Space Weather

RESEARCH ARTICLE
10.1029/2020SW002561

Reconstruction of Extreme Geomagnetic Storms: Breaking the Data Paucity Curse

M. I. Sitnov[1], G. K. Stephens[1], N. A. Tsyganenko[2], H. Korth[1], E. C. Roelof[1], P. C. Brandt[1], V. G. Merkin[1], and A. Y. Ukhorskiy[1]

[1]Applied Physics Laboratory, The Johns Hopkins University, Laurel, MD, USA, [2]Institute and Department of Physics, Saint Petersburg State University, Saint Petersburg, Russia

Key Points:
- Reconstruction of the magnetic field, electric current, and pressure distributions is made for strong storms using a new data mining method
- The use of weighted nearest neighbors in data mining significantly reduces the statistical bias for strong storms
- The pressure peak for strong storms may be as close as $\approx 2.1 R_E$ to Earth and its value may exceed 300 nPa

Supporting Information:
- Supporting Information S1

Correspondence to:
M. I. Sitnov,
Mikhail.Sitnov@jhuapl.edu

Citation:
Sitnov, M. I., Stephens, G. K., Tsyganenko, N. A., Korth, H., Roelof, E. C., Brandt, P. C., et al. (2020). Reconstruction of extreme geomagnetic storms: Breaking the data paucity curse. *Space Weather*, *18*, e2020SW002561. https://doi.org/10.1029/2020SW002561

Received 10 JUN 2020
Accepted 25 SEP 2020
Accepted article online 30 SEP 2020

Abstract Reconstruction of the magnetic field, electric current, and plasma pressure is provided using a new data mining (DM) method with weighted nearest neighbors (NN) for strong storms with the storm activity index *Sym-H* < -300 nT, the Bastille Day event (July 2000), and the 20 November 2003 superstorm. It is shown that the new method significantly reduces the statistical bias of the original NN algorithm toward weaker storms. In the DM approach the magnetic field is reconstructed using a small NN subset of the large historical database, with the subset number $K_{NN} \gg 1$ being still much larger than any simultaneous multiprobe observation number. This allows one to fit with observations a very flexible magnetic field model using basis function expansions for equatorial and field-aligned currents, and at the same time, to keep the model sensitive to storm variability. This also allows one to calculate the plasma pressure by integrating the quasi-static force balance equation with the isotropic plasma approximation. For strong storms of particular importance becomes the resolution of the eastward current, which prevents the divergence of the pressure integral. It is shown that in spite of the strong reduction of the dominant NN number in the new weighted NN algorithm to capture strong storm features, it is still possible to resolve the eastward current and to retrieve plasma pressure distributions. It is found that the pressure peak for strong storms may be as close as $\approx 2.1 R_E$ to Earth and its value may exceed 300 nPa.

Plain Language Summary Extreme geomagnetic storms represent the strongest perturbations of the magnetosphere, a gigantic cavity provided by the Earth's magnetic field in the flow of charged particles coming from our star, the solar wind. Their description using in situ observations with spaceborne magnetometers is strongly complicated by the sparsity of these observations and by the extreme rarity of these events. Here we test and demonstrate potency of a new technique, in which one forms a swarm of synthetic probes composed of historical records that were made at the moment most closely similar to the event of interest in terms of the solar wind impact, the global state of the magnetosphere and its trend. Those records are additionally weighted to increase the impact of the most similar events. The resulting reconstructions are shown to be more consistent with observations available at the moment of interest and global images of the magnetosphere.

1. Introduction

Understanding key physical processes in the magnetosphere during extreme geomagnetic storms (Buzulukova et al., 2018; Ebihara et al., 2005; Lakhina et al., 2013; Riley & Love, 2017) is exceedingly challenging for both modeling and observations. With the few exceptions (Ngwira et al., 2014), first-principle simulations of superstorms with *Sym-H* < -300 nT, such as the July 2000 Bastille Day event (Chen et al., 2003; Jordanova et al., 2001; Liemohn et al., 2002; Raeder, Wang, et al. 2001) and the 20 November 2003 storm (Ebihara et al., 2005; Fok et al., 2011), remain limited to magnetohydrodynamic (MHD) or standalone kinetic ring current simulations. At the same time, empirical investigation of extreme events (EEs) cannot be done using standard statistical methods, because EEs represent, by definition, tails of the corresponding data distributions (Riley & Love, 2017). In case of in situ observations, such as the magnetic field measurements, the problem of empirical reconstruction is further complicated because the Earth's magnetosphere is extremely sparsely sampled by spaceborne magnetometers with fewer than a dozen probes available at any moment.

To reconstruct the magnetic field and electric currents for strong storms, Tsyganenko et al. (2003) compiled a special storm time subset of multiyear spaceborne magnetometer records. It was used to fit an empirical magnetic field model with complex structure describing the main current systems and their amplitude coefficients being complex nonlinear functions of the solar wind and interplanetary magnetic field (IMF) input parameters. More recently, Sitnov et al. (2008) proposed to mine data in the whole multiyear database of K_{DB} records to select at any moment of interest a subset $K_{NN} \ll K_{DB}$ of the most relevant historical observations with the number being much larger than any simultaneous multiprobe observation number $K_{NN} \gg 1$. This allowed one to fit with observations a very flexible magnetic field model using basis function expansions for equatorial currents (Tsyganenko & Sitnov, 2007) and, at the same time, to keep the model sensitive to storm variability. The latter was described by the average value of the storm time index *Sym-H*, its time derivative, and the solar wind input parameter vB_z^{IMF}, where $-v$ is the X component of the solar wind velocity and B_z^{IMF} is the Z component of the IMF in the GSM coordinate system. The new empirical reconstruction algorithm called TS07D was successfully applied to investigate storms with different solar wind drivers (Sitnov et al., 2008, 2010).

Later it was found that the proposed data mining (DM) algorithms and basis function representations of the magnetic field provide TS07D (and its later substorm counterpart SST19 Stephens et al., 2019) with the property of *machine-learning tools* (e.g., Kubat, 2015): Given more data in the database and/or more dimensions in the description of the state of the magnetosphere, they provide more details about the magnetosphere structure and evolution. In particular, Stephens et al. (2016) showed that increasing the size of the model's database and the number of degrees of freedom for equatorial currents helps resolve more details of the ring current, including its eastward component and its closure paths with the main (westward) partial ring current forming a horseshoe- or banana-like current structure (Liemohn et al., 2013; Roelof, 1989). Application of this approach to the field-aligned current (FAC) system (Sitnov et al., 2017) revealed the spiral structure of the low-altitude FAC distributions, which can be independently reconstructed using low-altitude spacecraft observations (Iijima & Potemra, 1976).

The resolution of the eastward current during storms plays a key role in the reconstruction of storm time distributions of the plasma pressure (Sitnov et al., 2018), because it limits the pressure obtained by integrating the quasi-static isotropic plasma force balance equation $\nabla p = \mathbf{j} \times \mathbf{B}$ (Sergeev et al., 1994; Stephens et al., 2013). This reconstruction is so important because the plasma pressure distribution can be compared with the first-principle models (Wiltberger et al., 2017). Moreover, it can be ingested into the first-principles MHD models to adjust their storm time equation of state (Merkin et al., 2019), which is otherwise ad hoc (adiabatic) and hard to constrain by observations. Storm time pressure distributions derived from the empirical magnetic field can also be directly compared (Sitnov et al., 2018; Stephens et al., 2020) with the pressure profiles obtained by combining the energy channels in plasma instruments of the Van Allen Probes mission (Mauk et al., 2013).

In the DM approach (Sitnov et al., 2008), the magnetic field is reconstructed using a small subset of the large database of $K_{DB} \sim 4 \cdot 10^6$ historical magnetometer records. The subset with $1 \ll K_{NN} \ll K_{DB}$ is composed of the database points closest to the event of interest in the Euclidean space formed by the average *Sym-H* index, its time derivative and the solar wind input parameter vB_z^{IMF}. These closest points are called the nearest neighbors or NNs (Cover & Hart, 1967; Vassiliadis, 2006) and they are used as an instantaneous event-oriented database to fit the magnetic field model (Tsyganenko & Sitnov, 2007) at the moment of interest. Since $K_{NN} \gg 1$, this model may have many degrees of freedom to resolve, in particular, the storm time ring current, including its eastward component (Stephens et al., 2016).

However, for EEs, the NN subset has a strong bias toward weaker events because they are more prevalent, and as a result, the DM method may strongly underestimate the ring current strength and the plasma pressure (Sitnov et al., 2018). This problem cannot be resolved by reducing the K_{NN} number, because it forces one to reduce the number of degrees of freedom of the magnetic field model in order to avoid overfitting. As a result, the resolution of the model is degraded. Eventually, it may become impossible to resolve the eastward current and to limit the empirical pressure integral (Stephens et al., 2013). The fundamental underlying cause of this problem is the extreme paucity of in situ observations, which is in drastic contrast to the main class of machine learning problems that deal with data abundance (e.g., LeCun et al., 2015). This problem was realized already by Tsyganenko and Sitnov (2007) who noticed that with conventional binning approaches, getting a clearer spatial picture of the magnetosphere forces one to sacrifice the resolution of

its dynamics and variability. It becomes particularly compelling for EEs due to their peculiar variability and the necessity to resolve the ring current pressure peak, which most closely approaches the Earth for the strongest storms. Here we show that the use of weighted NNs in the DM algorithm (e.g., Mitchell, 1997) implemented for TS07D in Stephens et al. (2020) resolves this problem to a large extent and breaks the data paucity curse.

The structure of the paper is as follows. In section 2 we describe the magnetic field model architecture and the new weighted NN algorithm. In section 3 the results of application of this DM algorithm are presented for the July 2000 Bastille Day event and the 20 November 2003 superstorm. This includes validation using available probes, statistical analysis of the NN bins, distributions of the magnetic field, electric currents, low-altitude FACs, and plasma pressure. The results of this study are summarized and further discussed in section 4. Additional results of the error analysis are provided in the supporting information for this paper.

2. Methodology

2.1. Magnetic Field Model Architecture

In this study, following Tsyganenko and Sitnov (2007), we present the magnetic field of all equatorial current systems, including tail and ring currents, in the form

$$\mathbf{B}_{(ec)}(\rho,\phi,z)=\sum_{n=1}^{N}a_{0n}^{(s)}\mathbf{B}_{0n}^{(s)}+\sum_{m=1}^{M}\sum_{n=1}^{N}\left(a_{mn}^{(o)}\mathbf{B}_{mn}^{(o)}+a_{mn}^{(e)}\mathbf{B}_{mn}^{(e)}\right) \tag{1}$$

Here (ρ,ϕ,z) are cylindrical coordinates in a system with the z axis normal to the equatorial plane, $\mathbf{B}_{\alpha\beta}^{(\gamma)}$ are the basis functions and $a_{\alpha\beta}^{(\gamma)}$ are their weights in the expansion. The structure (1) follows from the solution of the Ampère's equation for the magnetic field of an infinitely thin current sheet (CS) above and below the equatorial plane $z=0$ by separation of variables. As was shown by Tsyganenko and Sitnov (2007), the solution contains axially symmetric radial expansions (the first group of terms in Equation 1 marked by the index (s)), as well as even and odd in azimuthal angle ϕ groups marked by the indices (e) and (o), respectively. The basis functions of the solution for the vector potential with an infinitely thin CS contain factors like $\exp(-k|z|)$. Their regularization assuming the finite CS half-thickness D can be provided by replacing the function $|z|$ by the smooth function $\zeta=\sqrt{z^2+D^2}$.

An example of the basis functions used in (1) can be given by the azimuthal component A_ϕ of the vector potential corresponding to the first group of basis functions $\mathbf{B}_{0n}^{(s)}$: $\left(A_\phi\right)_{0n}^{(s)}=J_1\left(k_n\rho\right)\exp\left(-k_n\sqrt{z^2+D^2}\right)$, where J_1 is the Bessel function of the first order, $k_n=n/\rho_0$, and ρ_0 is the radial scale, corresponding to the largest mode in the radial expansion. The parameters ρ_0, N, and M are fixed because they determine the adopted resolution of the expansions (1). Other parameters, such as the weights $a_{\alpha\beta}^{(\gamma)}$ and the CS thickness D, are determined from fitting the model to data. Thus, the spatial resolution of such an expansion is determined by the number of terms in (1) and can be increased to any desired level, commensurate with the data density. Each element in (1) is independently shielded (has its own subsystem of Chapman-Ferraro-type currents at the magnetopause). The solar wind dynamic pressure P_{dyn} is incorporated into the equatorial current system by replacing the scaling coefficient $a_{\alpha\beta}^{(\gamma)}$ with two groups of coefficients, one of which is a function of the dynamic pressure $a_{0,\alpha\beta}^{(\gamma)}+a_{1,\alpha\beta}^{(\gamma)}\sqrt{P_{dyn}}$, doubling the number of scaling coefficients listed in (1).

To take into account seasonal and diurnal variations of the Earth's dipole tilt angle Ψ with respect to the $X=0$, resulting in a periodic transverse motion and large-scale deformation of the tail current sheet (Russell, 1972), as well as the CS warping and twisting effects, the original tail magnetic field is deformed following the procedures described in Tsyganenko (2002). These deformations are controlled by the nonlinear parameters, the hinging distance R_H, the warping parameter G, and the twisting parameter TW, to be derived like the CS thickness D from data (since in the following we do not consider the effects of the finite B_y^{IMF} magnetic field, twisting effects are ignored and $TW=0$).

The model also includes a flexible FAC system (Sitnov et al., 2017), necessary to reproduce the spiral FAC structure at low latitudes (Iijima & Potemra, 1976) whose nightside part is likely associated with the Harang discontinuity (Harang, 1946). Each element of the FAC system is described as the magnetic field of two deformed conical surfaces corresponding to Region 1 (R1) and Region 2 (R2) FACs (Tsyganenko, 2002). The size of each system is an adjustable parameter, while their azimuthal distribution is controlled by the

relative contributions of two groups of basis functions with odd and even symmetry due to factors $\sin(l\phi)$ and $\cos(l\phi)$, $(l = 1, 2, \dots)$. The first group represents the main part of the FAC system, in which the duskside currents have the same magnitude but opposite direction to those at dawn, while the second group has an even distribution of currents with respect to the noon-midnight meridian plane, which allows one to model the azimuthal rotation of the FACs. In TS07D the FAC system was described by the first and second harmonics ($l = 1, 2$) with the odd ($\sin(l\phi)$) symmetry for R1 currents and the first harmonic ($l = 1$) with both odd and even and symmetry for R2 allowing its azimuthal rotation. Sitnov et al. (2017) proposed to use modules similar to the original TS07D FAC basis functions shifted in latitude to fill the possible gap between R1 and R2 distributions. The FAC modules are shielded similar to equatorial currents.

In this study, we use FAC systems with $N_{FAC} = 6$ and 12 basis functions. In the first case, we follow the original (Sitnov et al., 2017) scheme described above. In the second case, both the original R1, R2, and the intermediate in latitude FAC systems consist of four elements: the first and second harmonics ($l = 1, 2$) with the odd ($\sin(l\phi)$) symmetry and the first harmonic ($l = 1$) with both odd and even and symmetry. The magnetic field of all FAC basic elements is then deformed, to take into account the day-night asymmetry and tilt angle effects, and further rescaled using the formula $\mathbf{B}'(\mathbf{r}) = \kappa\mathbf{B}(\kappa\mathbf{r})$ with the nonlinear data-derived parameter κ for each latitudinal group of basis functions, consistent with the original algorithms proposed by Tsyganenko (2002).

The resulting model magnetic field of all external currents to be fit with data after its summation with the internal field $\mathbf{B}_{(int)}$ (the near-dipole field described by the IGRF model Thébault et al., 2015) can now be summarized in the form:

$$\mathbf{B}_{(ext)}(x, y, z) = \mathbf{B}_{(ec)} + \mathbf{B}^{(sh)}_{(ec)} + \mathbf{B}_{(fac)} + \mathbf{B}^{(sh)}_{(fac)} + \mathbf{B}^{(sh)}_{(int)}, \tag{2}$$

where $\mathbf{B}_{(ec)}$ is the magnetic field of near-equatorial currents (including tail and ring currents) described by (1), its magnetopause shielding counterpart $\mathbf{B}^{(sh)}_{(ec)}$, the corresponding combination of fields $\mathbf{B}_{(fac)}$ and $\mathbf{B}^{(sh)}_{(fac)}$ for the FAC system, as well as the magnetopause shielding part $\mathbf{B}^{(sh)}_{(int)}$ for the IGRF field also known as the Chapmen-Ferraro field (see for more details Sitnov et al., 2017; Stephens et al., 2020; Tsyganenko & Sitnov, 2007).

2.2. Mining Data With Weighted Nearest Neighbors

In the original DM method employed before (Sitnov et al., 2008, 2012; Stephens et al., 2019), the magnetic field model (2) is trained by adjusting its coefficients using the small NN subset of $K_{NN} \sim 10^4$ moments in time, "most relevant to the event of interest," from the database of $K_{DB} \sim 10^6$ historical magnetometer records. The relevance is determined by the distance of the NN from the moment of interest in the space of global state and input parameters of the magnetosphere. Based on the previous correlation analysis (Sitnov et al., 2012) and limiting our study by storm scales we define the following coordinates of this global parameter space:

$$G_1(t) = \langle Sym\text{-}H^* | \propto \int_{-\Pi_{st}/2}^{0} Sym\text{-}H^*(t+\tau)\cos(\pi\tau/\Pi_{st})\mathrm{d}\tau \tag{3}$$

$$G_2(t) = D\langle Sym\text{-}H^* | /Dt \propto \int_{-\Pi_{st}/2}^{0} Sym\text{-}H^*(t+\tau)\sin(2\pi\tau/\Pi_{st})\mathrm{d}\tau \tag{4}$$

$$G_3(t) = \langle vB_z^{IMF} | \propto \int_{-\Pi_{st}/2}^{0} vB_z^{IMF}(t+\tau)\cos(\pi\tau/\Pi_{st})\mathrm{d}\tau \tag{5}$$

Here $Sym\text{-}H^* = A \cdot Sym\text{-}H - B \cdot \sqrt{P_{dyn}}$ is the pressure-corrected *Sym-H* index (Tsyganenko, 1996), P_{dyn} is the solar wind dynamic pressure (in nPa), and the values of A and B are taken to be 0.8 and 13.0, respectively. The functions G_1 and G_3 describe weighted moving averages over the past moments of the index $Sym\text{-}H^*$ and the solar wind electric field parameter vB_z^{IMF}, while G_2 describes a smoothed time derivative of $Sym\text{-}H^*$. The asymmetric shape of the time averaging operators $\langle \dots |$ reflects the fact that the window function only averages over past data (for more details on the filter functions, see Sitnov et al., 2012, and in particular their

Figures 1 and 2). The averaging scale parameter $\Pi_{st} = 12$ hr reflects the characteristic storm scales and helps exclude substorm effects which are outside the scope of this study.

The NN subset in the 3-D Euclidean space with coordinates G_1–G_3 is formed by points $\mathbf{G}^{(i)} = [G_1(t_i), G_2(t_i), G_3(t_i)]$, $i = 1, \ldots K_{NN}$, that are closest to the query point (moment of interest) $\mathbf{G}^{(q)} = [G_1(t_q), G_2(t_q), G_3(t_q)]$ by the metric

$$R_q^{(i)} = \sqrt{\sum_{k=1}^{3} \left(G_k^{(i)} - G_k^{(q)} \right)^2 / \sigma_{G_k}^2}. \tag{6}$$

and hence are most similar to the event of interest. Here σ_{G_k} is the standard deviation of the component G_k.

In TS07D, the selected set of K_{NN} points in the global parameter space $\mathbf{G}$ is then used to find in the database of K_{DB} magnetometer measurements a subset of observations $K_{NN}^{(B)}$ (whose size is usually larger than K_{NN} by the average number of probes available at the fixed moment), which are most relevant to the event of interest. Then the subset $K_{NN}^{(B)}$ is used to fit the flexible magnetic field model (2) in the real space. Thus, the DM method in TS07D differs from simple KNN regressions (e.g., Mitchell, 1997) and more general conditional averages (Vassiliadis, 2006, and references therein) in that the model fitting (regression) is made not in the binning space but in the real space. The real-space fitting of the magnetic field for the NN subset of $K_{NN}^{(B)}$ magnetometer measurements is made using a combination of the least squares fitting for linear coefficients of the field model (for instance, $a_{\alpha\beta}^{(\gamma)}$ in (1)) and the downhill simplex method (Nelder & Mead, 1965) for nonlinear model parameters (e.g., the hinging distance R_H). To regularize the linear regression procedure and achieve a trade-off between the accuracy of the fitting and the noise in the best fit coefficients, the singular value decomposition (SVD) method (Press et al., 1992) was applied following Tsyganenko and Sitnov (2007). In the SVD approach to the regression problem (Press et al., 1992), the least squares normal equation matrix is presented as a product of two orthogonal matrices and one diagonal matrix containing positive numbers, the so-called singular values. Then, in the process of the matrix inversion the smallest singular values are excluded, so that their inverse values are replaced by zeros. The number of singular values to be excluded is controlled by the tolerance parameter *TOL*, which is the ratio between the smallest and the largest singular values to be retained. The specific value of *TOL* is selected to achieve a trade-off between the accuracy of the fitting and the noise in the best fit coefficients. In our present studies we use the tolerance value $TOL = 10^{-6}$, which safely eliminates the noise in inversions and which provides the accuracy far better than the accuracy of our basis function expansions described in the previous section.

For this study we use the multimission database covering more than two decades of observations (1995–2017) and described in more detail in Sitnov et al. (2019) and Stephens et al. (2019). It includes IMP-8, Geotail, Polar, GOES-08, GOES-09, GOES-10, GOES-12, Cluster, THEMIS, Van Allen Probes, and MMS missions with the total number of $K_{DB} = 3{,}668{,}101$ points.

Sitnov et al. (2018) found that while the use of this binning procedure gives reasonable distributions of the magnetic field, current, and plasma, for EEs it has a strong bias toward weaker events. This is seen from the comparison of black and blue lines in Figure 1. This bias can be reduced by reducing the K_{NN} value to $2 \cdot 10^3$ (gray lines in Figure 1). However, this also results in overfitting of the spatial reconstructions because the number of the data points ($K_{NN}^{(B)} \sim K_{NN}$) becomes comparable to the number of degrees of freedom (>500).

To resolve this problem, Stephens et al. (2020) implemented a distance-weighted NN algorithm (e.g., Mitchell, 1997). In this algorithm, the model magnetic field $\mathbf{B}^{(mod)} = \mathbf{B}_{(int)} + \mathbf{B}_{(ext)}$, where $\mathbf{B}_{(ext)}$, defined by (2), is determined by minimizing the RMS of its deviation from observations

$$M_{err}^{(NN)} = \sqrt{\sum_{j \in S_{NN}} \sum_{i=x,y,z} w_j w_{(0)}(r) \left[B_i^{(mod)}(\mathbf{r}^{(j)}) - B_i^{j,obs} \right]^2}, \tag{7}$$

where S_{NN} is a set of $K_{NN}^{(B)}$ magnetometer measurements of the magnetic field components $B_i^{j,obs}$ with ephemeris $\mathbf{r}^{(j)}$, corresponding to the selected set of K_{NN} nearest neighbors; $w_{(0)}$ is the original weighting factor, which is a function of the real-space distance r from Earth, introduced in Tsyganenko and Sitnov (2007) to mitigate the inhomogeneity of observations, especially at geosynchronous orbit. A distinctive feature of the weighted KNN algorithm is that each term in the sum in (7) has now an additional weighting factor

$$w_j = \exp\left[-\left(R_q^{(j)} / \sigma R_{NN} \right)^2 / 2 \right], \tag{8}$$

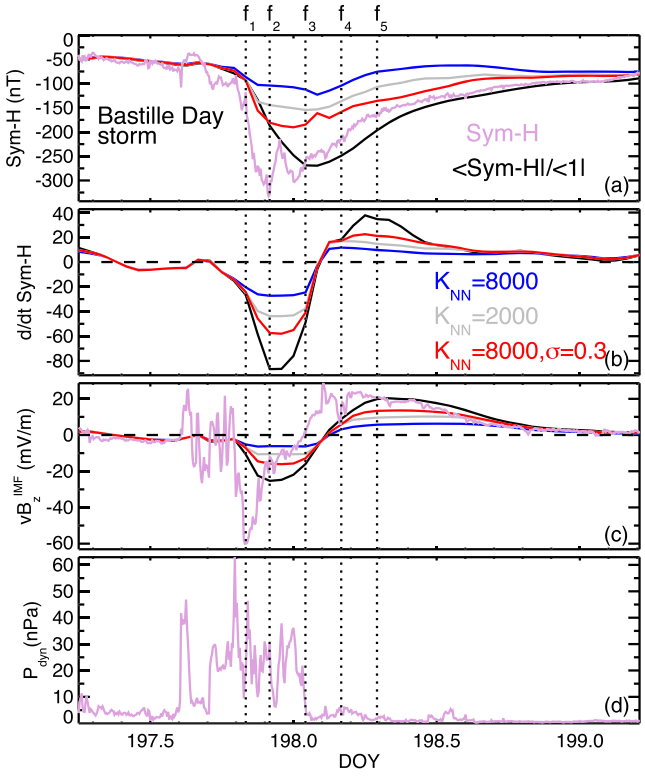

Figure 1. Analysis of the NN bins and input parameters for the Bastille Day event reconstruction: (a) The original index *Sym-H** (plum line), its nowcasting convolution $\langle Sym\text{-}H^*|$ (3) (black line) and means of the latter over bins with $K_{NN} = 8 \cdot 10^3$ (blue line) and $2 \cdot 10^3$ (gray line), as well as the mean over $K_{NN} = 8 \cdot 10^3$ weighted with $\sigma = 0.3$ (red line). Panels (b) and (c) represent analogs of panel (a) for the binning parameters $D\langle Sym\text{-}H^*|/Dt$ and $\langle vB_z^{IMF}|$. (d) The dynamical pressure P_{dyn}. Vertical dashed lines mark moments in this substorm $f_1 = 197$, 20:00 UT, $f_2 = 197$, 22:00 UT, $f_3 = 198$, 01:00 UT, $f_4 = 198$, 04:00 UT, and $f_5 = 198$, 07:00 UT, which are discussed in more detail later.

Here $R_q^{(j)}$ is the distance (6) of the corresponding NN from the query point q and R_{NN} is the radius of the sphere containing NNs in the binning space (G_1, G_2, G_3). When $\sigma \gg 1$, all distance-modulated weights $w_j \approx 1$ and NNs are not weighted. In contrast, for $\sigma < 1$, the new weighs w_j are well modulated within the sphere $R_q^{(j)} < R_{NN}$. This increases the statistical weight of measurements that were made at the more similar state and input conditions of the magnetosphere, according to the metric (6).

As one can see from Figure 1 (red lines), the use of weighted NNs with $\sigma = 0.3$ reduces the bias even stronger than the fourfold reduction of the K_{NN} value. At the same time, as it will be shown below, such a modification of the NN DM procedure keeps the model free from overfitting and it allows one to reconstruct the spatial picture of the magnetic field and currents with the resolution sufficient for determining the eastward current.

3. Reconstruction Results

3.1. Bastille Day Event Reconstruction Using Standard KNN Algorithm

The 15–16 July 2000 Bastille Day superstorm was discussed earlier in many studies (Brandt et al., 2001; Chen et al., 2003; Jordanova et al., 2001; Lepping et al., 2001; Liemohn et al., 2002; Raeder, McPherron, et al. 2001; Tsyganenko et al., 2003) and its first DM reconstruction using the standard (nonweighted) KNN algorithm has already been performed in Sitnov et al. (2018). This storm belongs to the highest G5 level in NOAA's classification of the geomagnetic activity. It was caused by the full-halo coronal mass ejection (CME), which arrived at the Earth's vicinity late on 15 July 2000, being preceded by a shock wave at 14:37 UT. The CME was associated with the X5.7 class solar flare that occurred a day before, 14 July, the National Day of France. It occurred near the peak of the solar maximum in solar cycle 23.

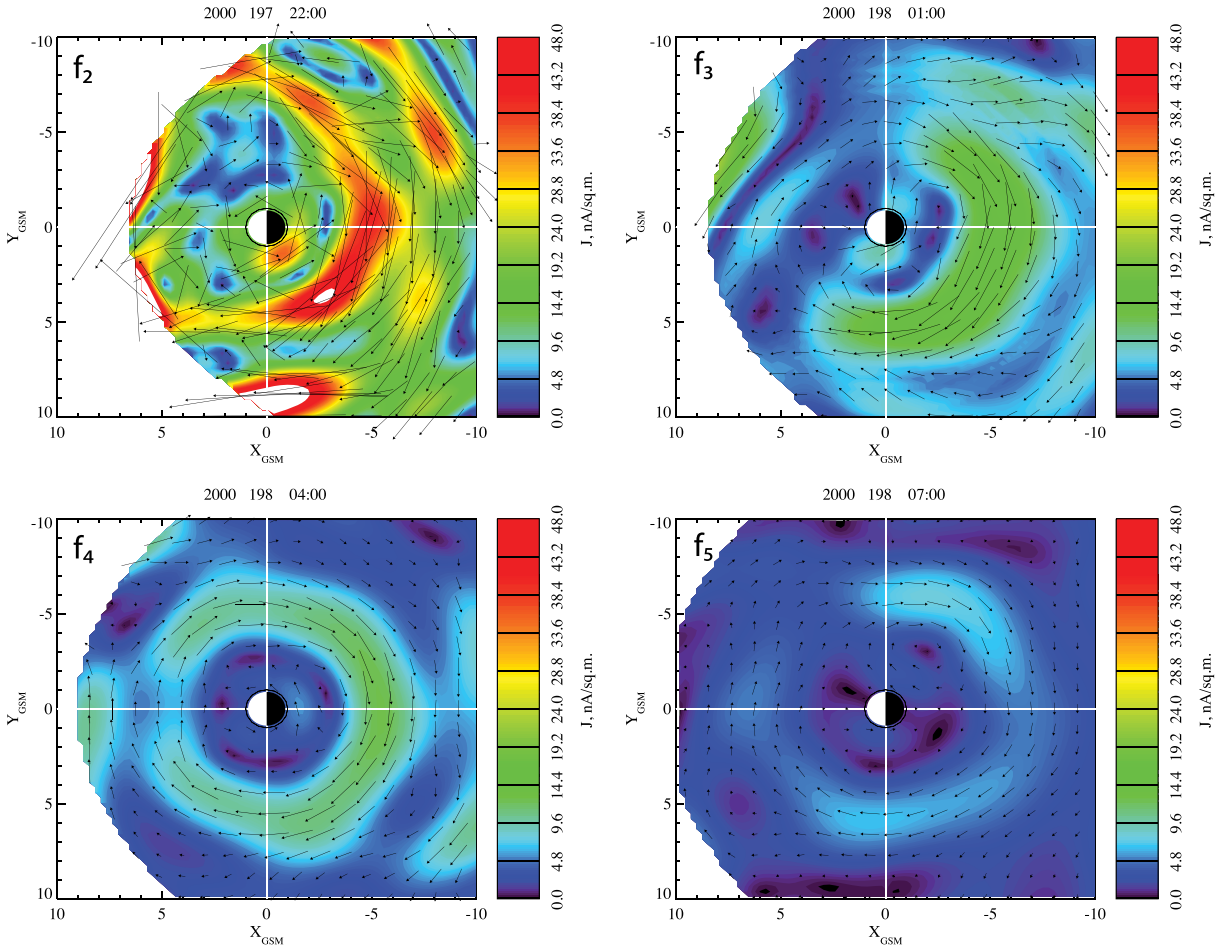

Figure 2. Equatorial current distributions (using zero tilt angle) for four moments of the Bastille Day superstorm (marked f_2–f_5 in Figure 1) obtained without NN weighting. The arrows are vectors of the current density whose absolute value is color coded.

The *Sym-H** index and the solar wind input parameters for this event are shown in Figure 1. The DM reconstruction results without weighting optimization for the equatorial current density and the derived isotropic plasma pressure at five moments f_1–f_5 marked in Figure 1 are provided in Figures 2 and 3. The reconstructions are made for similar parameters of the magnetic field model with $\rho_0 = 20R_E$, $(N, M) = (20, 6)$, and $N_{FAC} = 6$, and they only slightly differ from Figures 6 and 7 in Sitnov et al. (2018) due to the larger database used after (Sitnov et al., 2019), the number of iterations for nonlinear coefficients ($N_{it} = 100$), and different time frames.

Current distributions for this reconstruction reveal the classical picture of a storm development with the formation of the partial ring current (PRC) in the main phase reaching its peak near the *Sym-H** minimum (Figure 2f_2) and its further reduction in strength and transformation into a symmetric ring current during the recovery phase (Figures 2f_3–2f_5). All panels in Figure 2 show that the reconstruction resolves the eastward current in the innermost region (counterclockwise arrow loops). It also shows that in the main phase a part of the eastward current is connected with the westward PRC to form the banana current pattern, which was first resolved in data by Stephens et al. (2016). A distinctive feature of this superstorm is the strong compression of the magnetosphere, such that the Chapman-Ferraro currents appear already at geosynchronous orbit (Figure 2f_2). At the same time, the ring current peak exceeds those of the medium-strength storms (e.g., Sitnov et al., 2008; Stephens et al., 2016) by the factor of 5 or even more.

The plasma pressure reconstruction is made, using the quasi-static balance between the Lorentz force and the gradient of the plasma pressure, assuming its isotropy $\nabla p = \mathbf{j} \times \mathbf{B}$ (Sergeev et al., 1994;

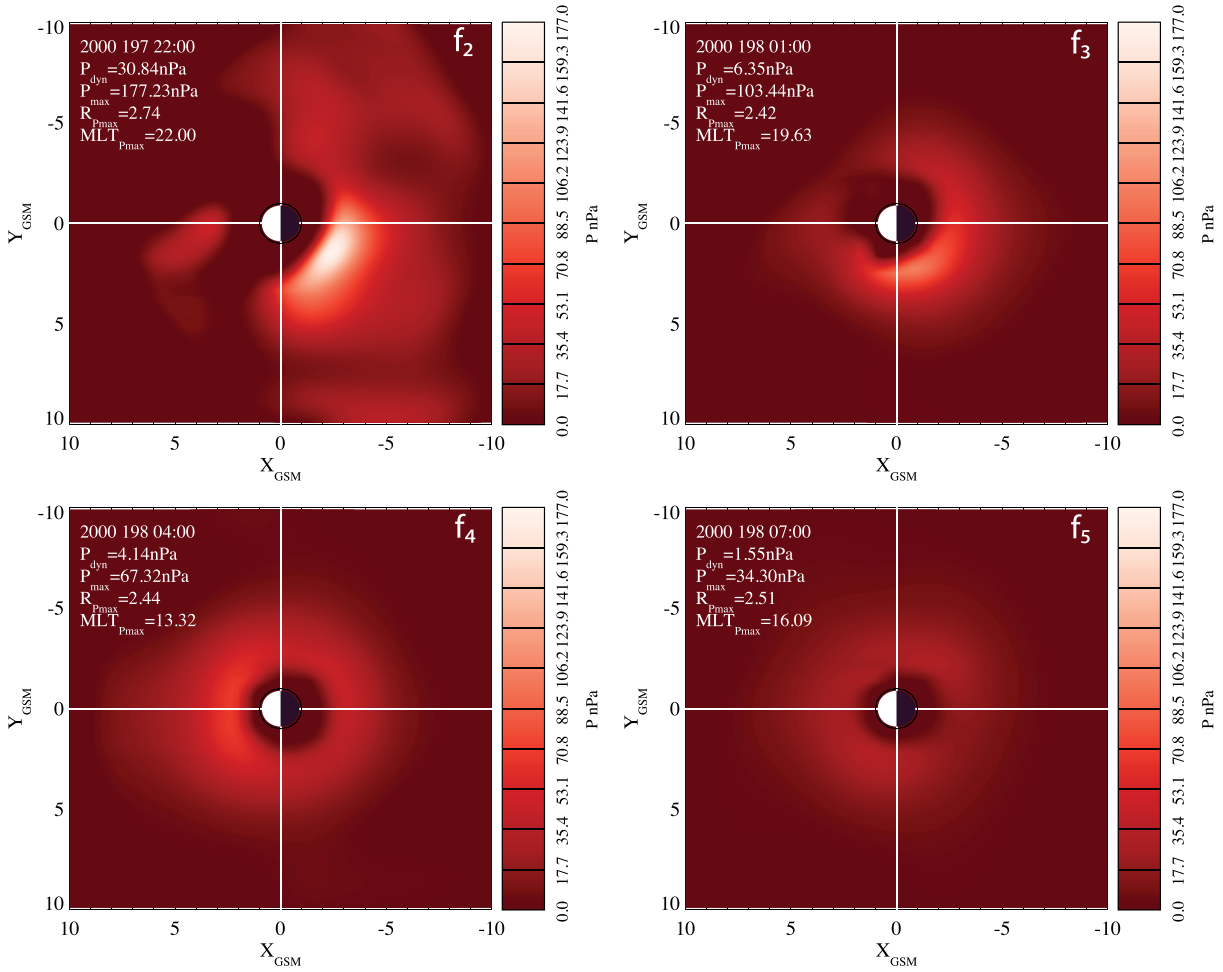

Figure 3. Color-coded equatorial plasma distributions (using zero tilt angle) for four moments of the Bastille Day superstorm (marked f_2–f_5 in Figure 1) obtained without NN weighting.

Sitnov et al., 2018; Stephens et al., 2013). Its right-hand side is integrated in the equatorial plane from the midtail boundary $x_0 = -20R_E$ to the point of interest (x, y):

$$p(x, y) = p_0 + \int_{x_0}^{x} \mathbf{j} \times \mathbf{B} \cdot \mathrm{d}\mathbf{x} \tag{9}$$

The initial value $p_0 = p(x_0, y)$ of the pressure can be estimated using the known lobe magnetic field B_0 and assuming a similar pressure balance across the plasma sheet $p_0 = B_0^2/2\mu_0$. Since previous estimates (e.g., Stephens et al., 2013) suggest that $p_0 \sim 1$ nPa, whereas typical storm pressure values and uncertainties are much larger, in the following calculations we use an ad hoc value $p_0 = 1$ nPa.

The pressure reconstruction uncertainty is caused both by the model magnetic field reconstruction errors (see, the next section for more details) and by the inherent limitations of the quasi-adiabatic force balance with the isotropic plasma approximation. In particular, the result of the path integration (9) may depend on the specific path from start to end integration points (Zaharia & Cheng, 2003). In this study, following Sitnov et al. (2018), we sew together two groups of the integration paths, earthward from $x_0 = -20R_E$ and tailward from the dayside magnetopause (assuming the same value of p_0), resulting in pressure distributions $p_n(x, y)$ and $p_d(x, y)$. They are combined near the terminator using the following matching formula: $p = \nu p_n + (1-\nu)p_d$ with $\nu = (1 - \tanh(x/s_0))/2$ and the the smoothing zone $s_0 = 1.5R_E$. The matching error is estimated as $p^{(err)} = (p_n - p_d)/[2\cosh(x/s_0)^2]$, and its distributions are provided in the supporting information.

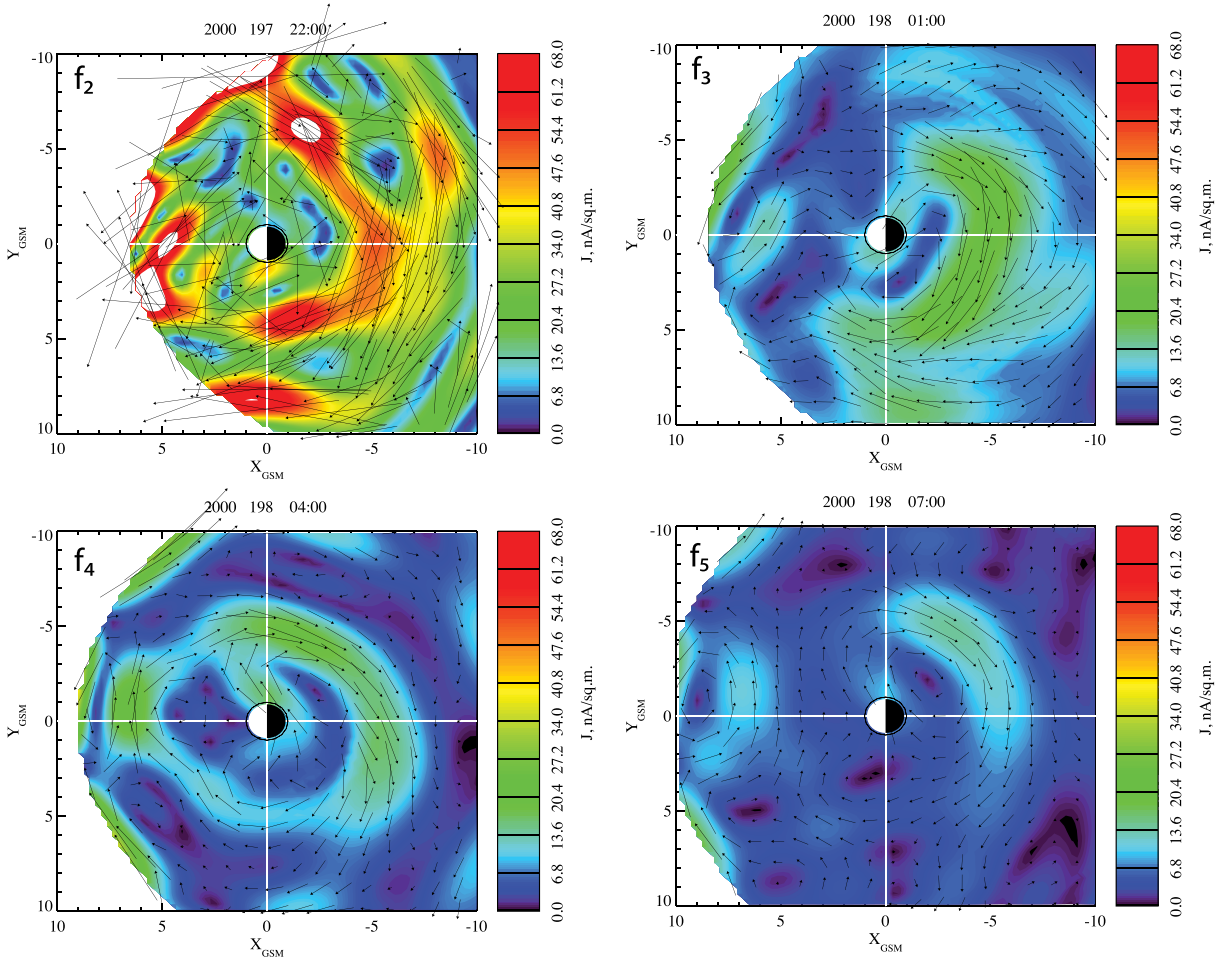

Figure 4. Equatorial current distributions for four moments of the Bastille Day superstorm (marked f_2–f_5 in Figure 1) in the format similar to Figure 2 but now obtained with the new weighted NN algorithm (7). Note the different color-coding scale, compared to Figure 2.

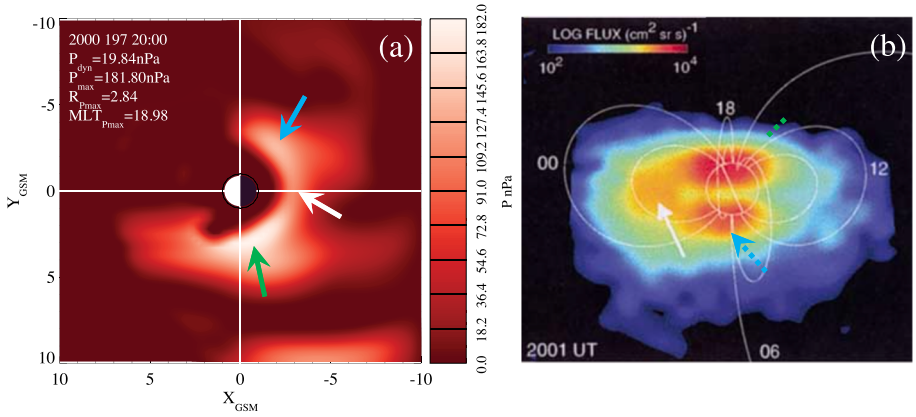

Figure 5. (a) The pressure distribution in the main phase of the Bastille Day event (frame f_1 in Figure 1), which is obtained using the distance-weighted KNN DM procedure, and (b) the ENA image in the 39–50 keV energy range at almost the same moment 20:01 UT, adapted from (Brandt et al., 2001, Figure 3b). The white dipole field lines for $L = 4$ and $L = 8$ are marked for reference at dawn, noon, dusk, and midnight.

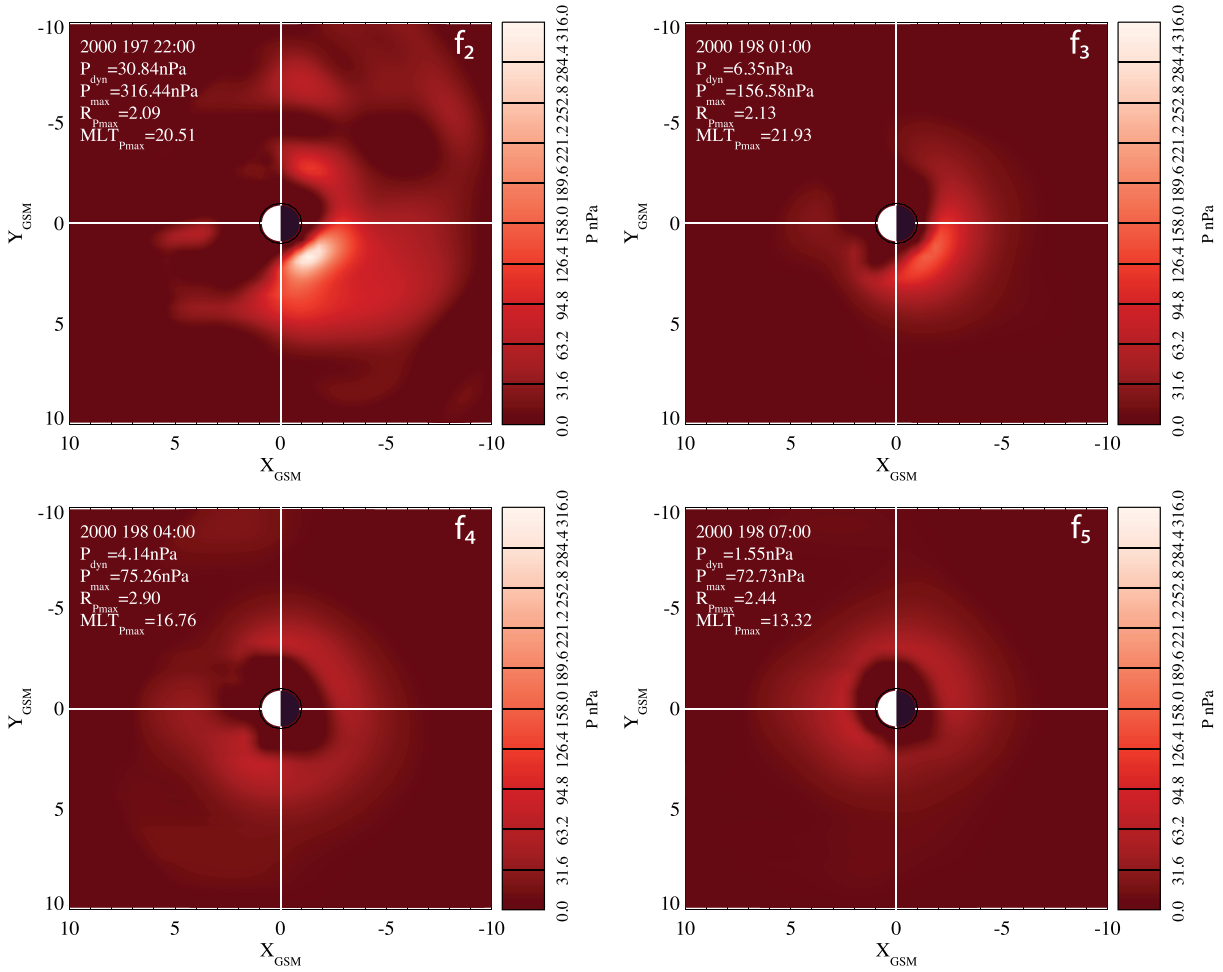

Figure 6. Color-coded equatorial plasma distributions for four moments of the Bastille Day superstorm (marked f_2–f_5 in Figure 1) similar to Figure 3 but now obtained using weighted NN algorithm.

Plasma distributions presented in Figure 3 show a conventional picture of the storm development with the pressure peak located in the premidnight sector at the end of the main phase (Figure 3f_2) and the subsequent symmetrization of the plasma distribution in the recovery phase (Figures 3f_3–3f_5). This picture is consistent with the similar evolution of the hot plasma distributions in the earlier statistical analysis (Ebihara et al., 2002), kinetic ring current models (e.g., Ebihara & Ejiri, 2003; Liemohn et al., 2001; Zaharia et al., 2006), and ENA observations (Brandt et al., 2001).

3.2. Bastille Day Event Reconstruction Using Distance-Weighted NNs

Equatorial distributions of the current density for the reconstruction of the moments f_2–f_5 of the Bastille Day event with the use of $8 \cdot 10^3$ radially weighted NNs with $\sigma = 0.3$, and the same magnetic field model parameters are presented in Figure 4. It shows that, in spite of the use of effectively $\sim K_{NN}\sigma^3 \sim 200$ NNs, the method still reveals a consistent picture of the current evolution and morphology, which is qualitatively similar to the nonweighted NN case shown in Figure 2. Moreover, it is still possible to resolve eastward and banana currents. The main difference between Figures 2 and 4 is the ≈50% increase of the current strength.

The corresponding changes in the plasma distributions are more substantial. But before discussing their main series f_2–f_5 we show in Figure 5 the plasma distribution in the main phase of this storm and compare it with the corresponding ENA image provided by the IMAGE mission (Brandt et al., 2001). ENA observation is one of the main complementary tools for global imaging of this event and they need to be properly interpreted. The fact is that the direct ENA emission from the ring current plasma may be strongly

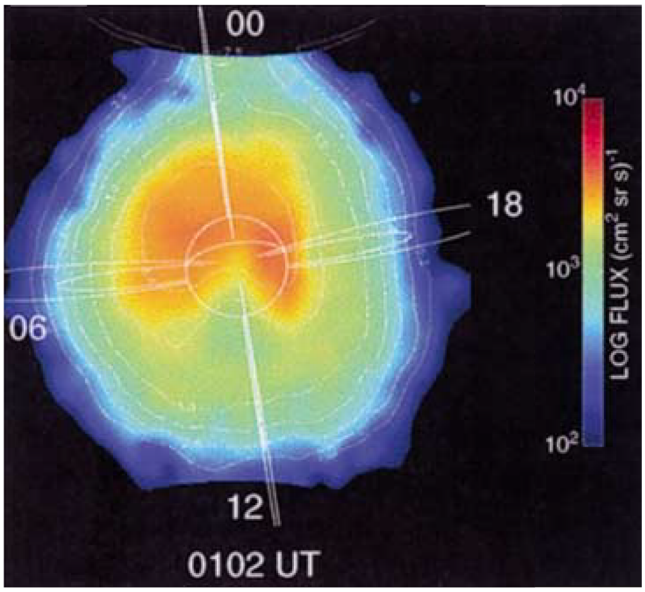

Figure 7. The ENA image in the 39–50 keV energy range at the moment ($DOY = 198, 01:02$ UT), adapted from Figure 4a in Brandt et al. (2001) and close to the moment shown in Figure 6f_3.

contaminated by the so-called low-altitude emission (LAE). The latter are produced by nearly mirroring ions at low altitudes (~500 km) where the exospheric O and N_2 start dominating (Roelof, 1997). The resulting ENA intensity can be 2–3 orders of magnitude brighter than the "genuine" ENA emissions coming from higher altitudes where the ions only charge exchange with the much more tenuous hydrogen geocorona. The LAE mainly shows up when the line of sight (LOS) in a pixel cuts field lines at ~500 km altitude at a pitch angle close to 90°.

The equatorial emission (white arrow) in Figure 5b indicates uncontaminated premidnight ring current emission. Moreover, its peak near midnight located at $L \approx 3$ is consistent with the location of the midnight current maximum of the DM reconstruction shown by the white arrow in Figure 5a. At the same time, the two brightest blobs in Figure 5b marked by dashed green and blue arrows must be LAE. Yet they are consistent with the ring current pressure enhancements at dusk and dawn marked in Figure 5a by solid green and blue arrows. In particular, the weaker morning ENA enhancement is consistent with weaker pressure in the morning sector (blue arrows).

The main set of the pressure distributions corresponding to moments f_2–f_5 in Figure 1 and obtained using the new weighted KNN algorithm, which is described in the previous section, is presented in Figure 6. Figure 6f_2 corresponds to the moment with the minimum value of $Sym\text{-}H < -300$ nT. At the same time, the comparison of black and red lines in Figure 1a shows that with the new weighted KNN algorithm the DM bin has practically no bias toward weaker events, as it was the case with the original KNN algorithm (blue line). It is even better than the nonweighted KNN with $K_{NN} = 2 \cdot 10^3$ points (gray lines in Figure 1). Nevertheless, the new weighted NN set provided the spatial current resolution, which is sufficient to resolve the eastward current and hence to determine the pressure peak. The latter turns out to be exceedingly strong ($p_{\max} \approx 316$ nPa) and to be located very close to the Earth ($R_{p_{\max}} \approx 2.09 R_E$). Note that the pressure matching errors provided in the supporting information are relatively small and localized along the terminator with their peak value $p_{max}^{(err)} \approx 192$ nPa.

The next panel (Figure 6f_3) is interesting because it can be compared with the corresponding IMAGE ENA distribution shown here in Figure 7 and adapted from Figure 4a in Brandt et al. (2001). The advantage of the latter is an almost complete absence of LAE contamination. Consistent with this ENA image, the pressure distribution in Figure 6f_3 has a similar crescent shape due to a significant decrease in both ENAs and pressure values in the morning sector. It is located in the premidnight sector with the peak $p_{\max} \approx 157$ nPa located at similar distance from the Earth ($R_{p_{\max}} \approx 2.13 R_E$) as 2 hr before, and with $MLT_{p_{\max}} \approx 22$ hr. While the ENA image is rather blurred because of the 10° transverse point spread width resulting in $>1 R_E$ uncertainty in its spatial resolution, the corresponding DM-based image in Figure 6f_3 is quite sharp,

Table 1
Pressure Parameter Summary for Nonweighted and Weighted KNN Algorithms in Case of the Bastille Day Event

Frame	DOY, UT	P_{max}	$P_{max}^{(w)}$ [a]	R_{Pmax}	$R_{Pmax}^{(w)}$	MLT_{Pmax}	$MLT_{Pmax}^{(w)}$
f_1	197, 20:00	119	182	2.9	2.8	21	19
f_2	197, 22:00	177	316	2.7	2.1	22	21
f_3	198, 01:00	103	157	2.4	2.1	20	22
f_4	198, 04:00	67	75	2.4	2.9	13	17
f_5	198, 07:00	34	73	2.5	2.4	16	13

[a] Index $^{(w)}$ denotes runs with weighted NNs with $\sigma = 0.3$. Pressure and MLT data are rounded off to an integer; R_{Pmax} data are rounded off to the first decimal.

especially at its earthward side. It is also interesting that the relative matching error for this moment is much smaller ($p_{max}^{(err)} \approx 35$ nPa) compared to the moment $DOY = 197$, $UT = 20:00$ (Figure 5a and the supporting information). This may be an indication of validity of the quasi-static force balance approximation with isotropic plasma (more discussion of this problem in section 4.

Other panels in Figure 6 describe symmetrization of the ring current in the growth phase, consistent with the classical storm picture derived from the statistical analysis (Ebihara et al., 2002), kinetic ring current models (e.g., Ebihara & Ejiri, 2003; Liemohn et al., 2001; Zaharia et al., 2006), and ENA observations (Brandt et al., 2001) (Figure 5). The corresponding matching error distributions, whose peaks are smaller than the pressure peaks by an order of magnitude, reveal a strong consistency of the different integration paths and suggest plasma isotropization and its steady evolution later in the recovery phase of this storm.

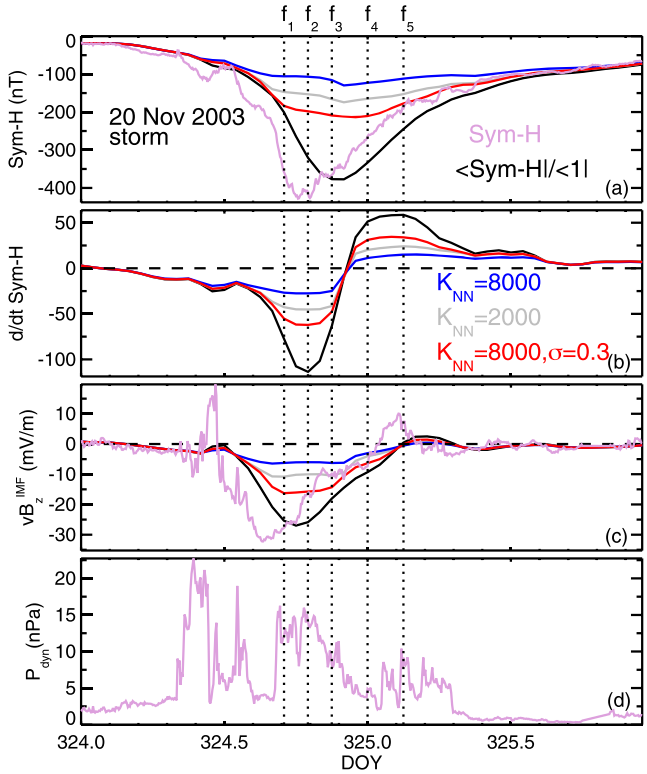

Figure 8. Analysis of the NN bins and input parameters for the 20 November 2003 superstorm with panels (a)–(d) made in the format similar to Figure 1. Vertical dashed lines mark moments in this substorm $f_1 = 324$, 17:00 UT, $f_2 = 324$, 19:00 UT, $f_3 = 324$, 21:00 UT, $f_4 = 325$, 00:00 UT, and $f_5 = 325$, 03:00 UT, which are discussed in more detail later.

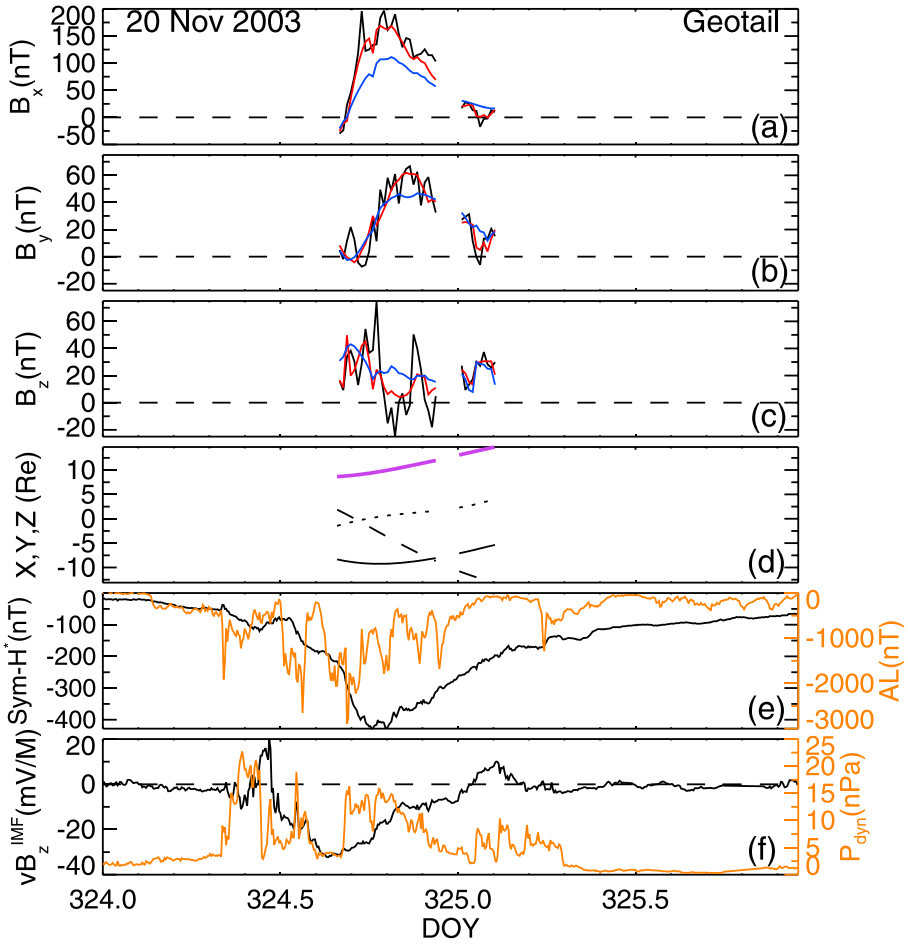

Figure 9. Magnetic field validation of the 20 November 2003 storm using Geotail observations: (a)–(c) GSM magnetic field components from Geotail (black), nonweighted (blue) and weighted NNs (red); (d) Geotail ephemeris (X, Y, Z, and R components correspond to the black solid, dashed, dotted, and purple lines); (e) *Sym-H** (black line) and *AL* (orange line) indices of storm and substorm activity; (f) the solar wind parameter vB_z^{IMF} (black line) and solar wind dynamic pressure (orange line).

The results of the pressure investigation for the Bastille Day event are summarized in Table 1. It shows that the use of the weighted KNN algorithm substantially changes both the values of the pressure peak, which exceeds 300 nPa, and its radial distance from the Earth, which becomes as small as ≈2.1 R_E. Thus, the pressure almost doubles, whereas the peak radial location reduces by more than half R_E. Also, in the advanced DM picture, the dip in the radial location the pressure peak is found earlier and closer to the minimum of the original *Sym-H** value (Figure 1). The changes in the azimuthal location of the pressure peak are less substantial. The improved DM picture shows a persistent clockwise (in time) shift of the peak in frames f_1–f_3 which is not seen in the original DM reconstructions. However, further MLT discrepancy in the recovery phase (frames f_4 and f_5) rather reflects symmetrization of the pressure distribution when the MLT value of the (relatively small) pressure peaks becomes unimportant.

3.3. The 20 November 2003 Storm Reconstruction Using Standard KNN Algorithm

Another EE in our study, the 20 November 2003 superstorm is by now the strongest storm since 1995, when the solar wind and IMF input started being continuously monitored around the L1 point. It is also the second largest storm in the space era (since 1957) with the minimum $Dst = -472$ nT. Compared to the Bastille Day event, this storm is also interesting because it allows validation of the geomagnetic field reconstructions using in situ observations from Geotail and Polar missions.

The storm was caused by the CME associated with an M-class solar flare that occurred on 18 November 2003 (e.g., Ebihara et al., 2005). Due to a high-speed solar wind $v \sim 700$ km/s) and a strong southward IMF up to

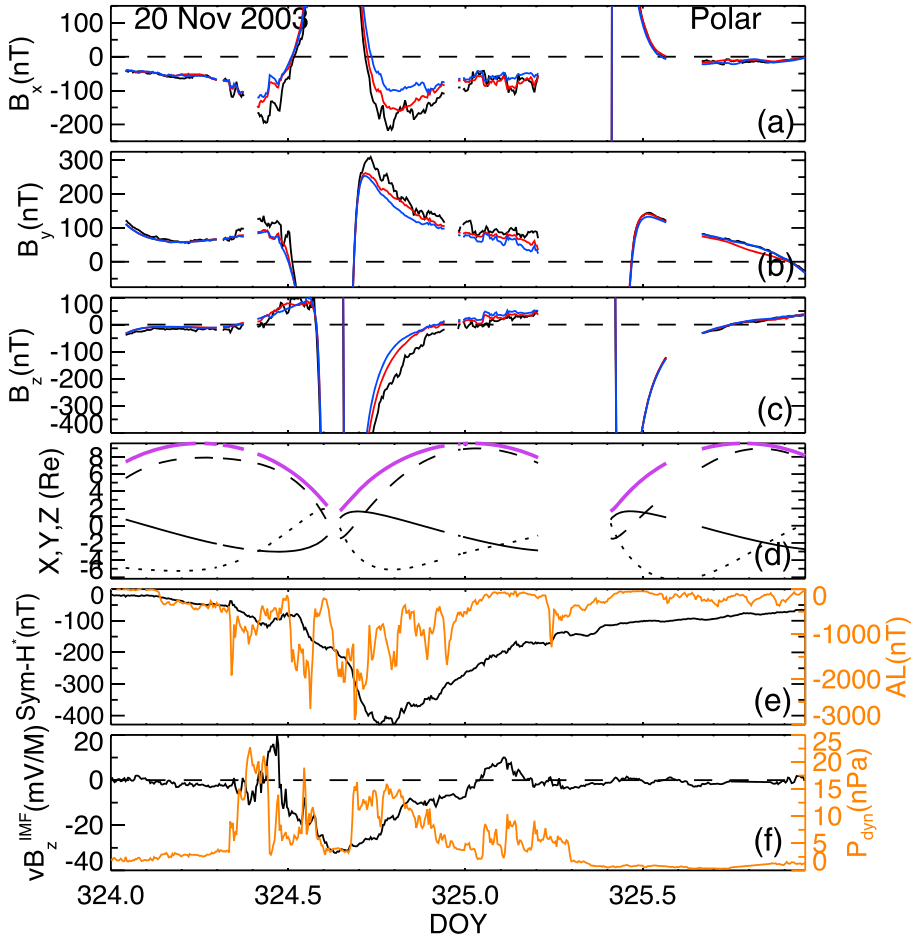

Figure 10. Magnetic field validation of the 20 November 2003 storm using Polar observations. Panels (a)–(f) are made in the format similar to Figure 9.

~60 nT, its driving electric field reached ≈30 mV/m. This peak is 2 times smaller, compared to the Bastille Day event. However, in this case the strong driving was longer, which is seen from the larger values of the convoluted input parameter $\langle vB_z^{IMF}|$ (black line in Figure 8c) and stronger dips of both the original index *Sym-H** and its convolution $\langle$*Sym-H**| (plum and black lines in Figure 8a).

The analysis of the DM bins presented in Figure 8 shows that, in the case of the Bastille Day event, the use of weighted NNs substantially reduces the statistical bias of the DM method. The amplitudes of key DM parameters are now smaller than their original values for this EE by only a factor of 2, with almost no bias for the moment f_1 in the main phase. Figures 9 and 10 show that this bias reduction is accompanied by the substantial improvement of the validation results for Geotail and Polar.

In Figure 11 we present the equatorial current distributions for frames f_1 and f_3–f_5 in this storm, which are obtained with nonweighted KNN algorithm. They show the structure and evolution of currents, which is qualitatively similar to the Bastille Day storm (Figure 2), with the formation of the PRC in the near-midnight and in the premidnight sectors in the main phase and its symmetrization and weakening in the recovery phase. Quantitative distinctions, compared to the Bastille Day, are the more gradual reduction of the current density and its smaller values near the *Sym-H** minimum. (To be more precise, the minimum $j \approx 27$ mV/m is achieved near frame f_2, which we reserved for a further separate discussion and comparison with ENA distributions.)

The pressure distribution derived using the integration (9) of the quasi-static force balance with the isotropic plasma approximation and nonweighted KNN algorithm for the magnetic field reconstruction are presented

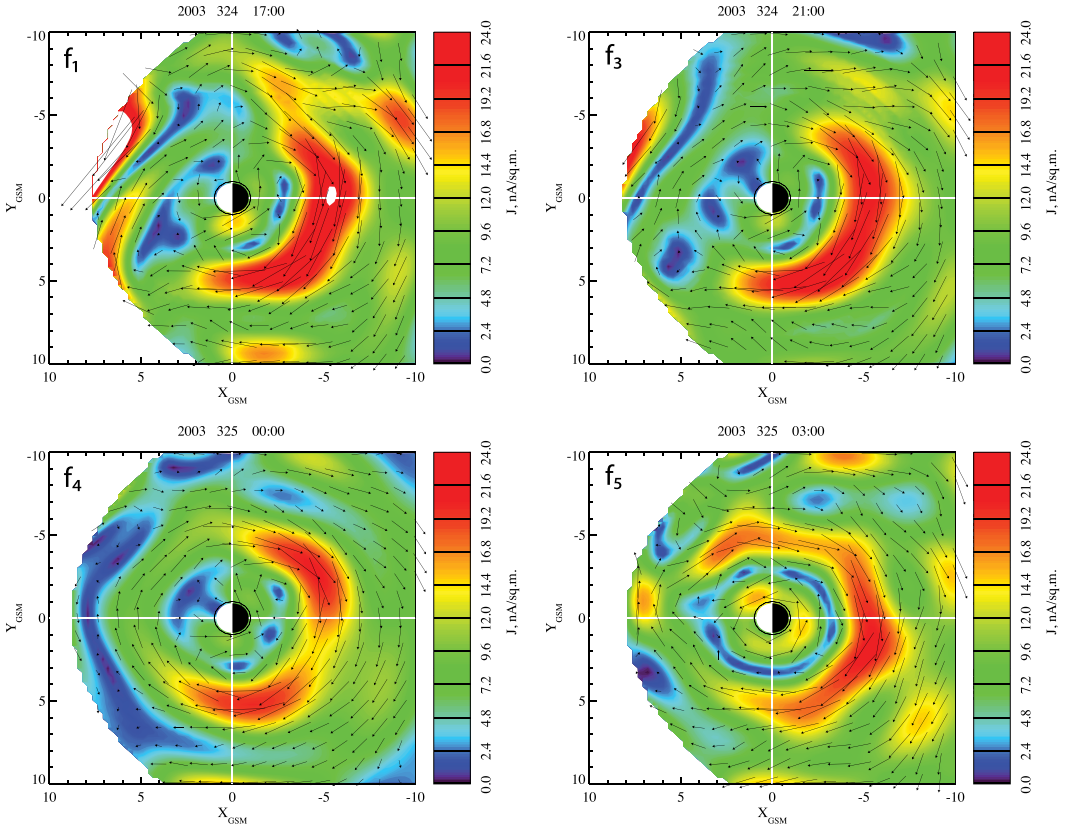

Figure 11. Equatorial current distributions for four moments f_1 and f_3–f_5 (in Figure 8) of the 20 November 2003 superstorm obtained without NN weighting in the format similar to Figure 2.

in Figure 12. It shows the formation of the pressure peak in the premidnight sector around the *Sym-H** minimum and further symmetrization of the pressure distribution in the recovery phase. However, in contrast to the Bastille Day storm, the strongest pressure peak is shifted to the early growth phase (frame f_4).

3.4. The 20 November 2003 Storm Reconstruction Using Distance-Weighted NNs

Analogs of current and pressure distributions shown in Figures 11 and 12 for weighted NNs with $\sigma = 0.3$ and the same magnetic field model architecture ($(N, M) = (20, 6)$ and $N_{FAC} = 6$) are provided in the supporting information file. They show a substantial increase of the current and pressure amplitudes, compared to nonweighted NN reconstructions discussed in the previous section. In particular, the largest pressure peak increases from 125 to 256 nPa for frames f_1 and f_3–f_5 and it is achieved now earlier in this storm. In fact, the strongest peak $P^{(w)}_{max}(f_2) = 390.76$ (not shown) is achieved for frame f_2. This is almost 4 times stronger than its value $P_{max}(f_2) = 96.84$ for the same frame in nonweighted reconstructions. The pressure distributions become more asymmetric in the azimuthal angle, being sharper in the radial directions and their peaks being located closer to the Earth, compared to the nonweighted NN case.

Here we show a further refinement of the reconstruction, which becomes possible for this event, namely the results of weighted KNN with $\sigma = 0.3$, the magnetic field model with $(N, M) = (20, 6)$, and a better resolution of the FACs due to $N_{FAC} = 12$ (note, that setting $N_{FAC} = 12$ for the Bastille Day storm results in the model overfitting, which is likely due to stronger compression of the magnetosphere in the main phase of that storm). The current distributions (Figure 13) are now more consistent without an anomalously strong current peak for frame f_5. Their morphology is similar to nonweighted reconstructions (Figure 11), including classical crescent-like PRC in frame f_3, double crescent configuration in frame f_4 (note that similar structures were reported in the original TS07D reconstructions for moderate storms; see, for instance, Figure 4c in Sitnov et al., 2008), and more ring-like azimuthally symmetric distributions in frame f_5.

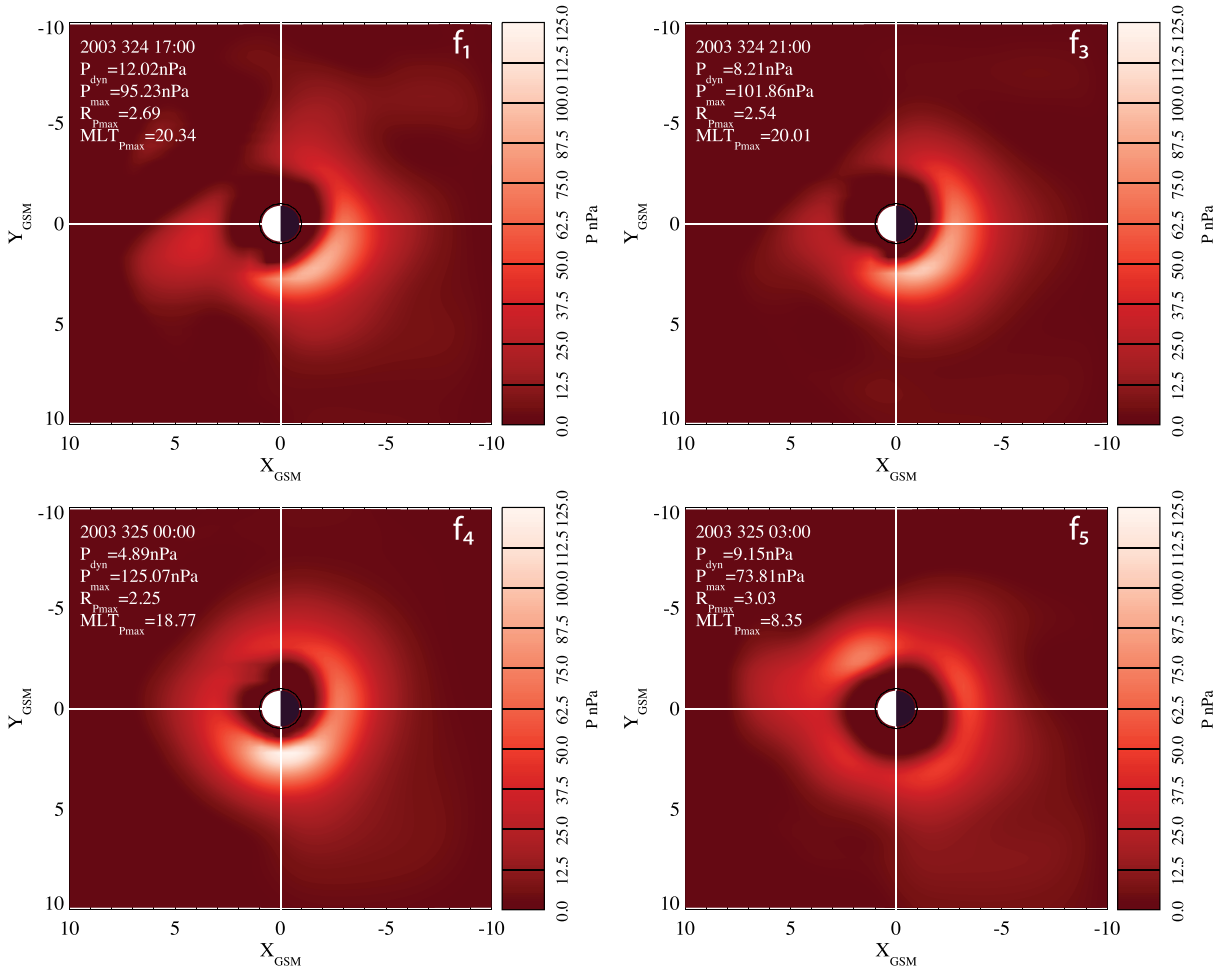

Figure 12. Equatorial plasma distributions for four moments f_1 and f_3–f_5 (in Figure 8) of the 20 November 2003 superstorm obtained without NN weighting in the format similar to Figure 3.

The corresponding pressure distributions are presented in Figure 14 and their matching errors are provided in the supporting information. As is seen from the comparison of Figure 14 with Figure 12, the pressure peak in the weighted NN reconstruction approximately triples. This is consistent with the estimate made in the analysis of nonweighted NN bins in Sitnov et al. (2018). In addition, the improved pressure distributions become sharper in the radial direction with their narrower peaks being located closer to Earth. They are also more asymmetric in MLT near the *Sym-H** minimum and in the main phase (frames f_1, f_3, and f_4).

In Figure 15a we show separately the pressure distribution for the frame f_2 when all three binning parameters G_1–G_3 defined in (3)–(5) and shown in Figure 8 by black lines reach their extrema. This frame is also close to the original *Sym-H** minimum (plume line in Figure 8a). This pressure distribution is particularly interesting, because it may be compared with the ENA image obtained from a polar vantage point with the minimal LAE contamination (Figure 15b). In spite of the peak values of several input parameters, this frame does not show the strongest pressure peak or its closest approach to Earth. Yet the peak is located so close to Earth that the corresponding ENA image is occulted by the planet. While the latter is relatively free from LAE, the image resolution (10° or $\sim 1R_E$) is apparently insufficient to determine the radial location of the pressure peak. Nevertheless, its general morphology with the brightest spot occupying the pre-midnight sector is consistent with our DM reconstruction.

Distributions of the near-terminator errors arising from mismatch between the pressure integrals p_n and p_d obtained from (9) for integration paths from the tail ($X = -20R_E$) and the dayside magnetopause for frames

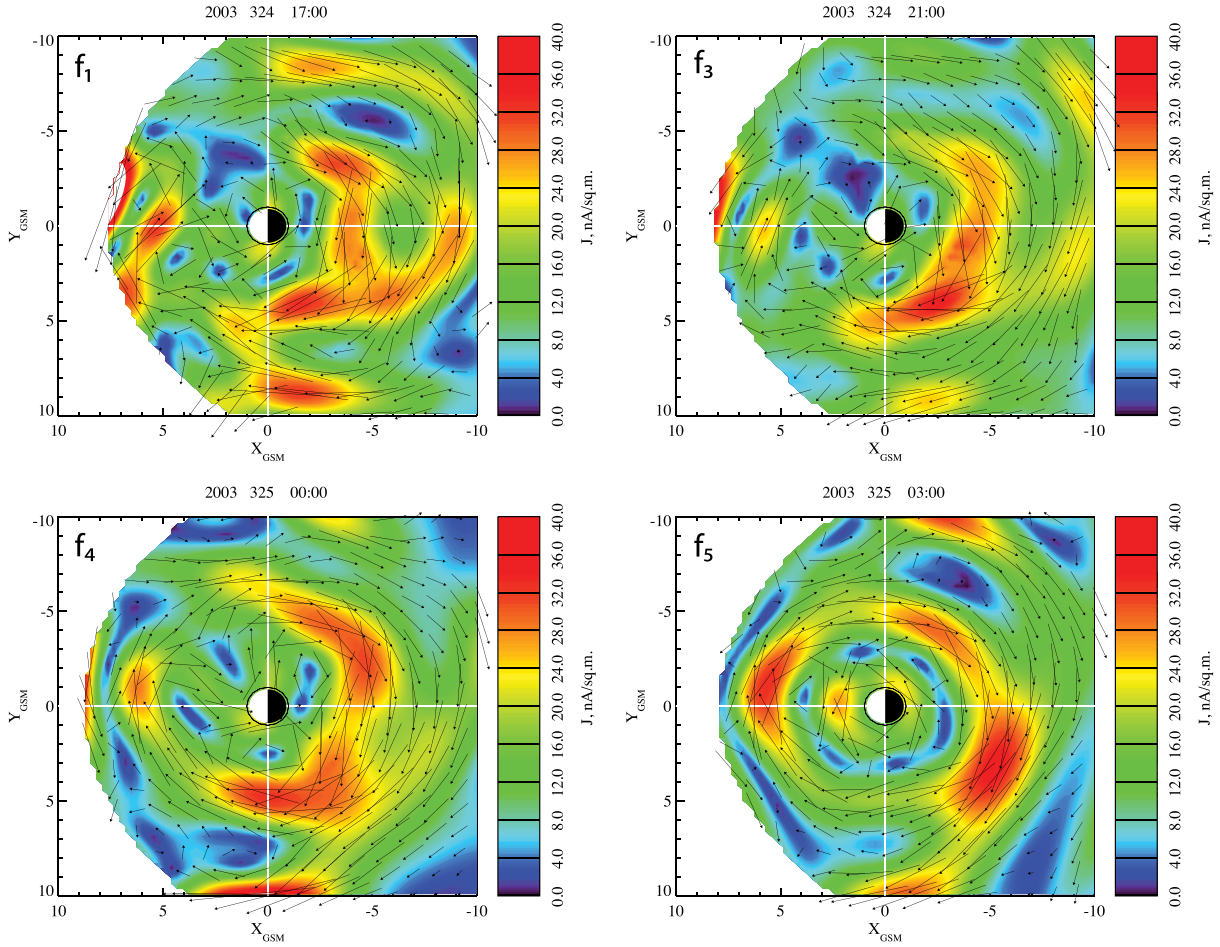

Figure 13. Equatorial current distributions for four moments f_1 and f_3–f_5 (in Figure 8) of the 20 November 2003 superstorm obtained with weighted NN using $\sigma = 0.3$ and an advanced magnetic field model with $(N, M) = (20, 6)$ and $N_{FAC} = 12$. The format is similar to Figure 2. Note the different color-coding scale, compared to Figure 11.

f_1 and f_3–f_5 (in Figure 8) are presented in the supporting information in the format similar to the original pressure distributions (Figure 14). They show that the errors are relatively small and their ratio to the local pressure peak reduces away from the *Sym-H** minimum.

We summarize the results of the pressure reconstruction for the 20 November 2003 storm in Table 2. It shows that in the new reconstruction, advanced due to the weighted KNN algorithm, the peak pressure triples and its radial distance becomes closer to Earth by $\approx 0.7 R_E$. These changes are more drastic, compared to the Bastille Day event (Table 1). Also, in contrast to the Bastille Day storm, the strongest pressure peak is achieved in the same phase as in the nonweighted reconstruction, ≈5 hr after the *Sym-H** minimum (frames f_4 vs. f_2). In addition, in MLT the peak shifts counterclockwise (in time) and it is overall shifted more to the midnight. At the same time, in both storms the strongest pressure peaks are found at the same local time MLT ≈ 21 hr.

The new magnetic field model architecture with 12 FAC elements allows one to reconstruct the FAC distributions for this storm. Consistent with DMSP observations, discussed by Ebihara et al. (2005), these distributions shown in Figure 16 reveal a structure, which is more complex, compared to classical sets of R1 and R2 systems (e.g., Iijima & Potemra, 1976), and it contains multiple current sheets. Further, consistent with DMSP data (Ebihara et al., 2005, Figure 9), FACs extend well beyond 50° latitude (Figure 16f_3) and they disappear shortly after the moment f_4 (Figure 16f_5) because the storm current becomes azimuthally symmetric (Figure 14f_5).

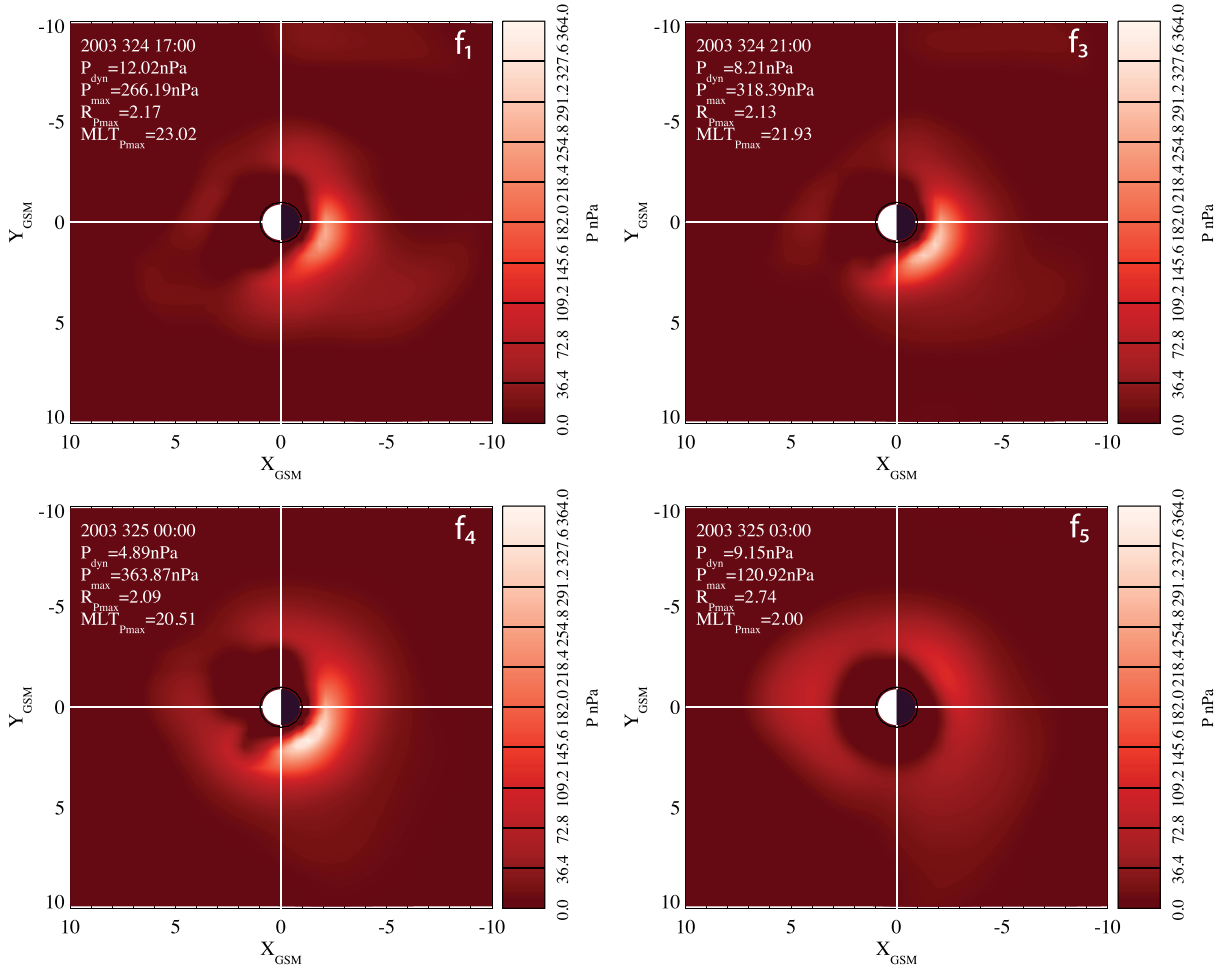

Figure 14. Equatorial plasma distributions for four moments f_1 and f_3–f_5 (in Figure 8) of the 20 November 2003 superstorm obtained with weighted NN using $\sigma = 0.3$ and an advanced magnetic field model with $(N, M) = (20, 6)$ and $N_{FAC} = 12$. The format is similar to Figure 12.

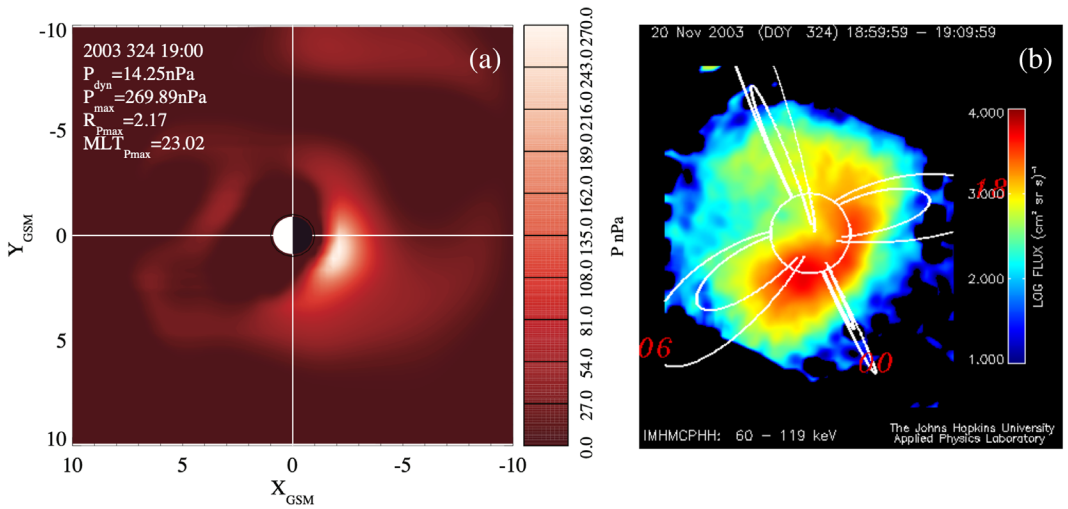

Figure 15. (a) The equatorial pressure near the *Sym-H** minimum (frame f_2 in Figure 8) of the 20 November 2003 storm obtained with weighted NN using $\sigma = 0.3$ and an advanced magnetic field model with $(N, M) = (20, 6)$ and $N_{FAC} = 12$. The maximum pressure in the color table here is different from the panels in Figure 14 to better reveal the pressure morphology. (b) The ENA image in the 60–119 keV energy range at almost the same moment (19:00–19:10 UT) with the minimal LAE contamination due to the appropriate vantage point.

Table 2
Pressure Parameter Summary for Nonweighted and Weighted KNN Algorithms in Case of the 20 November 2003 Storm

Frame	DOY, UT	P_{max}	$P^{(w)}_{max}$	R_{Pmax}	$R^{(w)}_{Pmax}$	MLT_{Pmax}	$MLT^{(w)}_{Pmax}$
f_1	324, 17:00	95	266	2.7	2.2	20	23
f_2	324, 19:00	97	270	2.9	2.2	21	23
f_3	324, 21:00	102	318	2.5	2.1	20	22
f_4	325, 00:00	125	364	2.3	2.1	19	21
f_5	325, 03:00	74	121	3.0	2.7	9	2

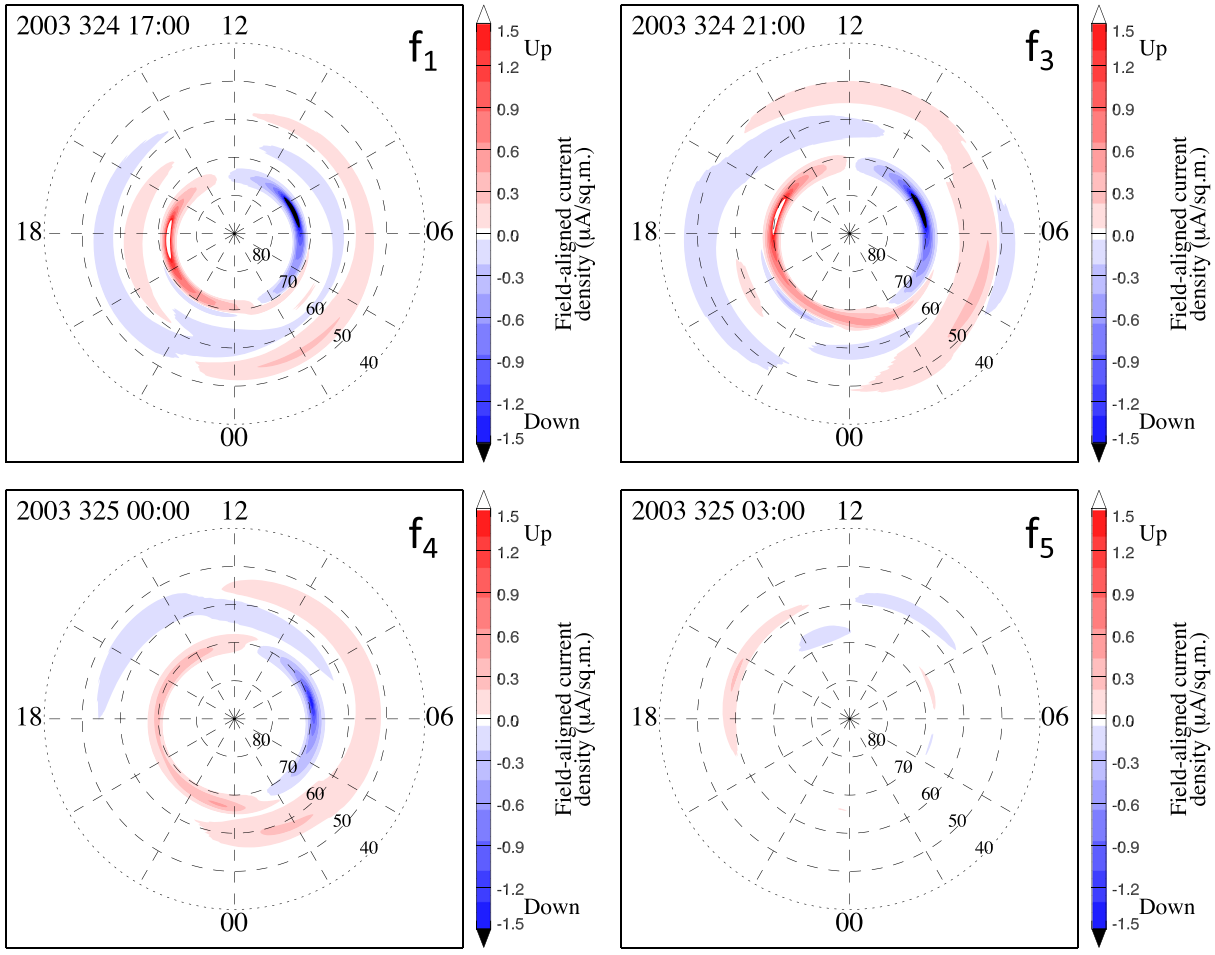

Figure 16. Low-altitude FAC distributions for four moments f_1 and f_3–f_5 (in Figure 8) of the 20 November 2003 superstorm obtained with weighted NN using $\sigma = 0.3$ and an advanced magnetic field model with $(N, M) = (20, 6)$ and $N_{FAC} = 12$. Upward (positive) and downward (negative) currents are shown, respectively, by red and blue shading.

4. Discussion and Conclusion

Extremely strong storms represent a double challenge for their empirical reconstruction. On the one hand, because of their rarity, there is a little information from in situ observations of similar events. On the other hand, the reconstruction of their spatial structure requires extreme refinement to resolve the ring current (particularly its innermost eastward component), which is most closely located to Earth. The common underlying problem causing these challenges is the data paucity of in situ observations when only a few probes are available at any moment. DM based on the KNN algorithm mitigates this problem because it allows one to form swarms of virtual probes whose number is much larger than the number of observations at the moment of interest, and at the same time, much smaller than the whole database number. Yet for EEs the statistical analysis of KNN bins reveals strong bias toward weaker events. Moreover, it cannot be reduced by reduction of the bin size K_{NN}, because at some point their number becomes insufficient to resolve the eastward current and hence to reconstruct the ring current pressure peak.

In this study we showed that the use of DM with distance-weighted NNs provides a substantial advance in the resolution of this data paucity curse. Figures 1 and 8 demonstrate the reduction of bias in the NN bins, which is equivalent to the reduction of their effective number by a factor of σ^3, where σ is the weighting exponent. In particular, with $\sigma = 0.3$ and $K_{NN} = 8 \cdot 10^3$, which reduces the effective K_{NN} value to $K_{NN}^{(eff)} \sim \sigma^3 K_{NN} \approx 200$, it becomes possible to avoid overfitting using a model with the number of free parameters larger than $K_{NN}^{(eff)}$. As a result, it becomes possible to resolve the eastward current, ring current pressure peak, and even the complex FAC structure, which is a distinctive feature of these EEs.

It is interesting that pressure peaks in both EEs considered are observed in the same phases of the corresponding storms for both weighted and nonweighted cases. For the Bastille Day event this is the *Sym-H** minimum, whereas for the 20 November 2003 storm the largest pressure in found in the early recovery phase. This morphological consistency of the original and advanced DM methods suggests that one can further renormalize data using known *Sym-H** values as was suggested by Sitnov et al. (2018). They suggested to use the remaining mismatch between *Sym-H** and its NN average (black and red lines in Figures 1a and 8a) to renormalize the reconstructed pressure. An additional problem is that this can be done by both rescaling the pressure values and by changing the radial location of the pressure peak. The present study shows that the pressure peak for strong storms may be as close as $R_{Pmax} \approx 2.1 R_E$ to Earth and its value exceeds 300 nPa. At the same time, for the 20 November 2003 storm, observations of the flux of 30–80 keV ions measured by the polar-orbiting NOAA 17 satellite near the 2200 MLT meridian showed that the inner edge of the ion plasma sheet moved inward during the main phase reaching the region $L \approx 1.5$ (Ebihara et al., 2005). This suggests that the variation range for the parameter R_{Pmax} does not exceed ~25% and hence the main renormalization is expected for the pressure value, which may be therefore as high as ~600 nPa at its maximum. At the same time, for the Bastille Day storm, the strongest pressure peak (Figure 6f_2) with $P_{max} = 316$ nPa does not even require any further rescaling, because the corresponding bias (Figure 1) is absent.

In addition, weighted NN reconstructions reveal sharper radial pressure profiles, which are also more asymmetric in MLT. The corresponding FACs extend well beyond 50° in latitude, consistent with DMSP observations (Ebihara et al., 2005). In situ observations of Polar and Geotail during the 20 November 2003 storm show that the new DM method not only reduces the NN bias but also substantially improves validation results. Comparison of the new pressure reconstructions with ENA observations shows their qualitative consistency for selected moments with near-polar vantage points excluding LAE contamination. At the same time, further quantitative comparison requires further increase of the ENA angular resolution and comparison of the ENA images with their modeling using the obtained pressure profiles (e.g., DeMajistre et al., 2004; Ebihara & Fok, 2004), which we plan to perform in the future studies.

In this study we use a relatively simple algorithm of the pressure evaluation (9) proposed by Sergeev et al. (1994) and successfully used in our earlier investigations of SMCs and storms (Sitnov et al., 2018; Stephens et al., 2013). It is complemented here by the description of the matching error distributions provided in the supporting information. Other approaches to the pressure reconstruction from the force balance equation, including the Poisson's equation approach proposed by Zaharia and Cheng (2003), are discussed in the supporting information and in Stephens et al. (2020).

The validity of the original approach is seen from the comparison of its results with the pressure values obtained from particle instruments of the Van Allen Probes mission (Sitnov et al., 2018; Stephens et al., 2020).

This comparison shows not only the overall consistency of two pressure evaluation methods but also relatively mild impact of the pressure anisotropy on its spatial distribution. This explains important results obtained in the storm physics using first-principles kinetic models with isotropic plasma distributions (Wolf et al., 2007, 2016). Moreover, the empirical magnetic field and pressure distributions can be directly assimilated with first-principles models as was demonstrated for global MHD simulations with the empirical TS07D pressure by Merkin et al. (2019). They can be further combined with both modern isotropic (Yang et al., 2019) and anisotropic (Fok et al., 2014) ring current models to refine the corresponding energy distributions and anisotropy effects by providing those kinetic storm models with global magnetic field distributions and plasma pressure distributions at their boundaries.

The presented reconstruction of the plasma pressure profiles may look surprising and counterintuitive. Indeed, even though NNs are defined in the global binning space (3)–(5), for sufficiently small values of K_{NN} or weighting strength factor σ, they should correspond to a very few observations that are available at the moment of interest. Their distributions in the real space may be singular and the singularity may result in artifacts in the obtained current and pressure distributions, which should be exacerbated by the extreme nature of the considered strong storms. However, the analysis of the NN data distributions in the real space provided in Stephens et al. (2020) and in the supporting information of this study suggests that the used numbers of K_{NN} provide sufficiently dense coverage of the magnetosphere in the region under scrutiny. Furthermore, it appears that the moderately small value of the weighting factor σ used in our study does not reduce the resolution in space too much. At the same time, the inhomogeneity of the NN data distributions is mitigated in the magnetic field fitting process by using the basis function expansions (1) and the SVD method in finding the linear amplitude coefficients in (1), which strongly reduces the noise in data. In addition, the original data are weighted in the RMS function (7) to reduce the radial inhomogeneity of observations (weighting coefficients $w_{(0)}$, which are described in more detail in Tsyganenko & Sitnov, 2007). This allows, in particular, to mitigate the singularity in space provided by geosynchronous missions. The combination of all these methods, among which we would like to single out the distance-weighted KNN method, gives an incredibly detailed and consistent empirical description of the strongest observed storms.

Data Availability Statement

The data and data products related to the paper are archived on Zenodo (https://doi.org/10.5281/zenodo.3953089).

Acknowledgments
This work was funded by NASA grants NNX16AB78G, NNX16AB80G, and 80NSSC19K0074, as well as NSF grants AGS-1702147 and AGS-1744269. N. A. T. acknowledges support of the RFBR grant 20-05-00218. We thank the many spacecraft and instrument teams and their PIs who produced the data sets we used in this study, including the Cluster, Geotail, Polar, IMP-8, GOES, THEMIS, Van Allen Probes, and MMS, particularly their magnetometer teams. We also thank the SPDF for the OMNI database for solar wind values, which is composed of data sets from the IMP-8, ACE, WIND, and Geotail missions, and also the WDC in Kyoto for the Geomagnetic indices. This study strongly benefited from interesting discussions within the framework of the NASA LWS Team "Understanding Physical Processes in the Magnetosphere–Ionosphere / Thermosphere/Mesosphere System During Extreme Events."

References

Brandt, P. C. S., Mitchell, D., Roelof, E., & Burch, J. (2001). Bastille Day storm: Global response of the terrestrial ring current. *Solar Physics*, *204*(1), 377–386.

Buzulukova, N., Fok, M. C., Glocer, A., Komar, C., Kang, S. B., Martin, S., et al. (2018). Chapter 10—Geomagnetic storms: First-principles models for extreme geospace environment. In N. Buzulukova (Ed.), *Extreme events in geospace* (pp. 231–258). Cambridge, MA: Elsevier. https://doi.org/10.1016/B978-0-12-812700-1.00010-8

Chen, M. W., Schulz, M., Lu, G., & Lyons, L. R. (2003). Quasi-steady drift paths in a model magnetosphere with AMIE electric field: Implications for ring current formation. *Journal of Geophysical Research*, *108*(A5), 1180. https://doi.org/10.1029/2002JA009584

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

DeMajistre, R., Roelof, E. C., C:son Brandt, P., & Mitchell, D. G. (2004). Retrieval of global magnetospheric ion distributions from high-energy neutral atom measurements made by the IMAGE/HENA instrument. *Journal of Geophysical Research*, *109*, A04214. https://doi.org/10.1029/2003JA010322

Ebihara, Y., & Ejiri, M. (2003). Numerical simulation of the ring current: Review. *Space Science Reviews*, *105*(1), 377–452. https://doi.org/10.1023/A:1023905607888

Ebihara, Y., Ejiri, M., Nilsson, H., Sandahl, I., Milillo, A., Grande, M., & Roeder, J. L. (2002). Statistical distribution of the storm-time proton ring current: POLAR measurements. *Geophysical Research Letters*, *29*(20), 1969. https://doi.org/10.1029/2002GL015430

Ebihara, Y., & Fok, M. C. (2004). Postmidnight storm-time enhancement of tens-of-keV proton flux. *Journal of Geophysical Research*, *109*, A12209. https://doi.org/10.1029/2004JA010523

Ebihara, Y., Fok, M. C., Sazykin, S., Thomsen, M. F., Hairston, M. R., Evans, D. S., & Ejiri, M. (2005). Ring current and the magnetosphere-ionosphere coupling during the superstorm of 20 November 2003. *Journal of Geophysical Research*, *110*, A09S22. https://doi.org/10.1029/2004JA010924

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

Fok, M. C., Moore, T. E., Slinker, S. P., Fedder, J. A., Delcourt, D. C., Nos, M., & Chen, S. H. (2011). Modeling the superstorm in November 2003. *Journal of Geophysical Research*, *116*, A00J17. https://doi.org/10.1029/2010JA015720

Harang, L. (1946). The mean field of disturbance of polar geomagnetic storms. *Terrestrial Magnetism and Atmospheric Electricity*, *51*(3), 353–380. https://doi.org/10.1029/TE051i003p00353

Iijima, T., & Potemra, T. A. (1976). The amplitude distribution of field-aligned currents at northern high latitudes observed by Triad. *Journal of Geophysical Research*, *81*(13), 2165–2174. https://doi.org/10.1029/JA081i013p02165

Jordanova, V. K., Thorne, R. M., Farrugia, C. J., Dotan, Y., Fennell, J. F., Thomsen, M. F., & McComas, D. J. (2001). Ring current dynamics during the 13–18 July 2000 storm period. *Solar Physics*, *204*(1), 361–375. https://doi.org/10.1023/A:1014241527043

Kubat, M. (2015). An introduction to machine learning. Springer Publishing Company, Incorporated.

Lakhina, G. S., Alex, S., Tsurutani, B. T., & Gonzalez, W. D. (2013). Supermagnetic storms: Hazard to society. *Extreme events and natural hazards: The complexity perspective* (pp. 267–278). Washington DC: American Geophysical Union (AGU). https://doi.org/10.1029/2011GM001073

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lepping, R., Berdichevsky, D., Burlaga, L., Lazarus, A., Kasper, J., Desch, M., & Ackerson, K. (2001). The Bastille Day magnetic clouds and upstream shocks: Near-earth interplanetary observations. *Solar Physics*, *204*(1), 285–303. https://doi.org/10.1023/A:1014264327855

Liemohn, M. W., Ganushkina, N. Y., Katus, R. M., De Zeeuw, D. L., & Welling, D. T. (2013). The magnetospheric banana current. *Journal of Geophysical Research: Space Physics*, *118*, 1009–1021. https://doi.org/10.1002/jgra.50153

Liemohn, M. W., Kozyra, J. U., Hairston, M. R., Weimer, D. R., Lu, G., Ridley, A. J., & Skoug, R. M. (2002). Consequences of a saturated convection electric field on the ring current. *Geophysical Research Letters*, *29*(9), 62–1-62-4. https://doi.org/10.1029/2001GL014270

Liemohn, M. W., Kozyra, J. U., Thomsen, M. F., Roeder, J. L., Lu, G., Borovsky, J. E., & Cayton, T. E. (2001). Dominant role of the asymmetric ring current in producing the stormtime Dst*. *Journal of Geophysical Research*, *106*(A6), 10,883–10,904. https://doi.org/10.1029/2000JA000326

Mauk, B. H., Fox, N. J., Kanekal, S. G., Kessel, R. L., Sibeck, D. G., & Ukhorskiy, A. (2013). Science objectives and rationale for the radiation belt storm probes mission. *Space Science Reviews*, *179*(1), 3–27. https://doi.org/10.1007/s11214-012-9908-y

Merkin, V. G., Sorathia, K., Sitnov, M. I., Stephens, G. K., Lyon, D., Garretson, J., & Ukhorskiy, A. Y. (2019). Global model of the storm-time magnetosphere with empirical ring current pressure. Fall AGU Meeting, Presentation SM12B-04.

Mitchell, T. (1997). *Machine learning*. New York: McGraw Hill.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.

Ngwira, C. M., Pulkkinen, A., Kuznetsova, M. M., & Glocer, A. (2014). Modeling extreme Carrington-type space weather events using three-dimensional global MHD simulations. *Journal of Geophysical Research: Space Physics*, *119*, 4456–4474. https://doi.org/10.1002/2013JA019661

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C* (2nd Ed.): *The art of scientific computing*. USA: Cambridge University Press.

Raeder, J., McPherron, R. L., Frank, L. A., Kokubun, S., Lu, G., Mukai, T., & Slavin, J. A. (2001). Global simulation of the Geospace Environment Modeling substorm challenge event. *Journal of Geophysical Research*, *106*(A1), 381–395. https://doi.org/10.1029/2000JA000605

Raeder, J., Wang, Y. L., Fuller-Rowell, T. J., & Singer, H. J. (2001). Global simulation of magnetospheric space weather effects of the Bastille Day storm. *Solar Physics*, *204*, 323–337. https://doi.org/10.1023/A:1014228230714

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

Roelof, E. C. (1989). Remote sensing of the ring current using energetic neutral atoms. *Advances in Space Research*, *9*(12), 195–203. https://doi.org/10.1016/0273-1177(89)90329-3

Roelof, E. (1997). ENA emission from nearly-mirroring magnetospheric ions interacting with the exosphere. *Advances in Space Research*, *20*(3), 361–366. https://doi.org/10.1016/S0273-1177(97)00692-3

Russell, C. T. (1972). The configuration of the magnetosphere. In E. R. Dyer (Ed.), *Critical problems of magnetospheric physics* (pp. 1–16). Washington, D. C., USA: National Academy of Sciences.

Sergeev, V. A., Pulkkinen, T. I., Pellinen, R. J., & Tsyganenko, N. A. (1994). Hybrid state of the tail magnetic configuration during steady convection events. *Journal of Geophysical Research*, *99*(A12), 23,571–23,582. https://doi.org/10.1029/94JA01980

Sitnov, M. I., Stephens, G. K., Gkioulidou, M., Merkin, V., Ukhorskiy, A. Y., Korth, H., & Tsyganenko, N. A. (2018). Empirical modeling of extreme events: Storm-time geomagnetic field, electric current, and pressure distributions. In N. Buzulukova (Ed.), *Extreme Events in Geospace* (pp. 259–279). Cambridge, MA: Elsevier. https://doi.org/10.1016/B978-0-12-812700-1.00011-X

Sitnov, M. I., Stephens, G. K., Tsyganenko, N. A., Miyashita, Y., Merkin, V. G., Motoba, T., & Genestreti, K. J. (2019). Signatures of nonideal plasma evolution during substorms obtained by mining multimission magnetometer data. *Journal of Geophysical Research: Space Physics*, *124*, 8427–8456. https://doi.org/10.1029/2019JA027037

Sitnov, M. I., Stephens, G. K., Tsyganenko, N. A., Ukhorskiy, A. Y., Wing, S., Korth, H., & Anderson, B. J. (2017). Spatial structure and asymmetries of magnetospheric currents inferred from high-resolution empirical geomagnetic field models, *Dawn-dusk asymmetries in planetary plasma environments* (pp. 199–212). Washington DC: American Geophysical Union (AGU). https://doi.org/10.1002/9781119216346.ch15

Sitnov, M. I., Tsyganenko, N. A., Ukhorskiy, A. Y., Anderson, B. J., Korth, H., Lui, A. T. Y., & Brandt, P. C. (2010). Empirical modeling of a CIR-driven magnetic storm. *Journal of Geophysical Research*, *115*, A07231. https://doi.org/10.1029/2009JA015169

Sitnov, M. I., Tsyganenko, N. A., Ukhorskiy, A. Y., & Brandt, P. C. (2008). Dynamical data-based modeling of the storm-time geomagnetic field with enhanced spatial resolution. *Journal of Geophysical Research*, *113*, A07218. https://doi.org/10.1029/2007JA013003

Sitnov, M. I., Ukhorskiy, A. Y., & Stephens, G. K. (2012). Forecasting of global data-binning parameters for high-resolution empirical geomagnetic field models. *Space Weather*, *10*, S09001. https://doi.org/10.1029/2012SW000783

Stephens, G. K., Bingham, S. T., Sitnov, M. I., Gkioulidou, M., Merkin, S., Korth, H., & Ukhorskiy, A. Y. (2020). Storm-time plasma pressure inferred from multi-mission magnetometer measurements and its validation using Van Allen Probes particle data. *Space Weather*.

Stephens, G. K., Sitnov, M. I., Kissinger, J., Tsyganenko, N. A., McPherron, R. L., Korth, H., & Anderson, B. J. (2013). Empirical reconstruction of storm time steady magnetospheric convection events. *Journal of Geophysical Research: Space Physics*, *118*, 6434–6456. https://doi.org/10.1002/jgra.50592

Stephens, G. K., Sitnov, M. I., Korth, H., Tsyganenko, N. A., Ohtani, S., Gkioulidou, M., & Ukhorskiy, A. Y. (2019). Global empirical picture of magnetospheric substorms inferred from multimission magnetometer data. *Journal of Geophysical Research: Space Physics*, *124*, 1085–1110. https://doi.org/10.1029/2018JA025843

Stephens, G. K., Sitnov, M. I., Ukhorskiy, A. Y., Roelof, E. C., Tsyganenko, N. A., & Le, G. (2016). Empirical modeling of the storm time innermost magnetosphere using Van Allen Probes and THEMIS data: Eastward and banana currents. *Journal of Geophysical Research: Space Physics*, *121*, 157–170. https://doi.org/10.1002/2015JA021700

Thébault, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., & Zvereva, T. (2015). International geomagnetic reference field: The 12th generation. *Earth, Planets and Space*, *67*(1), 79. https://doi.org/10.1186/s40623-015-0228-9

Tsyganenko, N. A. (1996). Effects of the solar wind conditions on the global magnetospheric configurations as deduced from data-based field models. In *Proceedings of the Third International Conference on Substorms (ICS-3)* (pp. 181–185).
Tsyganenko, N. A. (2002). A model of the near magnetosphere with a dawn-dusk asymmetry 1. Mathematical structure. *Journal of Geophysical Research*, *107*(A8), 1179. https://doi.org/10.1029/2001JA000219
Tsyganenko, N. A., Singer, H. J., & Kasper, J. C. (2003). Storm-time distortion of the inner magnetosphere: How severe can it get? *Journal of Geophysical Research*, *108*(A5), 1209. https://doi.org/10.1029/2002JA009808
Tsyganenko, N. A., & Sitnov, M. I. (2007). Magnetospheric configurations from a high-resolution data-based magnetic field model. *Journal of Geophysical Research*, *112*, A06225. https://doi.org/10.1029/2007JA012260
Vassiliadis, D. (2006). Systems theory for geospace plasma dynamics. *Reviews of Geophysics*, *44*, RG2002. https://doi.org/10.1029/2004RG000161
Wiltberger, M., Merkin, V., Zhang, B., Toffoletto, F., Oppenheim, M., Wang, W., & Stephens, G. K. (2017). Effects of electrojet turbulence on a magnetosphere-ionosphere simulation of a geomagnetic storm. *Journal of Geophysical Research: Space Physics*, *122*, 5008–5027. https://doi.org/10.1002/2016JA023700
Wolf, R., Spiro, R., Sazykin, S., & Toffoletto, F. (2007). How the Earth's inner magnetosphere works: An evolving picture. *Journal of Atmospheric and Solar-Terrestrial Physics*, *69*(3), 288–302. https://doi.org/10.1016/j.jastp.2006.07.026
Wolf, R. A., Spiro, R. W., Sazykin, S., Toffoletto, F. R., & Yang, J. (2016). Forty-seven years of the rice convection model, *Magnetosphere-ionosphere coupling in the solar system* (pp. 215–225). Washington DC: American Geophysical Union (AGU). https://doi.org/10.1002/9781119066880.ch17
Yang, J., Wolf, R., Toffoletto, F., Sazykin, S., Wang, W., & Cui, J. (2019). The inertialized rice convection model. *Journal of Geophysical Research: Space Physics*, *124*, 10,294–10,317. https://doi.org/10.1029/2019JA026811
Zaharia, S., & Cheng, C. Z. (2003). Can an isotropic plasma pressure distribution be in force balance with the T96 model field? *Journal of Geophysical Research*, *108*(A11), 1412. https://doi.org/10.1029/2002JA009501
Zaharia, S., Jordanova, V. K., Thomsen, M. F., & Reeves, G. D. (2006). Self-consistent modeling of magnetic fields and plasmas in the inner magnetosphere: Application to a geomagnetic storm. *Journal of Geophysical Research*, *111*, A11S14. https://doi.org/10.1029/2006JA011619