsymbol{y}_{\mathrm{LOS}}$), the potential field can be reconstructed using Equation 13.

In summary, we reduce the electric field (vector) modeling problem into the fitting of its LOS component (scalar) assuming the curl-free condition and use the fitting information of $\hat{\mathrm{d}}$ and $\hat{\boldsymbol{c}}$ from the LOS measurements to simulate electric potentials for all locations over the domain. This approach is a simplified formulation of Fan et al. (2018), which uses Helmholtz-Hodge decomposition to study a broader range of vector fields on a sphere.

2.3. Validation Test Using Synthetic Inputs

The validation of electric field Lattice Kriging is performed using an artificial two-cell potential map with an arbitrary statistical background model in a two-dimensional (2D) plane. The test domain is $-4 < x < 4$ and $-2 < y < 2$. In order to apply Equation 26 to estimate the BLUPs of $\hat{\mathrm{d}}$ and $\hat{\boldsymbol{c}}$, we need to input $\boldsymbol{y}_{\mathrm{LOS}}$ (LOS electric fields), $\boldsymbol{Z}_{\mathrm{LOS}}$ (LOS projection of the background model), and $\boldsymbol{\epsilon}_{\mathrm{LOS}}$.

$\boldsymbol{y}_{\mathrm{LOS}}$ is obtained from a reference potential ϕ, which is a combination of two cells with equal magnitudes centered symmetrically around origin

$$\phi(x, y) = \frac{1}{1 + (x-1)^2 + y^2} - \frac{1}{1 + (x+1)^2 + y^2} \tag{27}$$

Then, the two electric field components are

$$E_x(x, y) = -\frac{\partial \phi(x, y)}{\partial x} = 2\left(\frac{x-1}{\left[1 + (x-1)^2 + y^2\right]^2} - \frac{x+1}{\left[1 + (x+1)^2 + y^2\right]^2}\right) \tag{28}$$

$$E_y(x, y) = -\frac{\partial \phi(x, y)}{\partial y} = 2\left(\frac{y}{\left[1 + (x-1)^2 + y^2\right]^2} - \frac{y}{\left[1 + (x+1)^2 + y^2\right]^2}\right) \tag{29}$$

The locations of inputs into the fitting model are taken randomly within the whole test domain. The azimuth angle θ_i at each location is also randomly chosen from 0 to 2π. Then, the LOS direction is

$$\boldsymbol{k}(x_i, y_i) = \boldsymbol{e}_x \cos\theta_i + \boldsymbol{e}_y \sin\theta_i \tag{30}$$

The projection of the electric field on the LOS direction is

$$E_{\mathrm{LOS}}(x_i, y_i) = \boldsymbol{E}(x_i, y_i) \cdot \boldsymbol{k}(x_i, y_i) \tag{31}$$

Such LOS electric fields are then fed into the model as $\boldsymbol{y}_{\mathrm{LOS}}$. In real applications, these chosen LOS electric fields correspond to the spatially scattered LOS observations. All error terms ($\boldsymbol{\epsilon}_{\mathrm{LOS}}$) are taken as identity for simplicity.

In this validation test, the background potential model is taken as a linear function in the x direction

$$Z_\phi(x, y) = x \tag{32}$$

Then, the background LOS electric field model ($\boldsymbol{Z}_{\mathrm{LOS}}$) is

$$Z_{\mathrm{LOS}}(x_i, y_i) = -\nabla Z_\phi(x_i, y_i) \cdot \boldsymbol{k}(x_i, y_i) = -\boldsymbol{e}_x \cdot (\boldsymbol{e}_x \cos\theta_i + \boldsymbol{e}_y \sin\theta_i) = -\cos\theta_i \tag{33}$$

Following Equation 26, we obtain the BLUPs of the model parameters $\hat{\mathrm{d}}$ and $\hat{\boldsymbol{c}}$, then we simulate potential over the whole domain out of the obtained $\hat{\mathrm{d}}$ and $\hat{\boldsymbol{c}}$ using Equation 13. The simulated potential ϕ' is then compared with the reference potential ϕ (Figure 1).

Figure 1a shows the reference potential field over the whole domain. Figure 1b shows LOS electric fields chosen for the model fitting. The background potential model is shown in Figure 1c. The fitted potential is depicted in Figure 1d, with its uncertainty/standard deviation (SD) shown in Figure 1e. Figure 1f is the fitting error, which is defined as the difference between output (Figure 1d) and reference (Figure 1a). Note that both the SD and fitting error are magnitudes smaller than the field itself. The small errors and the agreement between original (reference) and fitted potentials confirm the validity of extending Lattice Kriging to assimilate vector fields.

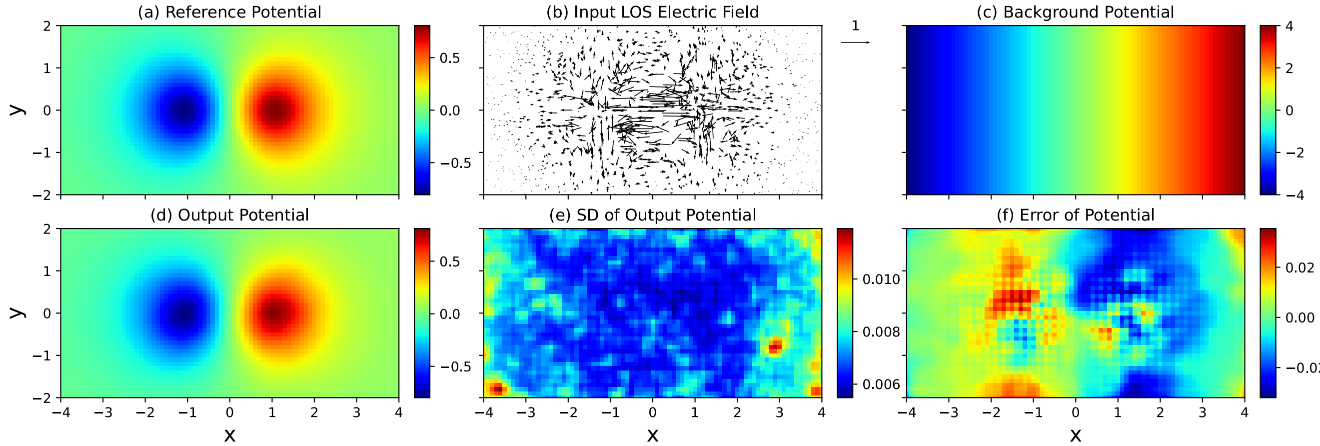

Figure 1. (a) Reference electric potential; (b) model inputs of LOS electric fields at selected locations; (c) background potential model; (d) fitted potential; (e) standard deviation (SD) of fitted potential; (f) errors of fitted potential, (d–a). Units are arbitrary.

Here, the background model is essentially arbitrary, but the fitted potential ϕ' resembles the input field ϕ to a large extent. We also tried other background models in the fitting process (not shown here), and the results are similar. This implies that with sufficient input data points available, the fitting results are not sensitive to the background model.

2.4. Multi-Resolution Modeling by Using Multiple Levels

Lattice Kriging can be used for multi-resolution modeling via implementing basis functions at multiple levels using different resolutions. We give an example of the multi-level configuration of basis functions in Figure 2. From the coarsest (yellow) to the finest (black) level, Figure 2a shows a setup of three-level basis functions in one dimension (1D). The bell-shape curves (RBFs in our case) with peak values of 1 are evenly distributed for each level. The coarsest level occupies the whole domain, and its RBFs have the longest separating distance and widest width. The finest level occupies the smallest region with the shortest distance and narrowest width of the RBFs. In the region where all the three levels overlap, the model has the highest resolution. The resolution relaxes out toward the boundary, which allows for the multi-resolution data assimilation.

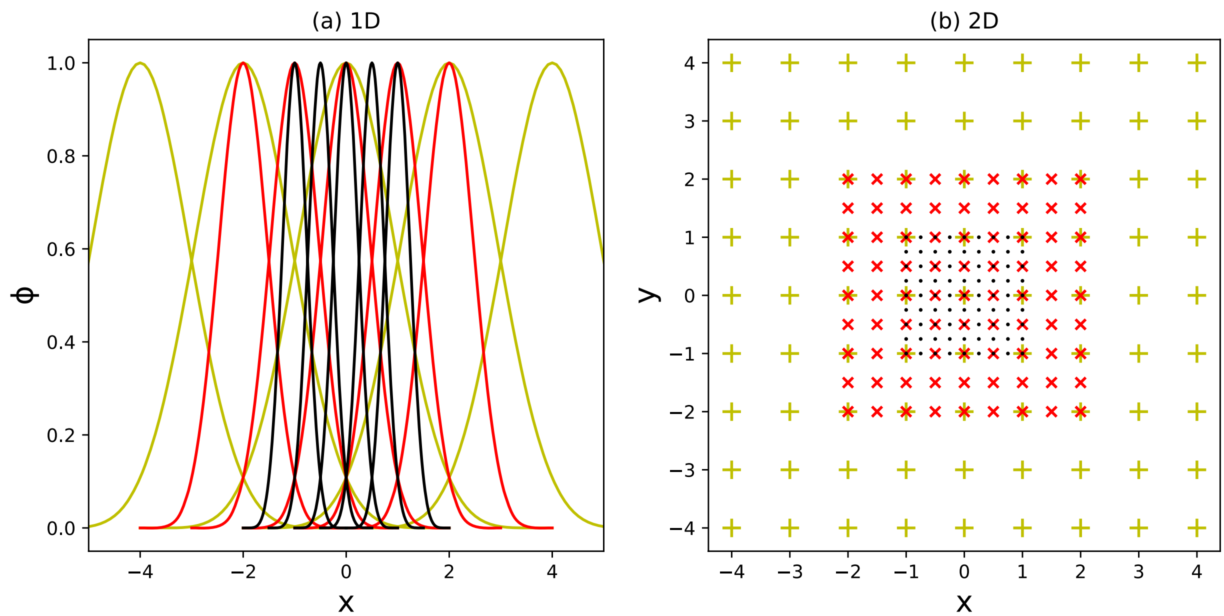

Figure 2. Examples of the configuration of multi-level basis functions in (a) 1D and (b) 2D. Units are arbitrary.

Figure 2b is an example of the 2D configuration of the multi-level basis functions. Every dot represents the center of each basis function. Similar to the 1D configuration, the basis functions are the sparsest and occupy the whole domain in the coarsest level (yellow); in the finest level (black), the basis functions are the densest and take place only at the center region of the whole domain.

Throughout the fitting process, the coefficients of every basis function for each level are obtained, which are then used for the reconstruction of the field at this level. The final fitting result is a weighted mean of the reconstructed fields at all levels. The weighting factors depend on the application and can be adjusted toward small scales by attributing large weights on fine levels or toward large scales by putting large weights on coarse levels. In real applications, the finer grids are suggested to be located at the regions with more observations.

2.5. Data Preparation and Boundary Treatment for Electric Field Assimilation

The setup of the model for assimilating real observations is similar to the validation test (Section 2.3) except that the coordinate system is different. The coordinate system used in simulating electric field is a scaled plane coordinate system on the surface of the sphere centered at the magnetic north pole. The coordinate transform from magnetic latitude (MLAT) and local time (MLT) to the model coordinate in the northern hemisphere is

$$x = \left(\frac{\pi}{2} - \frac{\pi\varphi}{180}\right)\cos\left(\frac{\pi t}{12}\right) \tag{34}$$

$$y = \left(\frac{\pi}{2} - \frac{\pi\varphi}{180}\right)\sin\left(\frac{\pi t}{12}\right) \tag{35}$$

where φ is MLAT in degree and t is MLT in hour. After the data assimilation, the inverse coordinate transform is performed to obtain magnetic coordinates.

In the high-latitude ionosphere, the plasma motion is dominated by the drift motion

$$\boldsymbol{v} = \frac{\boldsymbol{E} \times \boldsymbol{B}}{B^2} \tag{36}$$

So, all ion drift measurements used in this study are transformed to equivalent electric fields using

$$\boldsymbol{E} = -\boldsymbol{v} \times \boldsymbol{B} \tag{37}$$

For current development, the lower latitude boundary is set at 30° MLAT, where we assume the electric field vanishes. Since typical electric field observations are on the order of 10 mV/m and the errors are about 1/5–1/3 of the observations, the setup of the boundary condition consists of small values (10^{-3} mV/m) with large errors (100 mV/m). The small value term is used to force the fitting results to approach 0 at the boundary, while the large error term is to minimize the boundary impact on the fitting of the internal field (poleward of 50° MLAT), which is the focus of this study.

3. Application of Lattice Kriging to Real Observations

The validate test demonstrates that Lattice Kriging can recover the synthetic electric fields (Section 2.3); now, we apply it to real observations and examine its performance. We choose the St. Patrick's Day storm (17 March 2015) as an example, because it is the strongest geomagnetic storm in solar cycle 24 and has reasonable SuperDARN data coverage. The geomagnetic indices during the two-day storm period are shown in Figure 3. IMF B_z turns south at around 05 universal time (UT) on March 17 (Figure 3a), marking the start of the geomagnetic storm. The southward B_z lasts for almost one day before it returns to neutral at around 05UT on March 18, after which there are still minor IMF variations. IMF B_y is quite variable during the time. Figure 3b shows solar wind velocities and densities. The enhancement in the solar wind is clear during the southward B_z period. Figure 3c shows auroral electrojet (AE) indices. Strong AE variations indicate that the auroral activity is high during the time. The symmetric horizontal component of geomagnetic field (SYM-H) index is shown in Figure 3d, from which we can tell that the large substorm activity lasts until midnight in March 17 and the storm is still in the recovery phase till midnight in March 18.

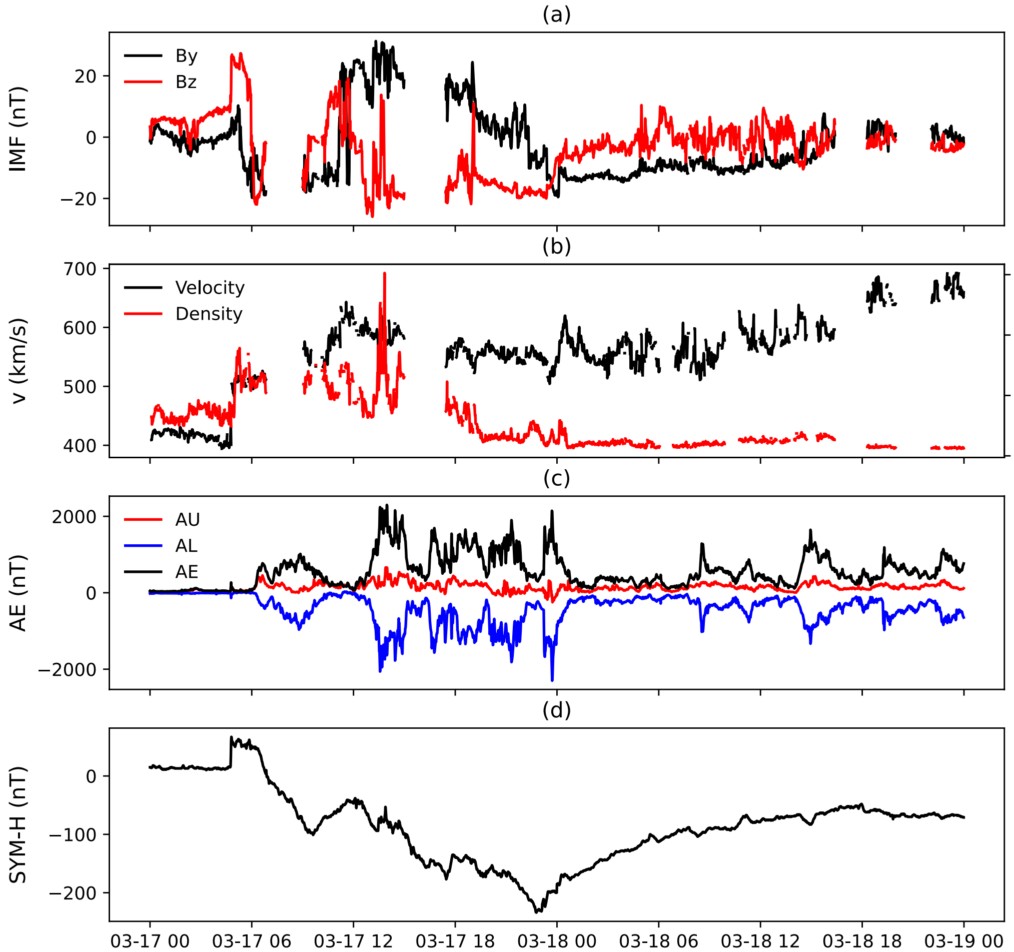

Figure 3. Geomagnetic indices on 17–18 March 2015: (a) IMF B_y and B_z, (b) solar wind velocity and density, (c) AE indices, and (d) SYM-H.

During the time, there are 19 northern hemisphere SuperDARN radars operating. The names and locations are given in the Supporting Information S1 for reference. The gridded ion velocity data are used, which have a spatial resolution of 0.5° and temporal resolution of 2 min. We will first use SuperDARN LOS ion drift data to perform data assimilation and analyze the errors using different statistical background models and resolutions in Section 3.1. Multi-level fitting is discussed in Section 3.2. Then, we add the measurements from PFISR into the data assimilation and compare the assimilated results locally (Section 3.3).

3.1. Assimilation of Electric Potential Using SuperDARN Data

We carry out six different settings for the fitting, which comes from the combination of two different background models (SuperDARN SHF and Weimer 2005 model) and three different resolutions (2°, 5°, and 8° in longitude and latitude). In all cases, the basis functions fill up the whole domain with equal distances, that is, a single resolution assimilation for each setting. Since the fitting domain is set as a square with edge length 60° × 2 = 120°, and the basis functions are equally spaced; the number of basis functions used is then $(120/r+1)^2$, where r is the resolution. For the three cases mentioned in this section, 256 basis functions are used for the 8° case, 625 basis functions are used for the 5° case, and 3,721 basis functions are used for the 2° case. Figure 4a shows SuperDARN LOS ion drift measurements at 09:37 UT on March 17. In this study, we examine the impact of the background model to the data assimilation results by using SuperDARN SHF potential maps versus Weimer model, which are shown in Figures 4b and 4c, respectively. Both SuperDARN SHF and Weimer potentials give two-cell patterns of similar magnitudes. In SuperDARN SHF potential, the peak magnitude is about 30 kV, while in the Weimer model, the peak magnitude is slightly larger at about 40 kV. The positive cell of SuperDARN SHF potential is

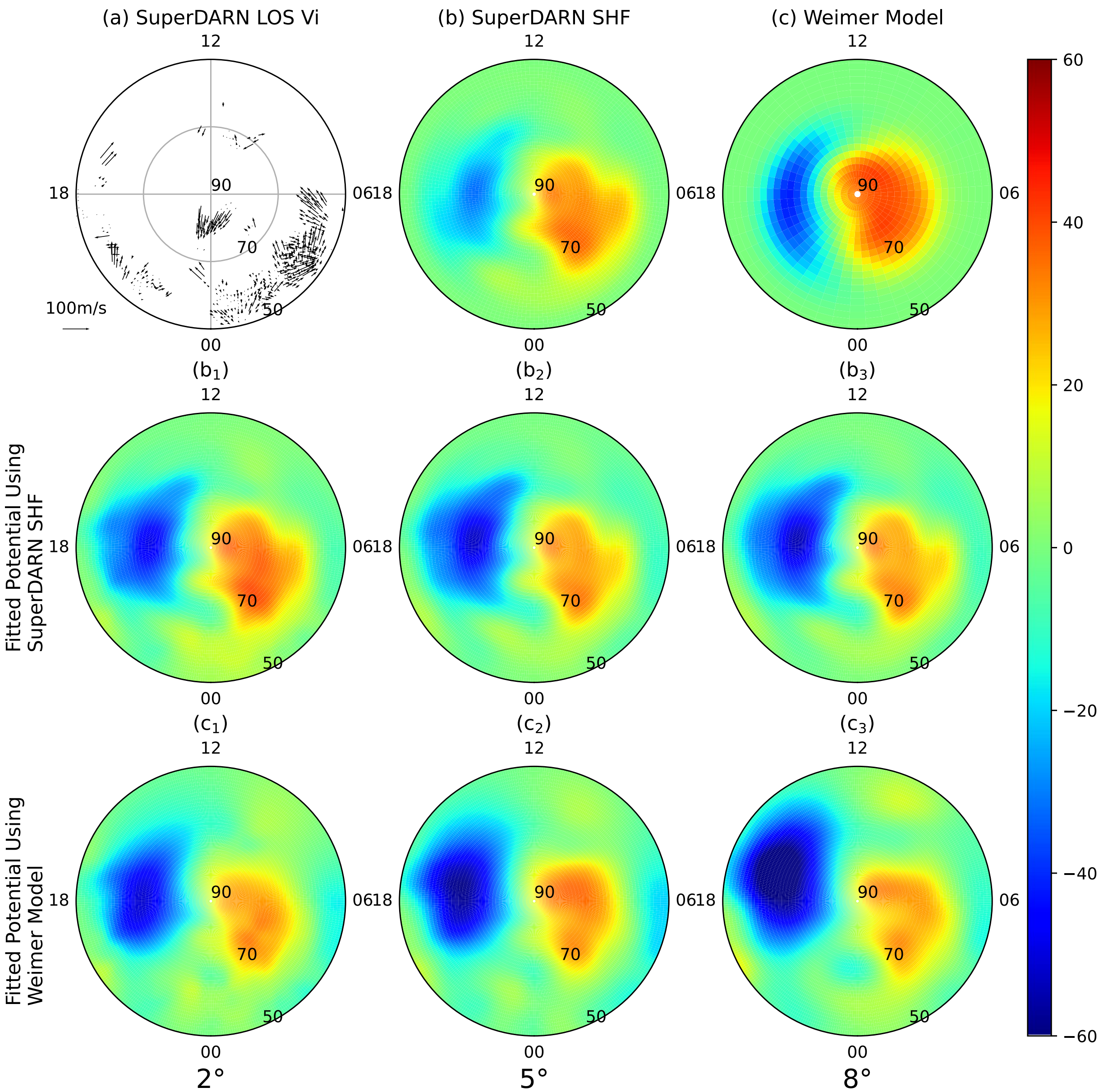

Figure 4. (a) SuperDARN LOS ion drift measurements, (b) SuperDARN SHF potential, (c) Weimer potential model, (b_1–b_3) fitted potentials using SuperDARN SHF potential as background model, (c_1–c_3) fitted potentials using the Weimer model as background model. Units are mV/m for electric fields and kV for potentials.

located at a lower MLAT compared to Weimer model. SuperDARN SHF potential shows more spatial variations while the Weimer model is more uniform. The background potential maps are used to derive LOS electric fields and used as $\mathbf{Z}_{\mathrm{LOS}}$ in Equation 26 for the assimilation procedure.

Figures 4b_1–4b_3 and Figures 4c_1–4c_3 show assimilated electric potentials for the six different settings with SuperDARN SHF and Weimer as background models, respectively. Using the same background model, the fitted results of different resolutions are generally similar except with slightly different magnitudes. A cross comparison of different background models (e.g., Figure 4b_1 vs. Figure 4c_1, Figure 4b_2 vs. Figure 4c_2, and Figure 4b_3 vs. Figure 4c_3) shows that the fitted potential is smoother and the negative cell has a larger amplitude using Weimer model than SuperDARN SHF, even though the general two-cell structure is similar.

To examine how the fitting results are related to the background model, we compare Figures 4b_1–4b_3 with Figure 4b (SuperDARN SHF) and Figures 4c_1–4c_3 with Figure 4c (Weimer model). Figures 4b_1–4b_3 are similar to Figure 4b to a large extent due to the fact that SuperDARN SHF is largely fitted upon SuperDARN LOS ion

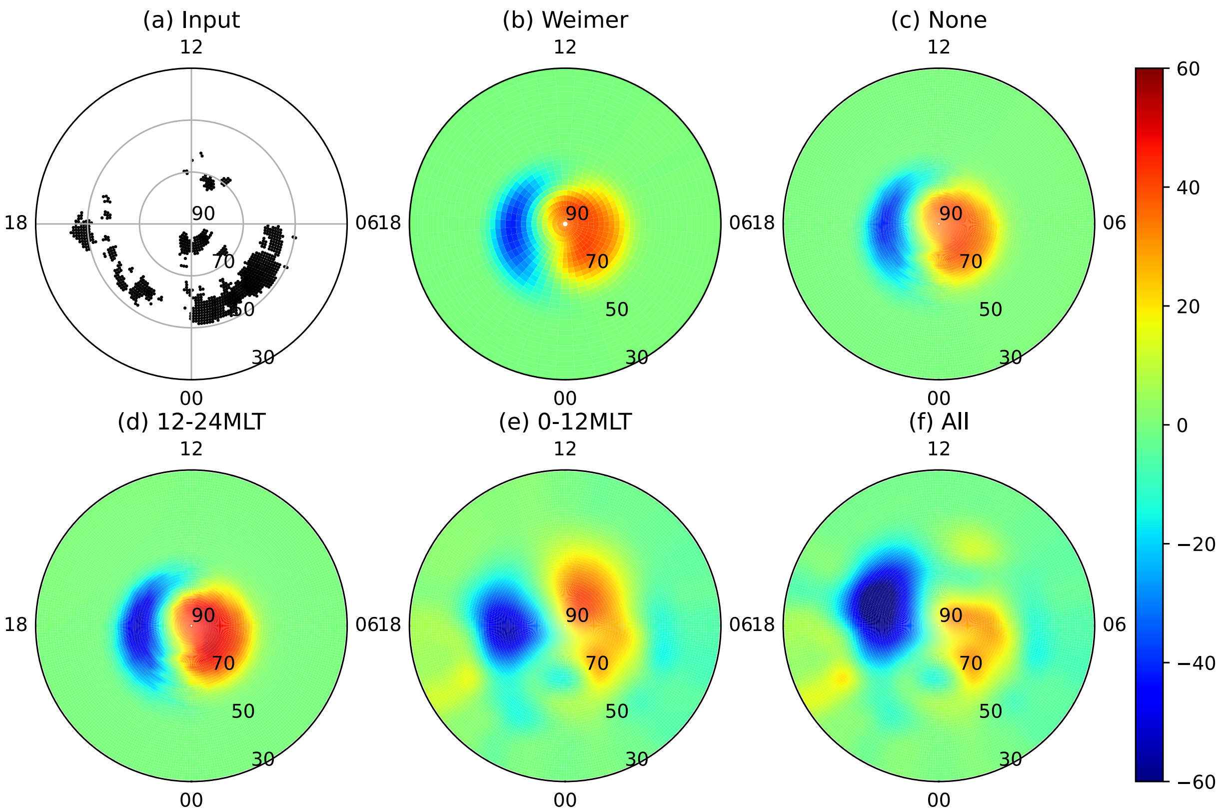

Figure 5. (a) Input data coverage at 09:37 UT, (b) electric potentials from the Weimer model, (c) fitting results with no input data, (d) fitting results with data only from 12 to 24 MLT, (d) fitting results with data only from 0 to 12 MLT, (e) fitting results with all available data.

drift measurements, which has already contained a large bulk of the observational information. One noticeable change is that the negative cell becomes more negative after applying Lattice Kriging. On the other hand, from the Weimer model to the fitting result using it as a background model (Figure 4c vs. Figures $4c_1$–$4c_3$), significant differences are seen: (a) the comparable magnitudes of the positive and negatives cells are modified to a very negative cell (below −80 kV) and a mild positive cell (∼30 kV), (b) the inequal areas of the two cells are modified to be approximately equal, and (c) the negative cell is moved to a lower MLAT. Such modifications result from the fusion of the real-time observed LOS electric fields, which are not captured in the Weimer empirical model. The fitting process weighs more on the observations than the background model; therefore, the fitting results are dominated by observations whenever they are available.

From Figure 4c (original Weimer model) to Figure $4c_3$ (fitted results using Weimer as background model), the potential pattern changes significantly even in regions where data coverage is sparse, for example, the post-noon sector from 12 to 18 MLT. To examine how the input data impact the fitting process, we perform several tests by varying the total amount of data inputs. The fitting results in 8° using the Weimer background model are shown in Figure 5. In the extreme case that no data are inputted into the model, the fitting result (Figure 5c) is literally the same as the background model. In this case, no data are incorporated into the fitting process, and the result converges to the background model with $d = 1$ and $\boldsymbol{c} = 0$ in Equation 9, which further validates our method. In Figure 5d where we only include sparse observations from 12 to 24 MLT, the fitting result still appears close to the background model. As more data are used (from Figure 5d to 5f), the fitting results gradually deviate from the background model and become more determined by the data. This test suggests that data can impose significant influences on the background patterns in the region away from the data. In other words, the impacts of data tend to be remote and global.

To demonstrate how the fitting process introduces mesoscale variations in electric fields and ion drifts (converted from electric fields using Equation 36), the fitting results in 2° using Weimer as the background model at 09:37 UT are shown in Figures $6a_1$ and $6a_2$. Electric fields and ion drifts solely from the Weimer model are shown

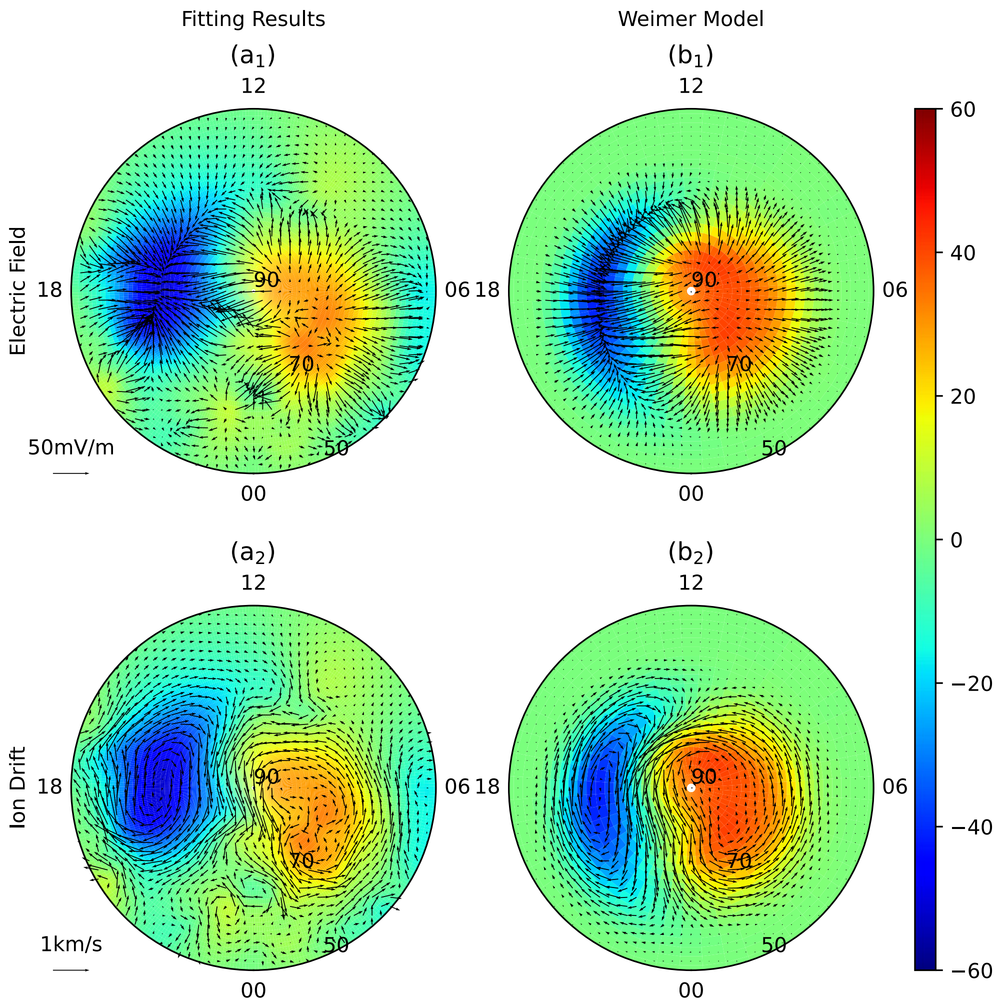

Figure 6. (a_1) Fitted electric field, (b_1) Weimer electric field, (a_2) fitted ion drift, (b_2) Weimer ion drift. Potential is overplotted in all subfigures. Arrows represent either electric field or ion drift, and the color contour represents the fitted potential. Units are mV/m for electric fields, km/s for ion drifts, and kV for potentials.

in Figures $6b_1$ and $6b_2$ for comparison. There are clear differences, and the fitted maps show more regional structures than the empirical Weimer model. For instance, around 70° MLAT and 00 MLT, the local divergence and convergence in electric fields (Figure $6a_1$) and local ion drift vortices (Figure $6a_2$) are only seen in the fitted maps. Mesoscale electric fields, which are missing in the empirical model, start to emerge when SuperDARN observations are included in the fitting process.

To better evaluate the fitting outcome and performance, we analyze the fitting error defined as the root-mean-square error (RMSE) in terms of the LOS electric field differences between fitted results and observations.

$$\varepsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(-\nabla\phi(\boldsymbol{x}_i)\cdot\boldsymbol{k}(\boldsymbol{x}_i) - E_{\mathrm{LOS}}(\boldsymbol{x}_i)\right)^2} \tag{38}$$

where n is the total points of observations, $\phi(\boldsymbol{x}_i)$ is the fitted potential, and $-\nabla\phi(\boldsymbol{x}_i)\cdot\boldsymbol{k}(\boldsymbol{x}_i)$ gives its projection onto the fitted LOS electric field. $E_{\mathrm{LOS}}(\boldsymbol{x}_i)$ is the observed LOS electric field.

Figure 7 shows comparisons of the fitting errors from all cases and the errors solely from background models are also shown as a reference. Figures $7a_1$ and $7b_1$ show fitting errors at 09:37 UT. In Figure $7a_1$, the fitting errors decrease as resolutions increase and the errors are all smaller than those from SuperDARN SHF, which means that the fitting process successfully reduces the errors and captures more information in the data as resolution increases. The fitting error of using 2° resolution decreases by 30% compared with the background model. In

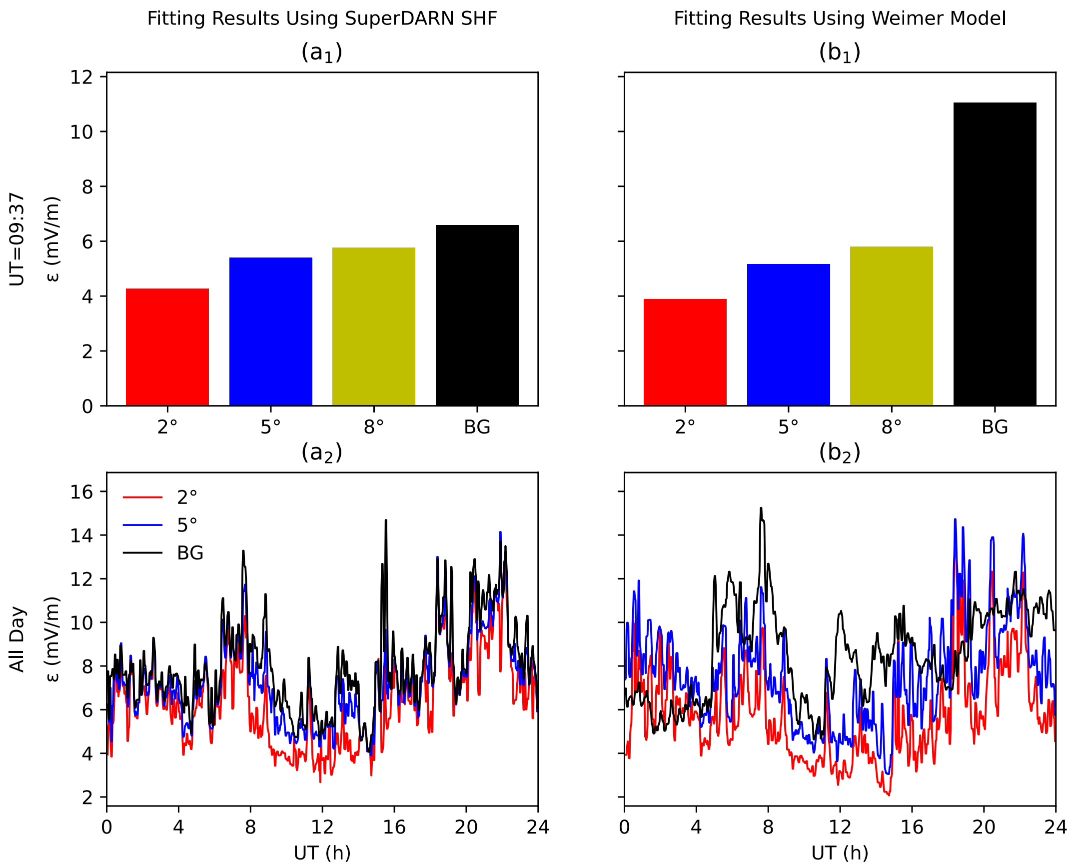

Figure 7. RMSE comparison of the fitting results: (a_1) RMSE of the LOS electric fields using SuperDARN SHF as background model at 09:37 UT; (b_1) same as (a_1) except for using the Weimer model as the background model. (a_2 and b_2) RMSE during the whole day using SuperDARN SHF and Weimer as background models, respectively. The errors of Weimer model are divided by 2 in (b_2) for displaying purposes. Red, blue, and yellow colors are for the fitting results under the resolutions of 2°, 5°, and 8°, respectively, while black is for the background model.

Figure 7b_1, the fitting errors decrease more than half compared with the Weimer model and more significant improvements are seen in the higher resolution assimilation. Cross comparing Figure 7a_1 with Figure 7b_1 in each resolution, the fitting errors are of the similar magnitude no matter which background model is invoked. This confirms that the fitting results are not sensitive to the background model when sufficient observations are available.

Figures 7a_2 and 7b_2 show the fitting errors during the whole day (the results of 2° and 5° are given as examples). The RMSE in Weimer model is divided by 2 for displaying purposes in Figure 7b_2. In general, Lattice Kriging reduces the LOS electric field error when 2° and 5° of fitting resolutions are used. In Figure 7a_2, the fitting results using SuperDARN SHF decrease the error by more than 30% during 07:00 and 15:00 UT, while in Figure 7b_2, the fitting results using the Weimer background model decrease by more than half during most of the time. The fitting errors decrease with increasing fitting resolutions in both cases in accordance with Figures 7a_1 and 7b_1. This implies that more structures in data are captured by using higher resolutions and Lattice Kriging generally performs better than the fitting using global harmonics.

3.2. Multi-Resolution Assimilation of Electric Fields Using SuperDARN Data

To demonstrate the capability of multi-resolution data assimilation (Section 2.4), we set up two-level basis functions to perform the fitting and analyze the results (Figure 8a). The basis functions of the coarse level are separated by 5° and cover the whole domain; those of the fine level are separated by 2° and only cover half of the domain. Note that for the pure 2° and 5° cases, a single resolution (either 2° or 5°) is used throughout the whole domain. In Figure 8a, the outer circle marks the low-latitude boundary at 30° MLAT, and the middle circle marks

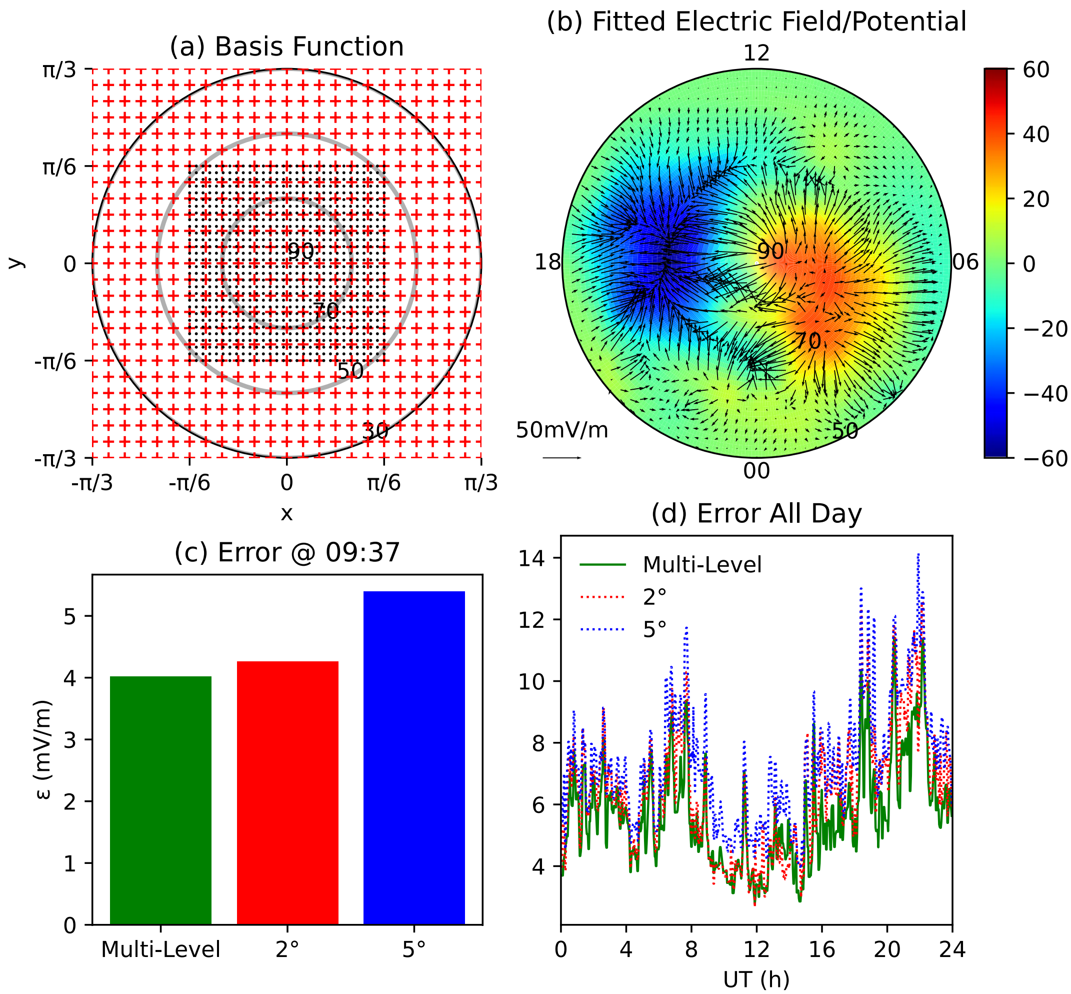

Figure 8. Multi-resolution fitting case: (a) Two-level basis function setup, (b) fitted potential and electric field, (c) comparison of fitting errors with 2° and 5° cases, and (d) fitting errors for all day. Units are mV/m for electric fields and kV for potentials. The correspondence of *x* and *y* in (a) with MLT and MLAT can be found in Section 2.5.

the 50° MLAT circle, which is of our concern and outputted. The multi-level fitting region (inner square in Figure 8a) largely overlaps with the output domain (middle circle in Figure 8a). Such configuration (referred to as fixed-hybrid fitting) is designed to take advantage of the most available SuperDARN observations, which take place mostly poleward of 50° MLAT.

Figure 8b shows the fitted potential and electric field at 09:37 UT. Comparing with Figures $4b_1$–$4b_3$, the large and mesoscale structures are similar. The fitting error shown in Figure 8c is slightly smaller than the 2° fitting error, which indicates that the fixed-hybrid fitting can better simulate the mesoscale structures of the electric field. Figure 8d shows the fitting errors for the whole day. Again, the fixed-hybrid fitting has the lowest errors throughout most of the time. We list the averaged RMSEs of LOS electric fields throughout the day for all cases in Table 1. The top row indicates the resolution used for each case and the first left column indicates the selection of the background model. The daily mean error from the background model itself is listed in the last column. Using SuperDARN SHF as the background model, the fixed-hybrid fitting decreases the RMSE by 29%, and the pure 2° case decreases by 19%, compared with the background model itself. Using Weimer model, these two decreases are 68% and 66%, respectively.

Table 1
Daily Means of RMSEs of LOS Electric Fields for All Cases (Unit Is mV/m)

	Fixed-hybrid	Auto-hybrid	2°	5°	8°	BG
SuperDARN	5.71	6.51	6.44	7.37	7.64	7.99
Weimer	5.46	6.45	5.80	7.67	8.81	17.11

Here, the multi-level case applies the high-resolution grids to a fixed region, but the real measurements and data coverage usually change with time. Applying high-resolution grids to the region without data coverage causes a

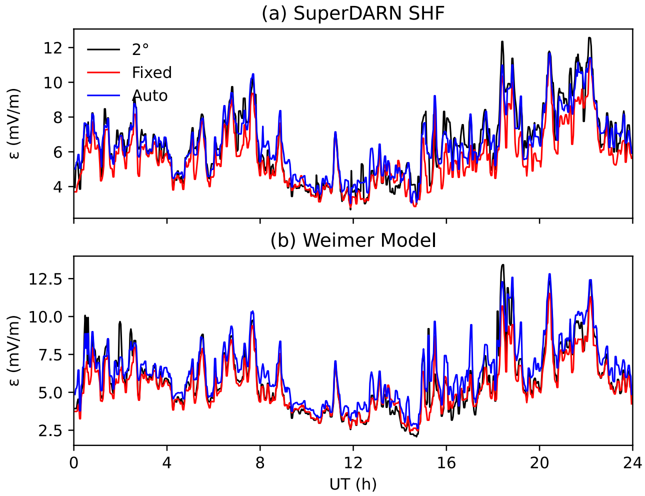

Figure 9. Fitting errors (RMSE) using (a) SuperDARN SHF as background model, (b) the Weimer model as background model. Red lines show fitting errors from the fixed multi-level setup, blue lines show fitting errors from the auto adjusted multi-level setup, and black lines show fitting errors from the uniform 2° setup for comparison.

waste of computing time and no substantial improvement is guaranteed. We perform a new test case with an auto-hybrid fitting approach (distinguished from the aforementioned fixed-hybrid fitting). The new setup of the auto-hybrid fitting consists of two levels: the low-resolution (5°) level covers the whole domain, while the high-resolution (2°) level is implemented in the region with observations, that is, determined by the data. We apply an auto adjustment to the high-resolution level every time to collocate the basis functions with observations. From the fixed-hybrid fitting to this auto-hybrid fitting, since the high-resolution basis functions in regions where observations are unavailable are removed, it has fewer basis functions than the fixed-hybrid fitting. The fitting error of the auto-hybrid fitting is shown in Figure 9 as the blue line. The fitting errors from the uniform 2° basis setup (black line) and the fixed-hybrid fitting (red line) are also plotted in the figure. The errors from the auto-hybrid fitting are only slightly larger than the fixed-hybrid fitting and of similar magnitude as the 2° case.

It is worth mentioning that the computation time for the fixed-hybrid fitting is ~40% shorter than the pure 2° case, and the auto-hybrid fitting enables a time decrease of ~80%, which suggests that with a proper setup of basis functions, the auto-hybrid fitting is likely the most affordable and efficient choice for the regional high-resolution assimilation.

3.3. Local Electric Field Modeling Using PFISR Data

On March 17, PFISR was operating in several different modes (IPY27_Tracking, Freg732, LTCS35, and WorldDay35) during the day, and ion drifts from all modes over Poker Flat as available are used to derive electric fields (Equation 37) for the entire storm period. The derived PFISR electric fields are then fed into the data assimilation model in combination with SuperDARN measurements to obtain the electric fields. The spatial resolution of PFISR measurements is 0.25° and the temporal resolution is typically every 1–2 min.

Shown in Figure 10 is the 2° fitted electric fields at Poker Flat when using SuperDARN SHF as the background model. PFISR measurements and the electric field from SuperDARN SHF are also plotted for comparison. Roughly speaking, the background model agrees well with PFISR measurements though some large fluctuations are missing (e.g., E_x at ~08:30, 12:00 UT and E_y at ~06:30 UT, after 11:00 UT). The fitted eastward and northward electric fields from Lattice Kriging, instead, capture such large fluctuations and follow more closely to the real PFISR observations. For example, the positive E_y peak at ~06:30 UT underestimated in SuperDARN SHF is largely elevated to a comparable level to PFISR measurements, while at ~07:30 UT the peak only present in the background model is attenuated and becomes more comparable with PFISR by the fitting process. Nevertheless, there are time periods where the fitting results still deviate from the observations such as E_x after 12:00 UT. The fitting results during these periods might be influenced by the ambient SuperDARN LOS electric field measurements, which show differences from PFISR.

In summary, Lattice Kriging can capture the short-period variations shown in the local data and meanwhile maintain the coherence to the ambient electric fields to a large extent. Compared with the method of just padding the local observations into the background model, Lattice Kriging avoids the problem of discontinuity and largely fuses the information from real local observations.

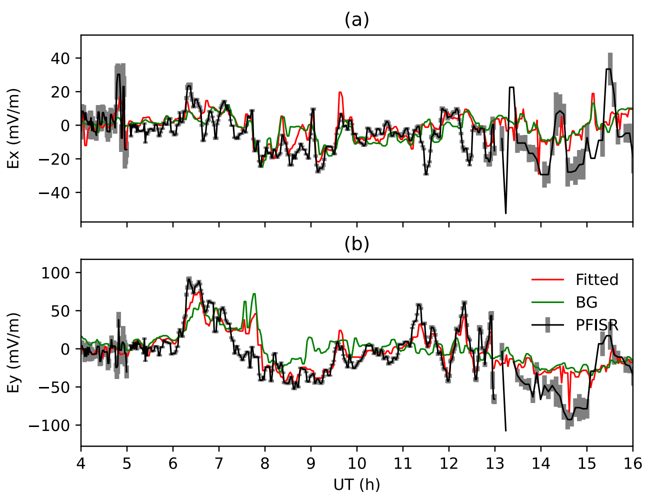

Figure 10. (a) Eastward and (b) northward electric fields at Poker Flat. The dotted black lines are PFISR observations, blue lines are the SuperDARN SHF, and red lines are the Lattice Kriging fitted results.

4. Conclusions and Outlook

This paper develops a new methodology to assimilate high-latitude electric fields via extending the Lattice Kriging framework to vector fields. This modeling assumes that the fitted field is a Gaussian process. By combining the background model (SuperDARN SHF or Weimer model in our case) used as "a priori" knowledge and the available observations (SuperDARN and PFISR), the means, variances, and covariances of the to-be-assimilated variables (electric fields and ion drifts) are calculated and further used to reconstruct the fields where no observations are available. By doing so, this methodology assimilates the observational data and provides the fitted results for the whole domain of interest.

We systematically evaluate the performance of Lattice Kriging using different resolutions and background models. We first assimilate the electric fields for the St. Patrick's Day storm (17 March 2015) applying the SuperDARN data. We find that when the same amount of data is provided, the fitting results are similar whichever background model is used, suggesting that data is a more important factor than the background model in fitting the electric fields. By varying the amount of input data, we find that the impacts of data tend to be remote and reach the regions without observations. Compared with the background model, data assimilation leads to considerable decreases in the RMSE of the LOS electric field. Such improvement is more significant against the Weimer model than SuperDARN SHF likely because the former is empirical, while the latter has incorporated some of the data information already. Comparing across the different fitting resolutions (2°, 5°, and 8° in our case), higher resolution always leads to smaller RMSE, suggesting that more details in the observations are captured with more basis functions (higher resolutions) used in the fitting model.

We demonstrate the capability and advantages of the multi-resolution modeling using multi-level basis functions. 2° in the fine level and 5° in the coarse level are adopted to form up a two-level framework as an example. Two types of configuration of the high-resolution (2° in this case) grids are tried. The fixed-hybrid grid covers the largest region that all the data appear and use it as a fixed region to do the high-resolution fitting. The auto-hybrid grid adjusts the high-resolution region according to the real-time data coverage. The fixed-hybrid fitting errors further decrease compared with the pure 2° case and the computation time shortens by 40%, while the auto-hybrid fitting has a similar fitting error as the pure 2° case and the decrease of computation time reaches ~80%, suggesting that (a) the multi-level basis function can further improve the fitting, (b) using the relatively coarse grid in the region without observations does not degrade the performance and saves computational cost, and (c) the auto-hybrid fitting provides an efficient way to perform the regionally high-resolution data assimilation. Using the SuperDARN SHF as the background model, the multi-level assimilation decreases the RMSE by 29%, and 2° case decreases by 19%, compared with the background model itself. By using the multi-level basis function setup (especially auto-hybrid fitting), high-resolution observations can be better assimilated with affordable computational resources.

Even though the multi-level basis function setup can effectively reduce errors compared with other setups, the errors are still substantial (daily mean of 5.71 mV/m for the auto-hybrid fitting and 6.51 mV/m for the fixed-hybrid fitting). To further reduce the errors, there are two possibilities: (a) The covariance matrix is to be improved. The covariance matrix used here is derived from a Gaussian Markov random field (which assumes two locations are correlated only if they are adjacent, Nychka et al., 2015). However, electric potentials/fields are not necessarily uncorrelated even if they are apart for some distance. Strictly, an additional term indicating the medium range electric field correlation needs to be included in the covariance matrix to describe the real-world electric field characteristics and (b) the mathematical formulation is to be modified based on non-Gaussian process models. The current development simply assumes the electric field is a Gaussian process, while in real world, the distribution of electric fields deviates from Gaussian (Golovchanskaya & Kozelov, 2010a, 2010b). Still, Gaussian statistics has good properties for fast computation, such as the sparse matrix calculation as aforementioned, which satisfies as a starting point.

Further, we include the PFISR observations in the model and the fitting results at Poker Flat better capture the short-period electric field variations shown in the data than those from the global spherical harmonics fitting (i.e., SuperDARN SHF). This indicates that our method can efficiently fuse and then recover the local measurements, and importantly, maintain the coherence of the patterns with the ambient electric fields.

The decreases in RMSE, flexibility of incorporating various data sources, and the benefits of the multi-level setup embedded in Lattice Kriging show that it is a powerful tool in the data assimilation application. The application of such method is not limited to electric field and ion drift, but can also be applied to other physical quantities, such as field-aligned current as a scalar field and wind as a vector field. For electric fields, we use the curl-free assumption to provide an additional constrain for the modeling. For neutral winds in relatively large scales (e.g., planetary and synoptic scales), the vertical gradient of the vertical wind is negligible and the horizontal winds are approximately divergence free. A stream function is well defined with the divergence-free assumption. Defining basis functions in a similar way as mentioned in Section 2, the modeling of the stream function using horizontal wind measurements can be similarly formulated. The new Michelson Interferometer for Global High-resolution Thermospheric Imaging (MIGHTI) instrument onboard Ionospheric Connection Explorer (ICON) provides neutral wind measurements over the midlatitude and low-latitude regions, which may provide an optimal data set to assimilate the multi-resolution structures of neutral dynamics using Lattice Kriging.

Data Availability Statement

SuperDARN data are obtained from http://vt.superdarn.org and PFISR data are obtained from https://amisr.com. The code of Lattice Kriging for electric fields is published at https://github.com/hzfywhn/LatticeKriging. The data used to produce the figures are available at https://data.mendeley.com/datasets/n6jsffz4n7.

Acknowledgments
The authors thank Xueling Shi from Virginia Tech, Yukitoshi Nishimura from Boston University, and Steve Kaeppler from Clemson University for providing observational data and insightful discussions. The authors are grateful for the advices from Qiong Zhang and Whitney Huang at Clemson University. This work is supported by NASA NNX17AG10G and NSF grants AGS-1705448, OPP-1705450, and CAREER-1753214.

References

Amm, O., Grocott, A., Lester, M., & Yeoman, T. K. (2010). Local determination of ionospheric plasma convection from coherent scatter radar data using the SECS technique. *Journal of Geophysical Research: Space Physics*, *115*(A3), A03304. https://doi.org/10.1029/2009JA014832

Bristow, W. A., Hampton, D. L., & Otto, A. (2016). High-spatial-resolution velocity measurements derived using local divergence-free fitting of SuperDARN observations. *Journal of Geophysical Research: Space Physics*, *121*(2), 1349–1361. https://doi.org/10.1002/2015JA021862

Bristow, W. A., Sibeck, D. G., Jacquey, C., Greenwald, R. A., Sofko, G. J., Mukai, T., et al. (1995). Observations of convection vortices in the afternoon sector using the SuperDARN HF radars. *Journal of Geophysical Research: Space Physics*, *100*(A10), 19743–19756. https://doi.org/10.1029/95JA01301

Buchau, J., Reinisch, B. W., Weber, E. J., & Moore, J. G. (1983). Structure and dynamics of the winter polar cap F region. *Radio Science*, *18*(6), 995–1010. https://doi.org/10.1029/RS018i006p00995

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. *Geophysical Research Letters*, *22*(17), 2393–2396. https://doi.org/10.1029/95GL01909

Cosgrove, R. B., & Codrescu, M. (2009). Electric field variability and model uncertainty: A classification of source terms in estimating the squared electric field from an electric field model. *Journal of Geophysical Research: Space Physics*, *114*(A6), A06301. https://doi.org/10.1029/2008JA013929

Cousins, E. D. P., & Shepherd, S. G. (2012a). Statistical characteristics of small-scale spatial and temporal electric field variability in the high-latitude ionosphere. *Journal of Geophysical Research: Space Physics*, *117*(A3), A03317. https://doi.org/10.1029/2011JA017383

Cousins, E. D. P., & Shepherd, S. G. (2012b). Statistical maps of small-scale electric field variability in the high-latitude ionosphere. *Journal of Geophysical Research*, *117*(A12), A12304. https://doi.org/10.1029/2012JA017929

Cressie, N. A. C. (1993). Statistics for spatial data. In *Statistics for spatial data. Wiley Series in Probability and Statistics* (pp. 1–26). John Wiley and Sons Incorporation. https://doi.org/10.1002/9781119115151.ch1

Eccles, J. V. (1998). Modeling investigation of the evening prereversal enhancement of the zonal electric field in the equatorial ionosphere. *Journal of Geophysical Research: Space Physics*, *103*(A11), 26709–26719. https://doi.org/10.1029/98JA02656

Fan, M., Paul, D., Lee, T. C. M., & Matsuo, T. (2018). Modeling tangential vector fields on a sphere. *Journal of the American Statistical Association*, *113*(524), 1625–1636. https://doi.org/10.1080/01621459.2017.1356322

Golovchanskaya, I. V., & Kozelov, B. V. (2010a). On the origin of electric turbulence in the polar cap ionosphere. *Journal of Geophysical Research*, *115*(A9), A09321. https://doi.org/10.1029/2009JA014632

Golovchanskaya, I. V., & Kozelov, B. V. (2010b). Properties of electric turbulence in the polar cap ionosphere. *Geomagnetism and Aeronomy*, *50*(5), 576–587. https://doi.org/10.1134/S001679321005004X

Hanuise, C., Senior, C., Cerisier, J.-C., Villain, J.-P., Greenwald, R. A., Ruohoniemi, J. M., & Baker, K. B. (1993). Instantaneous mapping of high-latitude convection with coherent HF radars. *Journal of Geophysical Research: Space Physics*, *98*(A10), 17387–17400. https://doi.org/10.1029/93JA00813

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guhaniyogi, R., Gerber, F., et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 398–425. https://doi.org/10.1007/s13253-018-00348-w

Hsu, C.-T., Matsuo, T., Maute, A., Stoneback, R., & Lien, C.-P. (2021). Data-driven ensemble modeling of equatorial ionospheric electrodynamics: A case study during a minor storm period under solar minimum conditions. *Journal of Geophysical Research: Space Physics*, *126*(2), e2020JA028539. https://doi.org/10.1029/2020JA028539

Kozelov, B. V., & Golovchanskaya, I. V. (2006). Scaling of electric field fluctuations associated with the aurora during northward IMF. *Geophysical Research Letters*, *33*(20), L20109. https://doi.org/10.1029/2006GL027798

Lu, G. (2017). Large scale high-latitude ionospheric electrodynamic fields and currents. *Space Science Reviews*, *206*(1), 431–450. https://doi.org/10.1007/s11214-016-0269-9

Lu, G., Zakharenkova, I., Cherniak, I., & Dang, T. (2020). Large-scale ionospheric disturbances during the 17 March 2015 storm: A model-data comparative study. *Journal of Geophysical Research: Space Physics*, *125*(5), e2019JA027726. https://doi.org/10.1029/2019JA027726

Maimaiti, M., Ruohoniemi, J. M., Baker, J. B. H., & Ribeiro, A. J. (2018). Statistical study of nightside quiet time midlatitude ionospheric convection. *Journal of Geophysical Research: Space Physics*, *123*(3), 2228–2240. https://doi.org/10.1002/2017JA024903

Matsuo, T. (2020). Recent progress on inverse and data assimilation procedure for high-latitude ionospheric electrodynamics. In *Ionospheric multi-spacecraft analysis tools: Approaches for deriving ionospheric parameters*. *ISSI Scientific Report Series* (pp. 219–232). Springer International Publishing. https://doi.org/10.1007/978-3-030-26732-2_10

Mayr, H. G., & Harris, I. (1978). Some characteristics of electric field momentum coupling with the neutral atmosphere. *Journal of Geophysical Research: Space Physics*, *83*(A7), 3327–3336. https://doi.org/10.1029/JA083iA07p03327

Nishimura, Y., Deng, Y., Lyons, L. R., McGranaghan, R. M., & Zettergren, M. D. (2021). Multiscale dynamics in the high-latitude ionosphere. In *Ionosphere dynamics and applications*. *Geophysical Monograph Series* (pp. 49–65). American Geophysical Union. https://doi.org/10.1002/9781119815617.ch3

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., & Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, *24*(2), 579–599. https://doi.org/10.1080/10618600.2014.914946

Richmond, A. D. (1992). Assimilative mapping of ionospheric electrodynamics. *Advances in Space Research*, *12*(6), 59–68. https://doi.org/10.1016/0273-1177(92)90040-5

Richmond, A. D., & Kamide, Y. (1988). Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique. *Journal of Geophysical Research*, *93*(A6), 5741–5759. https://doi.org/10.1029/JA093iA06p05741

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

Sánchez, E. R., Ruohoniemi, J. M., Meng, C.-I., & Friis-Christensen, E. (1996). Toward an observational synthesis of substorm models: Precipitation regions and high-latitude convection reversals observed in the nightside auroral oval by DMSP satellites and HF radars. *Journal of Geophysical Research: Space Physics*, *101*(A9), 19801–19837. https://doi.org/10.1029/96JA00363

Sheng, C., Deng, Y., Zhang, S.-R., Nishimura, Y., & Lyons, L. R. (2020). Relative contributions of ion convection and particle precipitation to exciting large-scale traveling atmospheric and ionospheric disturbances. *Journal of Geophysical Research: Space Physics*, *125*(2), e2019JA027342. https://doi.org/10.1029/2019JA027342

Weber, E. J., Buchau, J., Moore, J. G., Sharber, J. R., Livingston, R. C., Winningham, J. D., & Reinisch, B. W. (1984). F layer ionization patches in the polar cap. *Journal of Geophysical Research: Space Physics*, *89*(A3), 1683–1694. https://doi.org/10.1029/JA089iA03p01683

Weimer, D. R. (2005). Improved ionospheric electrodynamic models and application to calculating Joule heating rates. *Journal of Geophysical Research: Space Physics*, *110*(A05), 0148–0227. https://doi.org/10.1029/2004JA010884

Wiens, A., Nychka, D., & Kleiber, W. (2020). Modeling spatial data using local likelihood estimation and a Matérn to spatial autoregressive translation. *Environmetrics*, *31*(6), e2652. https://doi.org/10.1002/env.2652

Wu, H., Lu, X., Lu, G., Chu, X., Wang, W., Yu, Z., et al. (2020). Importance of regional-scale auroral precipitation and electrical field variability to the storm-time thermospheric temperature enhancement and inversion layer (TTEIL) in the Antarctic E region. *Journal of Geophysical Research: Space Physics*, *125*(9), e2020JA028224. https://doi.org/10.1029/2020JA028224

Zhang, J. J., Wang, W., Wang, C., Lan, A. L., Yan, J. Y., Xiang, D., et al. (2020). First observation of ionospheric convection from the Jiamusi HF radar during a strong geomagnetic storm. *Earth and Space Science*, *7*(1), e2019EA000911. https://doi.org/10.1029/2019EA000911