JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1002/,

Reconstruction of magnetospheric storm-time dynamics using cylindrical basis functions and multi-mission data mining

N. A. Tsyganenko[1], V. A. Andreeva[1], and M. I. Sitnov[2]

Key points:

- A novel method is developed to represent distant geomagnetic field by generalizing the previously proposed radial basis function approach
- The new model is combined with the nearest-neighbor data mining to reconstruct magnetic structures from a large pool of multi-mission data
- Modeling a strong storm with sudden commencement revealed a dramatic redistribution of magnetic flux due to the external impact and driving

Corresponding author: N. A. Tsyganenko, Institute and Department of Physics, Saint Petersburg State University, Saint-Petersburg, Russia (nikolai.tsyganenko@gmail.com or n.tsyganenko@spbu.ru)

[1]Saint-Petersburg State University, Institute and Department of Physics, Saint-Petersburg, Russia.

[2]Applied Physics Laboratory, The Johns Hopkins University, Laurel, Maryland, USA.

Abstract. First results are presented of the modeling of magnetospheric storm events, based on: (i) a new method to represent the magnetic field by means of the so-called cylindrical basis functions (CBF), (ii) the data mining approach by Sitnov et al. (2008), and (iii) upgraded and extended pool of multi-mission data taken in 1995–2019. The study is focused on the low-latitude magnetospheric domain in the distance range 3–18R_E bounded by field line shells with footpoint latitudes $\pm 70°$. The magnetic configurations are reconstructed from data subsets, selected from the grand database by the nearest-neighbor method, using both interplanetary data and the ground disturbance indices. A strong storm of May 27–29, 2017, has been studied in relation to its effect on the reconfiguration of the low-latitude magnetosphere. The modeling reproduces the main features of the magnetosphere dynamics in terms of the geomagnetic field depression/compression and extremely variable field line stretching. The initial contraction of the magnetosphere during the storm sudden commencement results in a local transient surge of the inner tail current and a dramatic antisunward discharge of the magnetic flux. As the storm progresses, the ring current buildup results in a strongly depressed magnetic field in the inner magnetosphere, which expels the magnetic flux to larger distances and increases the field line connection across the more distant tail plasma sheet. At the same time, a strong dawn-dusk asymmetry is developed due to the formation of the duskside partial ring current, in agreement with previous independent results.

1. Introduction

Since their inception in the early 1970s (Mead and Fairfield, 1975), the data-based magnetospheric models made a remarkable progress, from relatively simple analytical formulations describing the average geomagnetic field for several bins of the Kp-index (e.g., Tsyganenko and Usmanov, 1982), to much more sophisticated algorithms, capable to reproduce the storm-time dynamics of the magnetosphere (e.g., Tsyganenko and Sitnov, 2005; henceforth TS05). In early models (see a comprehensive bibliography in a review by Tsyganenko, 2013), a standard approach was to represent the magnetospheric field as a sum of contributions from only a few principal current systems, often referred to as "modules". The rapidly growing wealth of spacecraft data collected in the recent decades made it possible to cardinally revise the traditional method, which allowed to greatly increase the spatial and temporal resolution of the models. Tsyganenko and Sitnov (2007; henceforth TS07) partially abandoned the modular approach by replacing the custom-tailored field of equatorial currents with a much more flexible representation by sums of quasi-orthogonal basis functions.

The next significant step forward was to completely give up the custom-made modules and represent the magnetospheric field by expanding its toroidal and poloidal parts into series of terms based on radial basis functions (RBF). The new approach (Andreeva and Tsyganenko, 2016; henceforth AT16) was shown to successfully reproduce the magnetospheric field on the basis of then available data; in addition, a possibility was demonstrated to merge the RBF and modular models into a single 'hybrid' model (Tsyganenko and Andreeva, 2017), which allowed to overcome limitations of the purely RBF models, in particular, to realistically represent the highly structured field of Birkeland currents at low altitudes. With respect to the parameterization of the models, a major breakthrough (Sitnov et al., 2008) was to abandon the formerly adopted global approach based on representing the model parameters with *a priori* postulated functions of external drivers, derived from the entire grand pool of data. Instead, it was suggested to dynamically model the system's evolution during a specific event on the basis of data, mined from past and future events with similar scenarios of external input, comparable magnitudes, and temporal trends of the disturbance level. In that kind of approach, the final product is a sequence of models, representing a specific event of interest with a series of 'quasi-instantaneous' configurations, rather than an all-purpose universal code. Each individual model in the sequence is derived from a relatively small subset of data, selected from the grand pool on the basis of its proximity to the modeled situation, quantified by the sliding average values of state parameters (such as $\langle$SYM-H$\rangle$ index or the solar wind driver $\langle VB_z \rangle$) and their temporal trends, represented by their time derivatives.

In this paper we describe first results of a modeling study, whose goal is to synergistically combine an advanced modification of the RBF-like model with the data-mining approach to extract the maximum information from data. As the mathematical structure of the present model is significantly different from those developed in previous works, the paper starts with separate sections 2 and 3, containing a detailed description of the new basis functions and the grid. Section 4 overviews the data sets used in this study; special attention is paid to the new additions to our database. Section 5 presents results of testing the new model's performance on artificial data, by comparing the target and reconstructed fields and currents. Section 6 addresses the basic principles of creating the nearest-neighbors (NN) data subsets. Section 7 presents main results of the modeling of an intense geomagnetic storm of May 27–29, 2017, and their validation on independent data of geosynchronous satellites. Section 8 discusses the obtained results in terms of the field line reconfiguration, and the last section 9 summarizes the paper.

2. Cylindrical basis functions (CBF)

Before getting to specifics of the new model, we briefly remind the essence of the earlier developed AT16 approach. The field of external sources is represented as a sum of toroidal and poloidal parts:

$$\mathbf{B}^{(\mathrm{ext})}(\mathbf{r}) = \nabla \times (\Psi_t \mathbf{r}) + \nabla \times \nabla \times (\Psi_p \mathbf{r}) = \nabla \Psi_t \times \mathbf{r} + \nabla \times (\nabla \Psi_p \times \mathbf{r}) \tag{1}$$

where both generating potentials Ψ_t and Ψ_p are expanded into linear combinations of the RBFs $\chi_i(|\mathbf{r} - \mathbf{R}_i|)$, each of which depends only on the radial distance from its grid node $\mathbf{R}_i$.

In this work we introduce more flexible generalized basis functions, similar to the RBFs, but depending on the solar-magnetic (SM) cylindrical coordinates $\{\rho, \phi, z\}$ and centered around a set of nodes $\mathbf{r}_i$ located at positions $\{\rho_i, \phi_i, z_i\}$. Their form could be adopted in exactly the same way as for the RBFs, were it not for a complication due to the multi-valued property of the longitude ϕ. A simple remedy is to replace the squared azimuthal distance $\rho^2|\phi - \phi_i|^2$ with $\rho^2 \sin^2[(\phi - \phi_i)/2]$. The final form of the basis functions is adopted as a product of Gaussian distributions in ρ, ϕ, and z:

$$\chi_i(\rho, \phi, z, \rho_i, \phi_i, z_i) = \exp\left\{ -\frac{(\rho - \rho_i)^2}{D_\rho^2} - \frac{\rho^2 \sin^2[(\phi - \phi_i)/2]}{D_\phi^2} - \frac{(z - z_i)^2}{D_z^2} \right\} \tag{2}$$

In order to distinguish them from the formerly used RBFs, the new basis functions (2) will be termed henceforth as cylindrical basis functions (CBFs). Aside from the difference in the coordinate systems, they have three independent scale lengths, D_ρ, D_ϕ, D_z, which adds more flexibility to the model field. The principal motivation behind the adoption of the new CBFs is the inherently large difference between the radial, azimuthal, and North-South variation scales of the low-latitude geomagnetic field; this issue will be further discussed in the next Section 3.

The rest of the modeling formalism is fully analogous to that described in the earlier works (e.g., AT16, Eqs.(7–8)). Namely, the toroidal and poloidal potentials are expanded into sums

$$\Psi_t = \cos\psi \sum_{i=1}^{N} a_i(\chi_i^+ + \chi_i^-) + \sin\psi \sum_{i=1}^{N} b_i(\chi_i^+ - \chi_i^-) \tag{3}$$

$$\Psi_p = \cos\psi \sum_{i=1}^{N} c_i(\chi_i^+ - \chi_i^-) + \sin\psi \sum_{i=1}^{N} d_i(\chi_i^+ + \chi_i^-) \tag{4}$$

where ψ is the dipole tilt angle, a_i, b_i, c_i, and d_i are unknown model coefficients, and the basis functions $\chi_i^+ = \chi_i(\rho,\phi,z,\rho_i,\phi_i,z_i)$ and $\chi_i^- = \chi_i(\rho,\phi,z,\rho_i,\phi_i,-z_i)$. Being substituted in (1), the potentials (3–4) provide a model field whose components have the required mirror symmetry properties with respect to transformation $z \to -z$, $\psi \to -\psi$ (e.g., Mead and Fairfield, 1975, Eqs.(4–6)): $B_{x,y}(x,y,-z,-\psi) = -B_{x,y}(x,y,z,\psi)$ and $B_z(x,y,-z,-\psi) = B_z(x,y,z,\psi)$.

The total number of free parameters of the model field (1)–(4) equals $4N$, where N is the number of the grid nodes. Note that, due to the above North-South symmetry requirements, only the nodes in one hemispace are counted in that number. More details on the grid are given in the next section.

3. Numerical grid and the modeling domain

In most of our previous studies based on the RBF representation, axially symmetric nested grids were employed, whose nodes were placed on a set of concentric spherical layers according to Kurihara's (1965) method. The main objective behind that choice was to keep a reasonable trade-off between two apparently conflicting requirements: on the one hand, minimize the number of the grid nodes (hence, the computation time) and, on the other hand, maintain the model's resolution sufficiently high in all three dimensions. Initially, the latter requirement prompted us to keep the grid locally equidistant; later on, however, that restriction was realized as unnecessary. The first evidence of a possibility to use non-uniform grids came to our attention from a local modeling of the geosynchronous magnetic field (Andreeva and Tsyganenko, 2018), where a much wider azimuthal separation of the nodes did not result in any significant loss of accuracy. Later on, modified RBFs with largely different internode spacing were proposed and successfully used by Chen et al. (2019) to reconstruct complex simulated magnetic field structures with neutral points.

Based on the above, in this study we set out to construct an economical grid of nodes, which would take full advantage of the large difference between the characteristic variation scales in the radial, azimuthal, and North-South directions. Owing to the dominance of the nearly axisymmetric field of the Earth's dipole, it is natural to associate the largest variation scale with the solar-magnetic (SM) longitude ϕ. On the opposite end of this hierarchy is the variation across the equatorial plane, which can be very short-scaled on the nightside due to the presence of thin tail current sheet. Finally, the radial variation in $\vec{\mathbf{e}}_\rho$ direction takes the intermediate place in that sense. Note that, in the above ordering, we leave aside such transient short-lived features as the longitudinally structured small-scale finger-like convection streams associated with BBFs (e.g., Liu et al., 2013, and refs. therein), as well as relatively rare events with sharp boundaries between the inner magnetosphere and magnetotail (Apatenkov et al., 2008).

In this work, we concentrate exclusively on the low-latitude domain and leave out the vast high-latitude magnetosphere. To accurately delineate the modeling region, we followed the approach of a recent study of the IMF B_y 'penetration' effect (Tsyganenko and Andreeva, 2020; Sect.3.1, Eq.8), where both data and the RBF grid nodes were restricted to the inside of a domain, bounded by two funnel-like surfaces, separating the low-latitude magnetosphere from the tail lobes. Although the analytical form of the surfaces was described in detail in the above cited article and in earlier publications, we nevertheless reproduce it once again, in order to keep this paper self-contained:

$$\theta(r,\phi) = \arcsin\left[\frac{\sqrt{r}}{\left(r^{\nu} + \sin^{-2\nu}\theta_{o} - 1\right)^{1/2\nu}}\right] \tag{5}$$

Here the angle $\theta_o = 20^\circ$ is the footpoint colatitude of the Northern surface at the ground level ($r = 1$) and the constant parameter $\nu = 3$ controls the flaring rate of the surface and thus its distance from the equatorial plane in the tail. The Southern surface is mirror-symmetric to the Northern one, as long as the geodipole tilt angle ψ is zero. For $\psi \neq 0$, both surfaces are subject to a tilt-dependent deformation, similar to a bowl-shaped bending/warping of the magnetospheric equatorial current, as analytically represented in Tsyganenko and Andreeva (2014). In more detail, each point $\{r,\theta\}$ of the originally undeformed surface (5) receives an additional shift in the polar angle θ:

$$\Delta\theta(r,\theta) = -\sin(\theta)\arcsin\frac{Z_S(r,\psi)}{r} \tag{6}$$

where $Z_S(r,\psi)$ is the deviation of the bowl-shaped current surface from the SM equatorial plane:

$$Z_S(r,\psi) = R_H \tan\psi\left\{1 - \left[1 + \left(\frac{r}{R_H}\right)^{\alpha}\right]^{1/\alpha}\right\} \tag{7}$$

The parameter $R_H = 8\,R_E$ is the hinging distance and the exponent $\alpha = 3$ controls the smoothness of transition from the dipole- to solar-wind-dominated regimes. At small $r \ll R_H$ the solar wind influence is virtually zero and both surfaces almost rigidly follow the geodipole orientation (i.e., remain nearly fixed in the SM coordinates), but start to gradually slip behind at $r \geq R_H$ and, asymptotically, become nearly parallel to the distant tail plasma sheet at $r \gg R_H$. The $\sin\theta$ factor in (6) is introduced to gradually decrease the deformation magnitude away from the SM equator. Equations (5–7) define the Northern and Southern boundaries, which confine the CBF grid and data within a limited low-latitude region and symmetrically enclose the equatorial current sheet at any dipole tilt angle (see left panels in Figure 1 below). The modeling domain is also restricted (i) by the inner/outer spherical boundaries, located at geocentric distances $R_{\mathrm{inn}} = 3\,R_E$ and $R_{\mathrm{out}} = 18\,R_E$, respectively, and (ii) by the model magnetopause (Lin et al., 2010), calculated using average values of the wind pressure $P_d = 2\,\mathrm{nPa}$ and IMF $B_z = 0\,\mathrm{nT}$.

The CBF grid was constructed by creating first a planar subset of nodes in a single meridian plane. To that end, a sequence of 10 radial distances R_n ($R_{\mathrm{inn}} \leq R_n \leq R_{\mathrm{out}}$) was defined, with a linearly growing separation $\Delta R_n = R_{n+1} - R_n = \varepsilon R_n$. Then, for each R_n, 8 equidistant nodes were placed along a circular segment between the Northern and Southern bounding surfaces, defined by the above Eq.(5).

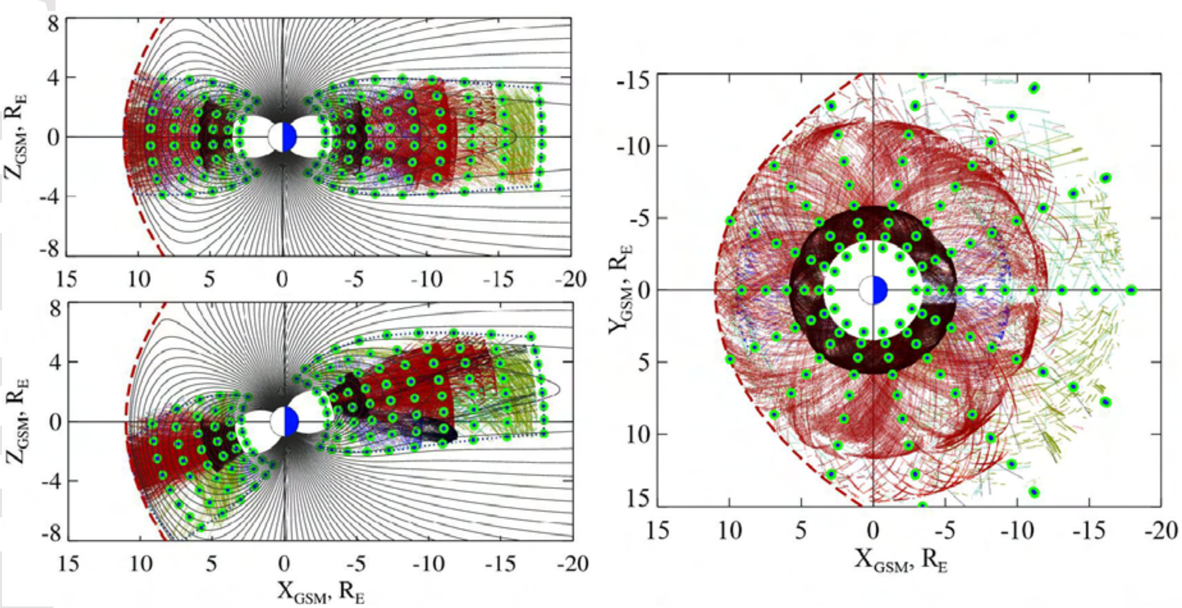

Figure 1. Distribution of spacecraft data (small colored specks) and CBF grid nodes (green circles) in the modeling region (delineated by dotted lines in the left panels). Left: meridional XZ views for untilted (top, $\psi = 0°$) and tilted (bottom, $\psi = 20°$) geodipole orientation. Right: equatorial XY view ($\psi = 0°$). Model magnetic field lines in the left panels and the magnetopause (dashed line) are shown only for the reader's orientation and correspond to an unrelated empirical model.

The generated meridional set of 80 nodes was then rotationally multiplied into 14 equally spaced longitude sectors. All the nodes that fell outside the model magnetopause were left out, which eventually produced a 3D grid with 960 nodes. Note that, due to the N-S symmetry properties (see end of previous section), only half of that number, $N = 480$, entered in the initial number of model parameters $4N = 1920$. The outcome of the above procedure is displayed in Figure 1, showing the nodes and the grid boundaries in the noon-midnight plane (left) and in the equatorial plane (right). The colored dots illustrate the distribution of data used in the model derivation, which we describe in more detail further below. To avoid excessive crowding of the data dots in the plots, only small fractions (∼10,000) of the entire set ($\geq 1,000,000$) lying in the immediate vicinity of meridional and equatorial planes were selected for plotting.

4. Data

Large multi-mission sets of historical data lie at the foundation of the empirical models and must be periodically upgraded, as long as new data become available with time. An essential part of this work was to significantly extend our previously compiled 1995–2016 archive by appending new data taken since 2016 through 2019, as well as adding older data that were left out in the earlier studies. Most of the data sets, contributing missions, and basic preparation procedures were already described at length in our previous papers (e.g., Tsyganenko and Andreeva, 2017, and refs. therein); to keep this article more or less self-contained, we briefly recapitulate and update that material below.

4.1. Interplanetary data

The solar wind, interplanetary magnetic field (IMF), and ground indices data were downloaded from the OMNI source (https://omniweb.gsfc.nasa.gov/form/omni_min.html) in the form of yearly files with 5-minute average values of all the parameters that characterize the external driving of the magnetosphere or its internal response. The next step was to partially degap the data by interpolating the IMF and plasma parameters over no-data intervals, whose duration did not exceed 6 hours. All such records were supplied with flags, in order to distinguish between the real and interpolated values.

In view of the need to use the OMNI data for mining from the magnetospheric data the nearest-neighbor (NN) subsets, the yearly OMNI files were merged into a single file, covering the entire 25-year-long interval from 1995 through 2019. From that interval, sliding sequences of gapless data “windows” were then formed, in order to calculate cosine-weighted averages of principal interplanetary drivers, ground disturbance indices, and their time derivatives, in a way similar to that detailed in (Sitnov et al., 2008, Eqs.4–6). More specifically, the applied averaging was either one-sided, i.e., based only on preceding data “trails” with length $T=2$ or $T=5$ hours, or symmetric, that is, including both previous and following intervals of the same duration. In effect, these options resulted in four alternative ‘grand’ OMNI files, covering the period 1995–2019 and containing in each record the following parameters: the ‘instantaneous’ (i.e., 5-minute average) IMF B_x, B_y, B_z, solar wind speed V, ram pressure P_d, proton density N_p, temperature T_p, $V \cdot B_z$, N-index (Newell et al., 2007), B-index (Boynton et al., 2011), ground-based indices Sym-H$_c$ (corrected for the magnetopause contribution), ASY-H (Iyemori, 1990), AL, and AE. Each of the above parameters was also supplied with its time average, calculated with the weight factor $\cos(\pi t/2T)$ both in the symmetric $\langle \ldots \rangle$ and in one-sided $\langle \ldots |$ mode, and, finally, with the time derivative, averaged over the same interval with the same cosine weight factor. The total numbers of records in the four output files are nearly the same, ranging from 2,559,885 for the symmetric averaging over the $[t-5\,\mathrm{hr}, t+5\,\mathrm{hr}]$ time intervals, to 2,582,451 for the one-sided averaging over $[t-2\,\mathrm{hr}, t]$ interval. As required by the NN technique, all the averaged parameters in the files were normalized by their r.m.s. values. It should finally be noted that about 12% of records in the files did not have valid AE/AL data, mostly because of the total gap from March 2018 to present.

4.2. Magnetospheric data

The magnetospheric data in our 'grand' archive are represented by space magnetometer observations onboard 14 spacecraft, most of which have already been used and described in previous publications. In the course of this study, the old database underwent a major extension by adding data obtained since 2016, as well as by including bulks of earlier data that were missing in previous versions of the archive. The magnetospheric data are represented by 5-minute average values of the magnetic field, along with the concurrent satellite positions, dates/UT, and the geodipole tilt angle values. The magnetospheric data were augmented with concurrent interplanetary parameters and ground disturbance indices, described above in Section 4.1. The data were organized into 14 files, each dedicated to an individual satellite. In total, the new database contains 8,840,460 records, equivalent to ∼84 years worth of space magnetometer observations, which is more than twice the size of the data pool used in our previous studies (4,377,329 data records; Tsyganenko and Andreeva, 2017).

As already detailed in Section 3, this study concentrates on only low-latitude magnetospheric region within a limited interval of distances. For that reason, only about a half of the entire grand data pool was used in the present work; more exactly, between 4,434,130 and 4,434,387 records for the symmetrical and one-sided weighted averaging schemes, respectively.

A detailed description of basic procedures involved in the data processing has already been given elsewhere (see a review: Tsyganenko, 2013, section 5). Below follows an updated concise synopsis of each mission and its share in the modeling database.

4.2.1. Geotail

This is the oldest mission represented in the database, launched in 1992 and, as of time of this writing, still operating and providing high-quality data. The timespan of all Geotail data included in the modeling database covers the period from 01/05/1995 through 11/09/2019 and the geocentric distance range from 8.08 to 49.09 R_E. The total number of 5-min data records contributed by Geotail equals 799,399, equivalent to $\sim$ 7.6 years worth of magnetospheric observations. For more details on the Geotail mission and its magnetometer experiment, the reader is referred to Nishida (1994) and Kokubun (1994).

4.2.2. Polar

The Polar mission was launched in March 1996 into a highly elliptical polar orbit with its apogee above the North polar cap. Due to the slow apsidal precession, the spacecraft apogee gradually rotated equatorward and into the Southern hemisphere, eventually covering the entire magnetosphere from low altitudes up to geocentric distance of 9.6 R_E during the whole cycle of solar activity, including its powerful maximum in the beginning of 2000s. The mission was terminated in April 2008 and contributed to the grand archive a total of 871,429 5-minute data records, equivalent to 8.3 years of observations inside the magnetosphere. A detailed description of the Polar magnetometer experiment can be found in (Russell et al., 1995).

4.2.3. Cluster

The Cluster mission (Balogh et al., 1997), launched in July-August of 2000, consists of four spacecraft with nearly the same high-inclination orbits and relatively close inter-probe separation, specially designed for studying small-scale magnetospheric phenomena. By contrast to our previous studies, which used only Cluster-3 data from 02/2001 through 02/2016 (606,609 five-minute records), in this work the Cluster part of the database was extended nearly 5.4-fold to 3,270,352: first, by adding Cluster-3 data through 09/2018 and, second, by including data of the other three spacecraft for the entire period from 01/30/2001 through 09/30/2018.

An important comment should be made here. Namely, in view of a relatively small inter-probe separation, one might question if the data of the four Cluster probes are sufficiently independent of each other and, hence, whether such an extension can really improve the informative value of the database for the large-scale modeling. In order to clarify this issue, we plotted in Figure 2 a histogram of maximum separation between the four probes, based on their one-minute ephemeris data for the entire time span 2001–2019 and for all radial distances between the perigee (from $\sim 1.3 R_E$ in 2010 to $\sim 6.7 R_E$ in 2018) and apogee (from $\sim 16.3 R_E$ in 2018 to $\sim 22.3 R_E$ in 2009).

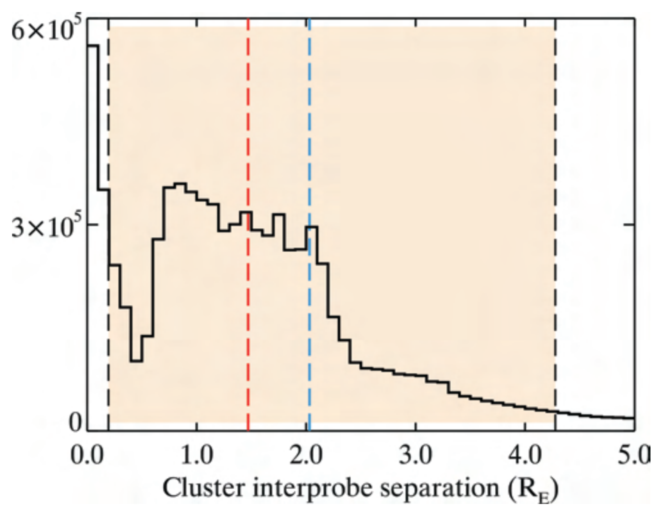

Figure 2. Histogram of maximum separation between the four probes for the entire timespan of the Cluster mission. The area between 10% and 90% percentile boundaries is highlighted with almond color, and the median/mean values are marked with red/blue dashed lines, respectively.

As can be seen from the plot, the mean value of the inter-probe separation (blue dashed line) is around $2R_E$, and 80% of its values lie in the range from $0.2R_E$ to $\sim 4R_E$. Since the satellites follow each other along the same trajectory at the speed from ∼1.5 km/s at apogee (where the separation is minimal) to ∼5–8 km/s at perigee (maximal separation) one roughly estimates their travel time lags to vary from ∼10–15 min to ∼1 hour, which is sufficient to treat the data of individual probes as largely independent from each other in space and time.

4.2.4. THEMIS

The data of five THEMIS probes used in this work covered the time interval from the mission launch date (February 2007) through August 31, 2019, and contributed a total of 2,863,887 five-minute records in the upgraded data archive. Most of the new data were those of THEMIS A, D, and E probes, while a relatively smaller portion came from B and C satellites, transferred in 2011 to lunar orbits under the ARTEMIS mission name. A comprehensive review of THEMIS mission, spacecraft orbits, and magnetometer experiment can be found in (Angelopoulos, 2008; Auster et al., 2008).

4.2.5. Van Allen Probes

Magnetometer data of the Van Allen Probes mission are an important asset for the empirical modeling, owing to their dense coverage of the low-latitude magnetosphere in the innermost distance range. The data came from two identical VAA and VAB spacecraft, following each other along the same orbit with apogees at $5.8 R_E$. The satellites were launched on August 30, 2012, and deactivated in July-October 2019. The total number of the VAA/VAB 5-minute data records in the grand database is 937,686, equivalent to 8.9 years of observations. More details on the magnetometer experiment onboard VAA and VAB spacecraft can be found in (Kletzing et al., 2013).

4.2.6. Magnetospheric Multiscale (MMS) data

The MMS data is the most recent addition to our archive. The mission is a constellation of four closely spaced satellites, similar to the Cluster tetrahedron, but with much tighter separation between the probes, which is why only one of them (MMS1) was used as a source of magnetometer data for this study. The mission was launched on March 12, 2015; the earliest and the most recent data included in the modeling database are dated by September 1, 2015, and July 31, 2019, respectively. The total number of MMS1 records in the present archive is 208,608; the data are concentrated mostly in the near-equatorial magnetosphere and cover a wide range of radial distances up to $R \approx 29 R_E$ on the nightside.

5. Testing the model on artificial data

The standard first step in implementing a new modeling method is to test its performance on a set of artificial data, created from a known model field as a source of synthetic 'data'. In this study, the new CBF representation (1)–(4) was tested by reconstructing the TS05 model field (Tsyganenko and Sitnov, 2005), generated at a set of locations randomly scattered inside the low-latitude modeling region shown above in Figure 1. The artificial data **B**-vectors were calculated using a set of TS05 input parameters, corresponding to a moderate disturbance; specifically, ram pressure $P_d = 2\,\text{nPa}$, IMF $B_z = -2\,\text{nT}$, SYM-H $= -40\,\text{nT}$, $W_{t1} = W_{t2} = 1.7$, $W_s = 1.3$, $W_p = 3.3$, $W_{b1} = 1.2$, $W_{b2} = 1.4$. This specific choice yielded a rather structured distribution of the target B-field and electric current, a sufficiently strong challenge for the model in terms of its flexibility and ability to reproduce complex configurations.

In the same fitting experiment, we also determined an optimal combination of the scaling factors D_ρ, D_ϕ, and D_z, entering in (2). More specifically, it was assumed that the above factors are not 'universal constants', but linearly vary with the radial distance r_i of each node, so that

$$D_\rho^{(i)} = D_\rho + \Delta D_\rho r_i\,, \quad D_\phi^{(i)} = D_\phi + \Delta D_\phi r_i\,, \quad \text{and} \quad D_z^{(i)} = D_z + \Delta D_z r_i. \tag{8}$$

The six nonlinear parameters D_ρ, ΔD_ρ, D_ϕ, ΔD_ϕ, D_z, and ΔD_z were treated as free unknowns and fitted to the artificial data along with the 1920 unknown coefficients entering linearly in (3)–(4). The fitting was carried out using a combined linear/nonlinear iterative code, based on a Nelder-Mead simplex algorithm for nonlinear parameters and a singular value decomposition (SVD) routine for the coefficients. The best-fit solution yielded the residual r.m.s. deviation between the model and target fields equal to 1.9 nT, that is, $\sim 3\%$ of the r.m.s magnitude 66.86 nT of the external target field. The obtained values of the six nonlinear scaling parameters in (8) were then assumed fixed for the adopted CBF grid and used in all further experiments on real data, described below in the following sections.

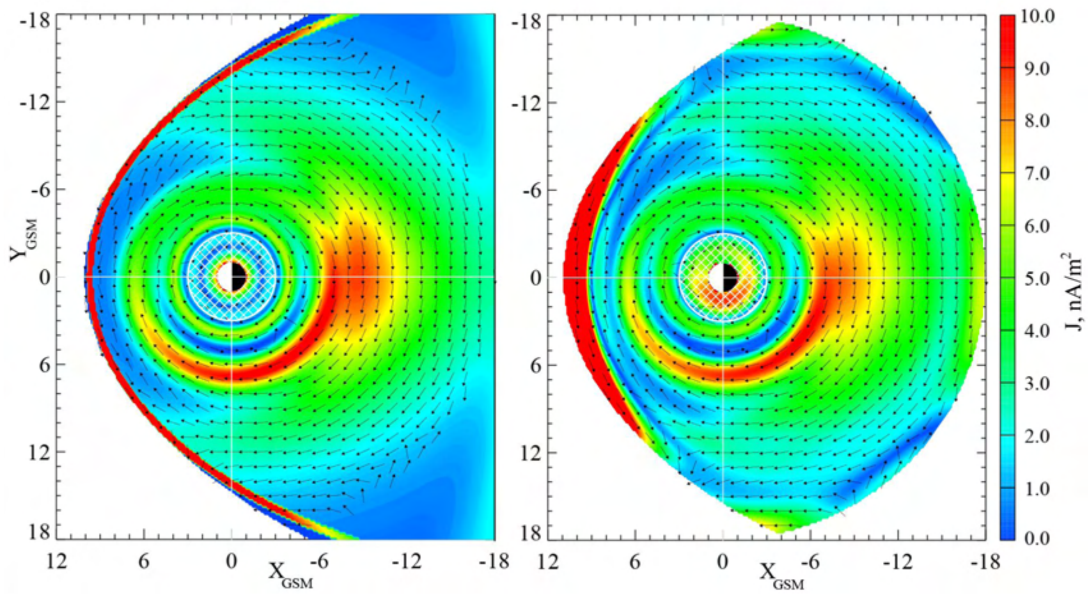

Figure 3. Comparing the equatorial electric currents, corresponding to the target TS05 field (left) and its reconstruction based on the CBF model (right). The current intensity is color-coded and the flow direction indicated by arrows of equal length. The innermost area within $r \leq 3R_E$ (white cross-hatching) is devoid of the grid nodes and does not belong to the modeling region.

Ideally, one might consider including the above six parameters (Eq.8) in the modeling of real events and find their values for each time moment of an event by the same nonlinear iterative code. In practice, however, that option is unfeasible due to the large number of the model coefficients and, hence, long computation times, even for the purely linear model. For that reason, those parameters and their radial gradients were derived only once, on the basis of an average simulated configuration. The purpose here was to approximately evaluate the parameters, on assumption that in individual real situations they remain reasonably close to the obtained values.

As an illustration of the model's performance, Figure 3 compares equatorial distributions of the volume density of the electric current $\mathbf{j} = (\nabla \times \mathbf{B})/\mu_o$, derived from the target $\mathbf{B}$ vector (left) and from the fitted CBF model field (right). In general, due to the curl operation the model current patterns are usually more structured than those of the corresponding magnetic field. In this case, however, the diagrams are surprisingly similar to each other throughout the entire modeling domain, except for the innermost extrapolation area $r \leq 3\,R_E$ (cross-hatched white circle), which does not belong to the modeling region and, as such, does not contain any grid nodes nor the 'data'. The other extrapolation area is the dayside boundary around the TS05 magnetopause. Here the eastward Chapman-Ferraro current calculated from the target field is significantly thinner than that reproduced by the CBF model, which is naturally explained by the absence of the magnetopause as such in the latter case. Similar plots have also been created to compare the target/model distributions of the magnetic field itself in the form of $\Delta B = |\mathbf{B}_{\text{total}}| - |\mathbf{B}_{\text{dipole}}|$ diagrams. The obtained distributions were found virtually identical to each other and are not reproduced here to save page space.

6. Parameters for the NN selection

As explained in detail in earlier publications (Sitnov et al., 2008, 2018, 2019; Stephens et al., 2019), the essence of the nearest-neighbor data mining is based on defining a state vector

$$\mathbf{G} = \{\langle A_1\rangle, \langle A_2\rangle, ..., \langle A_M\rangle, \langle \dot{\mathrm{A}}_1\rangle, \langle \dot{\mathrm{A}}_2\rangle, ..., \langle \dot{\mathrm{A}}_N\rangle\} \quad (N \leq M) \tag{9}$$

where the slide-averaged parameters $\langle A_i\rangle$ and their time derivatives $\langle \dot{\mathrm{A}}_i\rangle$ determine the current state and evolution trends (i) of the magnetosphere itself, usually derived from the ground-based activity indices, and (ii) of the incoming solar wind, observed by the interplanetary monitors.

The availability of a large pool of archived past observations allows one to compensate the lack of data for an event of interest by complementing them with data taken during different events, but similar to the situation under study in terms of proximity of their vectors (9) to that for the current time moment. To that end, all the components of the state vector (9) are normalized by their r.m.s. magnitudes, which allows to quantify the proximity of an archived data sample represented by a state vector $\mathbf{G}_{\mathrm{NN}}$ to the modeled state $\mathbf{G}$ in terms of the distance between them in the parametric hyperspace

$$\Delta G = \|\mathbf{G} - \mathbf{G}_{\mathrm{NN}}\| = \left\{ \sum_{i=1}^{M} \left[\langle A_i \rangle - \langle A_i \rangle_{\mathrm{NN}} \right]^2 + \sum_{i=1}^{N} \left[\langle \dot{\mathrm{A}}_i \rangle - \langle \dot{\mathrm{A}}_i \rangle_{\mathrm{NN}} \right]^2 \right\}^{1/2} \quad (10)$$

A data record (either from the current event or belonging to another one from the historical archive) is deemed close to the modeled state and, hence, is selected into a NN subset on the condition that the distance ΔG defined from (10) is less than a critical threshold ΔG_c. The optimal choice of a specific ΔG_c value depends on the complexity of a model, quantified by the number of its free parameters (degrees of freedom) and on the local density of data points in the parametric space for the time moment of interest. Too large values of ΔG_c lead to undesirable mixture of too different states in the subset, which smears the obtained solution and results in a loss of interesting details in the reconstructed field. On the opposite extreme, too small ΔG_c provide too little data in the subset and, hence, result in overfitted unstable solutions and artificial unphysical features, often hard to tell from the real ones. These issues have been addressed in more detail by Sitnov et al. (2019); in the present study, specific values of the threshold ΔG_c were set experimentally in such a way that the number of records in the NN samples be roughly 5–10 times larger than the number of free model parameters.

As the state vector $\mathbf{G}$ and the corresponding selection hypersphere move in the parametric space with time, some neighbour data records exit from the NN subset, while new ones enter in, generating thus a sequence of NN subsets, covering the interval of interest. Fitting a model to each subset in the sequence yields consecutive sets of model parameters and field configurations, representing the dynamics of the magnetosphere during the event. A major bottleneck in the modeling of magnetospheric storms, recognized still in early efforts (e.g., Tsyganenko et al., 2003), is the scarcity of data taken during disturbed periods, which becomes progressively acute with growing storm intensity. In our recent study to be described in more detail in a separate publication, an efficient method has been proposed to tackle the data paucity. Its essence is to provide each data sample with a weight factor, quantifying the degree of its proximity to the modeled state of the magnetosphere. This allows one to increase the statistical contribution of data taken during the most similar events and, at the same time, to keep the number of data points in the NN subsets at a reasonable minimum and thus avoid the overfitting. In this work, we have implemented that approach by assigning to each NN data record a weight $W \sim \exp(-\Delta G/\Delta G_c)$, varying between $W = 1$ and $W = 1/e = 0.368$ for the closest and farthest data records in the NN subset, respectively.

As detailed above in Section 4.1, the compiled OMNI and magnetospheric data files include a number of external parameters and ground-based indices, which offers a large variety of possible candidate variables to be used in the NN subset selection. Choosing an optimal combination is also not quite straightforward; first, it is intuitively clear that preference should be given to weakly correlated parameters, which ensures a more objective identification of really close states of the system. Second, a less obvious and maybe even counterintuitive fact is that the total number of the selection parameters should be kept at a reasonable minimum. Indeed, as demonstrated by Verleysen and François (2005), the volume of a parametric hypersphere dramatically falls down with growing space dimensionality, which may result in problems when using the Euclidian norm (10) with too large M. In application to the empirical magnetospheric modeling, the above problem belongs to a vast and largely untapped area of research, extending far beyond the scope of the present work.

In this study, a number of possible options was tried with different combinations of the selection parameters, averaging lengths/modes, and cutoff thresholds G_c defining the NN subset size. All those variants provided qualitatively similar results in terms of the storm-time evolution of the magnetospheric configurations, but differed with respect to quantitative details. In Section 7 below, modeling results will be presented for two specific choices of the NN search parameters, the first of which (Variant 1, henceforth V1 for short) used only upstream interplanetary parameters, while the second one (V2) was based mainly on the ground activity indices; the most essential details of both variants are described in the next subsections.

6.1. Variant 1: NN selection based on only interplanetary parameters

In this case, we used the following set of four selection parameters: the solar wind speed $\langle V|$ and the coupling index $\langle \mathscr{B}|$, based on the coupling function by Boynton et al. (2011), as well as their average time derivatives, $\langle dV/dt|$ and $\langle d\mathscr{B}/dt|$. Here $\mathscr{B} \propto N_{\mathrm{p}}^{1/2} V^{5/2} B_{\mathrm{t}} \sin^6(\theta_{\mathrm{c}}/2)$, where N_{p}, B_{t}, and θ_{c} are the solar wind proton number density, IMF transverse component and its clock angle, respectively. The rationale behind using the $\mathscr{B}$-index was its highest correlation with the SYM-H index, reported in the original work by Boynton et al., which implies its closest relevance to the state of the low-latitude magnetosphere. As prompted by the shape of the brackets $\langle ...|$, the averaging was made over time intervals, immediately preceding the data record; the trailing interval length was set equal to 5 hours, in order to average out substorm effects (Sitnov et al., 2008). The solar wind ram pressure P_d was not included in the NN search parameters but, in analogy to the TS07 model, its effects were taken into account by expanding the CBF coefficients in (3)–(4) into binomials of the form $p+q(P_d/\langle P_d\rangle)^{1/2}$. Based on the fact of relatively fast magnetospheric response to the pressure variations, it was not averaged over long trailing intervals, but retained in the original form of 5-minute averages. The splitting of the CBF coefficients into binomials doubled the number of model unknowns from 1920 to 3840; accordingly, the numbers of records in the NN subsets were set in each case around 30,000, that is, within a factor of 7–10 larger than the number of degrees of freedom.

6.2. Variant 2: NN selection based on ground disturbance indices

In this case, most of the NN search parameters were represented by the averaged and normalized indices of the ground geomagnetic activity and their time derivatives. Specifically, the following five parameters were employed: $\langle$SYM-H$_c|$, $\langle$ASY-H$|$, $\langle d$SYM-H$_c/dt|$, $\langle d$ASy-H$/dt|$, and $\langle$AL$|$, with the same 5-hour length of the averaging interval. The SYM-H$_c$ index is corrected for the solar wind ram pressure effect and quantifies the axisymmetrical contribution of the ring current (RC) to the ground field. The ASY-H index serves as a quantitative measure of the RC asymmetry (partial RC), a sensitive indicator of the storm phase. As for the solar wind pressure, in the first experiments it was completely excluded from the model, which, as expected, resulted in significantly poorer correlations between the reconstructed magnetic field and data, larger residuals, and weaker storm effects in the model field. In view of that, the solar wind pressure was eventually retained in this version in the same way as in V1; in addition, it implicitly entered into the NN data mining via the pressure correction term in SYM-H$_c = 0.8 \cdot$ Sym-H $- 13\sqrt{P_d}$ (Tsyganenko, 1996).

7. Results

The above described two variants of the model, V1 and V2, were tested by reconstructing the evolution of the low-latitude magnetospheric configuration during an intense storm of May 27-29, 2017.

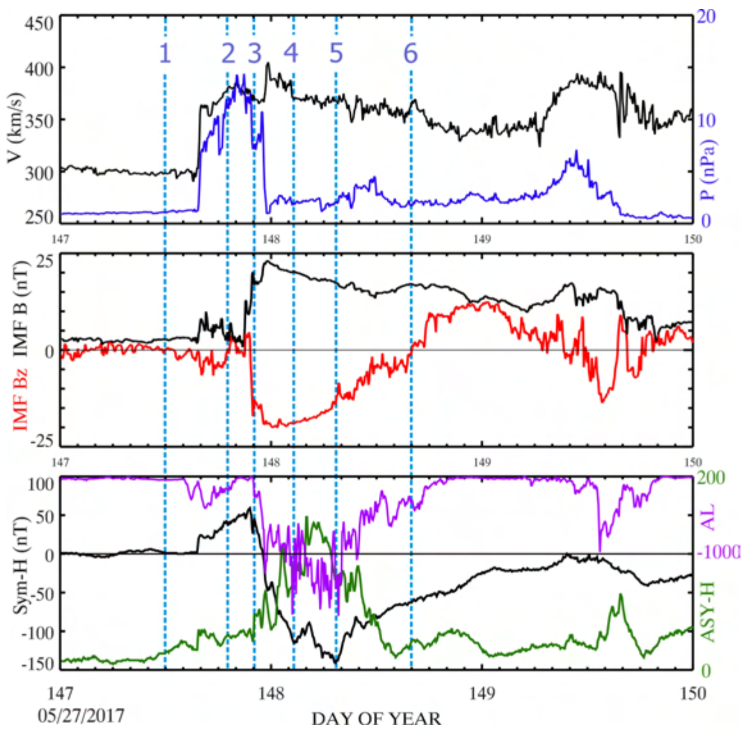

Figure 4. Interplanetary and ground disturbance data during the storm of May 27–31, 2017

This is a classic example of a CME-driven storm, preceded by a strong pulse of the solar wind pressure, followed by a sudden IMF reversal to south, and then its slow rotation to north. Three panels of Figure 4 show the variation of interplanetary parameters and the disturbance indices SYM-H, ASY-H, and AL during three days of the event (DOY 147–149). Six vertical dashed lines mark six time moments, for which the modeling was performed, from the quiet pre-storm state at 12:00 of DOY 147 (1), to sudden commencement with $P_d = 11.5\,\text{nPa}$ at 19:00 (2), to the southward IMF reversal at 22:00 (3), first negative peak of SYM-H (4), its second (absolute) minimum (5), and (6) middle of the recovery phase, coincident with the moment of IMF transition to northward orientation.

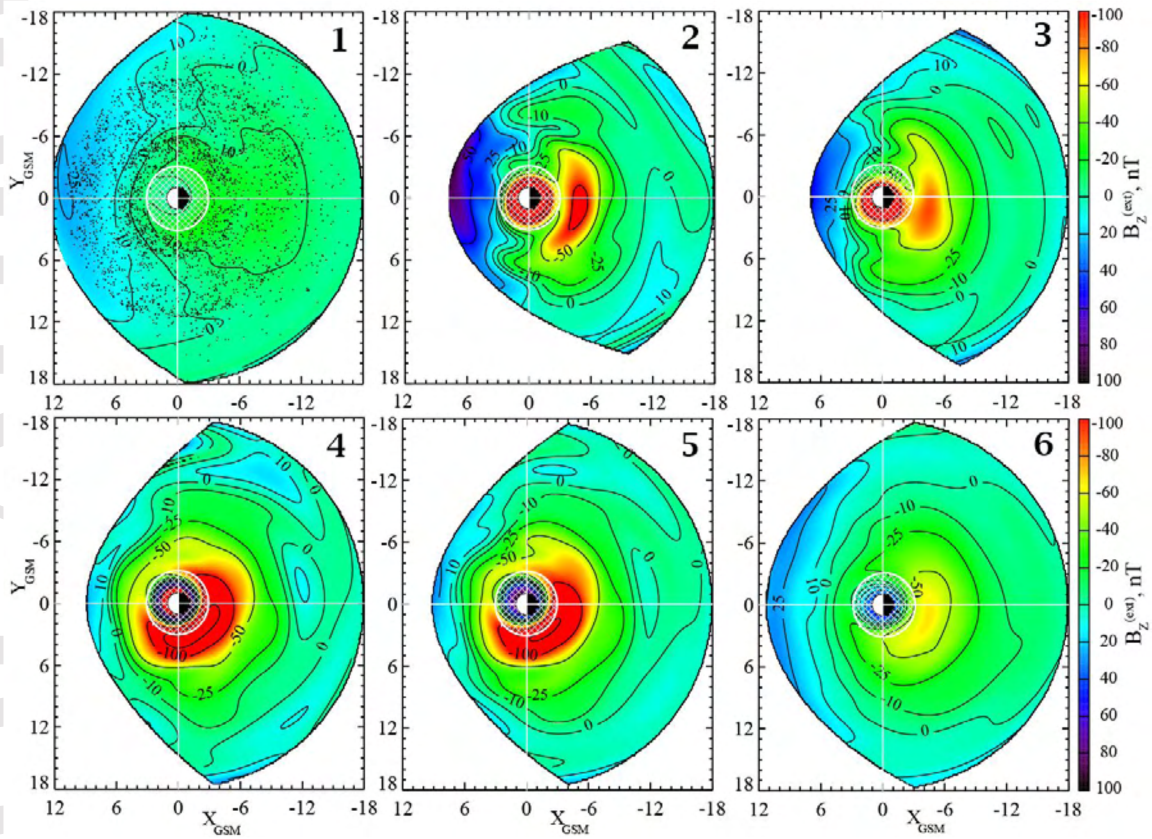

Figure 5. Equatorial plots of the external part $B_z^{(\text{ext})}$ of the total magnetic field for six time moments of the storm of 05/27–31, 2017, based on the V1 data subsets. Black specks in panel 1 illustrate the spatial density of the corresponding NN data samples (only every 10th data point is shown). The size/shape of the model magnetopause is based on Lin et al. (2010) model.

7.1. Reconstruction on the basis of only interplanetary data (variant V1)

Figure 5 shows equatorial plots of the external part $B_z^{(\text{ext})}$ of the total field, corresponding to contributions from all its external sources, derived from the model (1). The color coding represents the distribution of the North-South component of the model field and, at the same time, illustrates the degree $\Delta B = B_z^{(\text{total})} - B_z^{(\text{dipole})}$ of the magnetic field depression (black/blue) or compression (red/yellow). The six distributions are derived from six NN subsets and numbered in the same order as the corresponding time moments in Figure 4. The pre-storm configuration in panel 1 reveals a typical well-ordered quiet-time magnetic field, with a moderate compression on the day side (subsolar $\Delta B \sim +25\,\mathrm{nT}$) and a nearly symmetric near-tail depression ($\Delta B \sim -10\,\mathrm{nT}$ at $X \sim -12\,R_E$). To illustrate the coverage with the NN subset samples (each with ≈30,000 data points), locations of every 10th data record are shown by black dots. The data distributions in the rest five cases are qualitatively similar and are not shown to avoid obstructing the plots. The arrival of the shock front (panel 2, UT=19:00 on DOY=147) and the ensuing SC are accompanied by a compression of the magnetosphere with a strong field increase at the dayside ($\Delta B \sim 90\,\mathrm{nT}$ at the subsolar point). At the same time, one sees a sharp radially narrow depression at $X \sim -5\,R_E$ around midnight, extending by a few LT hours to dawn and dusk and exceeding $\Delta B \sim -100\,\mathrm{nT}$ in magnitude. At that time, the IMF B_z was still slightly positive and, as the ΔB diagrams do not reveal any sign of the RC development, the only conceivable reason for such a localized, abrupt, and dramatic depression is the sudden compression of the inner magnetotail, accompanied with formation of an intense and radially limited westward current in that area (see plots and a discussion in Section 8 below).

By ~22:00 UT (panel 3), the solar wind ram pressure subsides down to 7.4 nPa, the midnight depression weakens to $\Delta B \sim -80\,\mathrm{nT}$ and expands dawnward. We note in passing that in other experiments with different choices of the NN selection parameters (not described in this paper), a more azimuthally symmetric expansion of the inner midnight depression was found, propagating into both dawn and dusk sectors.

For the time moment of the first negative peak of SYM-H$\sim -110\,\mathrm{nT}$, reached at UT = 02:30 of DOY 148 after five hours of large southward IMF $B_z \sim -15\,\mathrm{nT}$ (though with rather low $V \leq 400\,\mathrm{km/s}$ and $P_d \sim 2\,\mathrm{nPa}$), the modeling reveals quite a different magnetic configuration (panel 4) with a deep field depression, which tightly envelops Earth over the entire 360° range of longitude and has a very strong dawn-dusk asymmetry. The peak $\Delta B \sim -180\,\mathrm{nT}$ is reached in the innermost post-dusk magnetosphere at $X_{\mathrm{SM}} = -1.7$ and $Y_{\mathrm{SM}} = 2.7\,R_E$. In this event, disturbance fields with comparable depressions between -150 and -200 nT were observed by Van Allen A in the same MLT sector and in the same range of the radial distance 2.6–3.0R_E, but about one hour earlier, between 01:00 and 01:30 UT. After that time period, the spacecraft moved well below the GSM equatorial plane, which precludes the direct comparison of its data with the equatorial plot 4 (UT = 02:30).

The lowest peak of SYM-H=–140nT was reached five more hours later at 07:30 of DOY 148 (panel 5 in Fig.5). By that time, the negative IMF B_z subsided from -18 to -13 nT, which is the most likely explanation why the equatorial depression has somewhat shrunk in size and decreased in magnitude, in spite of the significantly lower SYM-H. Also, note the difference in the midnight equatorial field: the intensity of the compression "island" at $X \sim -12\,R_E$ is clearly well correlated with the depression magnitude at $R \leq 6\,R_E$, which manifests the outward relocation of the magnetic flux from the innermost magnetosphere into the near magnetotail during the storm main phase. This effect is further confirmed in the next panel 6, corresponding to the early recovery phase: the magnetic field is now depressed in almost the entire equatorial magnetosphere on the nightside, even though the inner depression at $R \sim 4-5\,R_E$ is now much shallower. In the discussion below, we will address these effects from the field line mapping viewpoint.

The next Figure 6 presents a similar sequence of equatorial plots for the same time moments 1–6 of the same storm, showing distributions of the electric current volume density, calculated as the curl of the model $\mathbf{B}$. The color coding spans the interval from 0 to 12 nA/m^2 and displays the magnitude $|\mathbf{j}|$ of the current, while its direction patterns are shown by arrows of equal length.

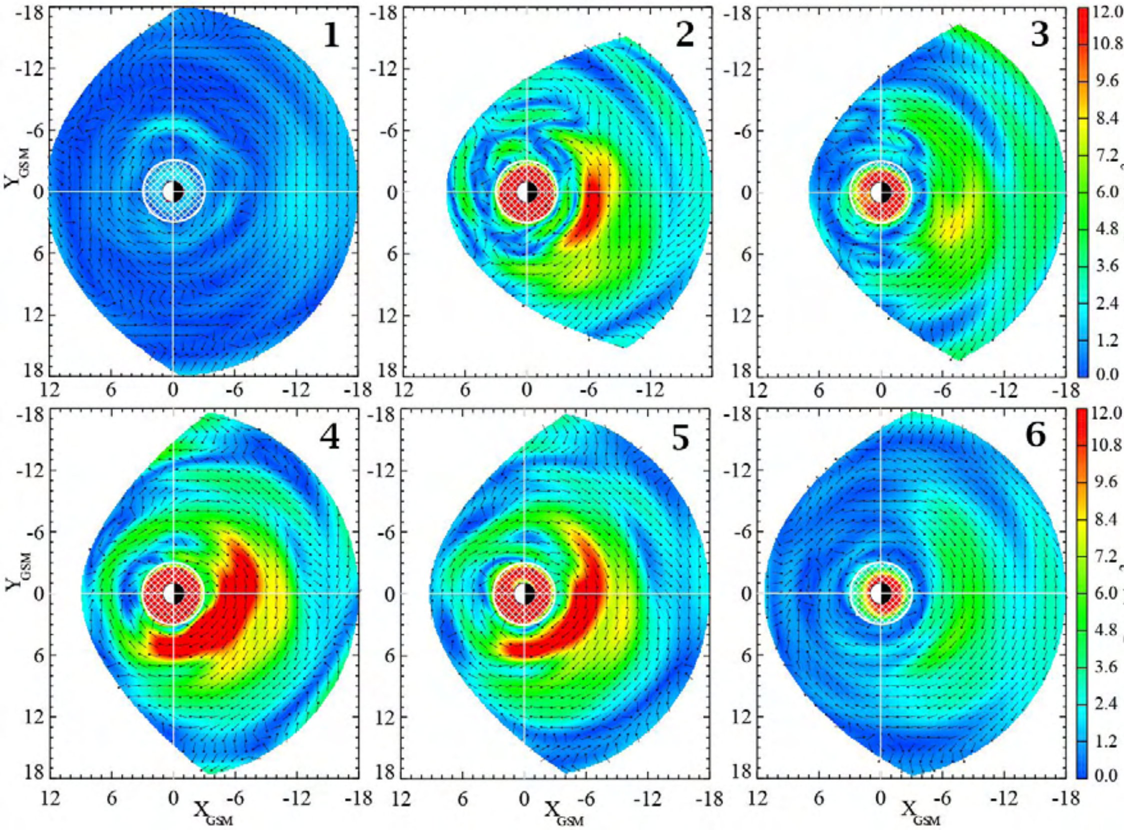

Figure 6. Equatorial plots of the volume density of the model electric current **j**, for the same time moments of the geomagnetic storm of 05/27–05/31, 2017. The color coding illustrates the current magnitude and the arrows of equal length indicate the direction of the current flow vectors. As in Fig.5, the cross-hatching in the area $R \leq 3\,R_E$ indicates the region where the model is invalid.

In contrast to the modular models with only a few custom-tailored smooth current systems, the high-resolution models include a large number of basis functions. This inevitably results in a somewhat structured magnetic field, whose degree of unevenness depends on the spatial non-uniformity of underlying data. Such structures are partially smoothed out in the mapping of field lines, since their tracing is effectively equivalent to the spatial integration (see Section 6.1 in Tsyganenko, 2013, for a more detailed discussion of the B-field mapping). By contrast, taking the curl of the model **B** magnifies the structures, which is why the electric current patterns usually look significantly bumpier than those for the magnetic field. Nevertheless, the plots in Figure 6 appear relatively regular and display an ordered westward current over the entire modeling region. The narrow azimuthal streaks with locally reduced or enhanced current density are most likely due to sharp gradients in the data coverage by individual missions.

An outstanding feature of the **j** patterns that repeatedly shows up for various combinations of the NN search parameters, is a transient radially narrow local surge of the current in the midnight sector at $R \sim 6-7R_E$, emerging during the storm SC (panel 2). As the solar wind pressure subsides three hours later (panel 3), the local current peak largely disappears, indicating its possible origin as the pressure wave induced by the CME shock passage and propagating toward the near-tail current sheet via the lobes. Later on, by the time of SYM-H first peak (panel 4, UT = 02:30, DOY 148), one sees a fully developed crescent-shaped area of strongly enhanced current in the post-dusk/pre-midnight sector, peaking at $R \sim 5-6R_E$. Five more hours later (panel 5; UT = 07:30) the current shrinks in space and decreases in magnitude, but remains localized in the same area. In the middle of the recovery phase (panel 6) the current further falls down, its distribution becomes much smoother and almost symmetric about the midnight meridian. In this case, one can also see the innermost eastward current at $R \sim 3-4R_E$. This is a fundamental feature, theoretically predicted still at the dawn of space era from the requirement of stress balance between plasma and magnetic field at the inner boundary of the RC (e.g., Akasofu et al., 1961). In the present model, the eastward current falls on the inner boundary of the modeling region and, hence, can hardly be analyzed nor quantified with a proper accuracy. It may be a subject of a future study, focused on the innermost magnetosphere and based on closer-range data and grid.

7.2. Reconstruction based on the ground disturbance indices (variant V2)

The ground-based indices have an advantage of reflecting the actual, rather than expected, state of the magnetosphere. Their disadvantage is the local nature of the information sources which, being tied to the Earth's surface, inevitably provide only a remote, spatially integrated and time-delayed monitoring of the complex magnetospheric processes, distributed over a vast domain of geospace.

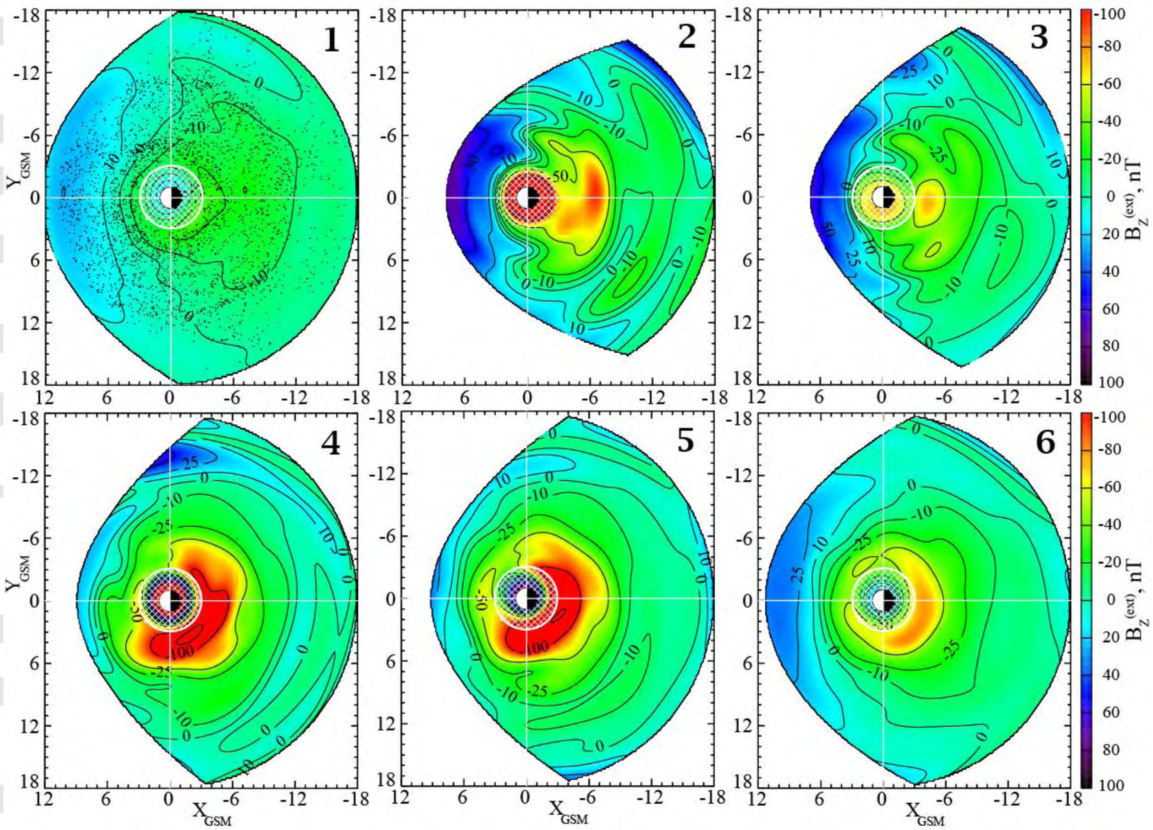

Figure 7. The CBF model $B_Z^{(\mathrm{ext})}$ plots, similar in format to those in Figure 5, but derived from the NN subsets compiled using $\langle$SYM-H$_c|$, $\langle$ASY-H$|$, $\langle d$SYM-H$_c/dt|$, $\langle d$ASY-H$/dt|$, and $\langle$AL$|$ as the data selection parameters.

It is nevertheless all the more interesting to test the ground indices as potential data mining source. Recent works (Sitnov et al., 2018 and refs. therein) provided a sound evidence in favor of that approach. Figure 7 shows a sequence of equatorial $B_Z^{(\mathrm{ext})}$ plots for the same 6 time moments of the storm of May 27–29, 2017, indicated in Figure 4. The plots are similar in format to those in Figure 5 but, as already said, instead of the upstream interplanetary data, are based on the NN data subsets selected using the ground indices listed above in Section 6.2.

As naturally expected, the diagrams in Fig.7 are not fully identical to those in Fig.5; the most interesting, however, is their surprising resemblance, in spite of the completely different data selection criteria. In particular, one again sees the quick buildup of the localized midnight depression (panel 2), coincident with the ram pressure pulse. Its following relaxation (panel 3) due to the solar wind pressure decrease is significantly more pronounced than that in the V1 case (Figure 5, panel 3).

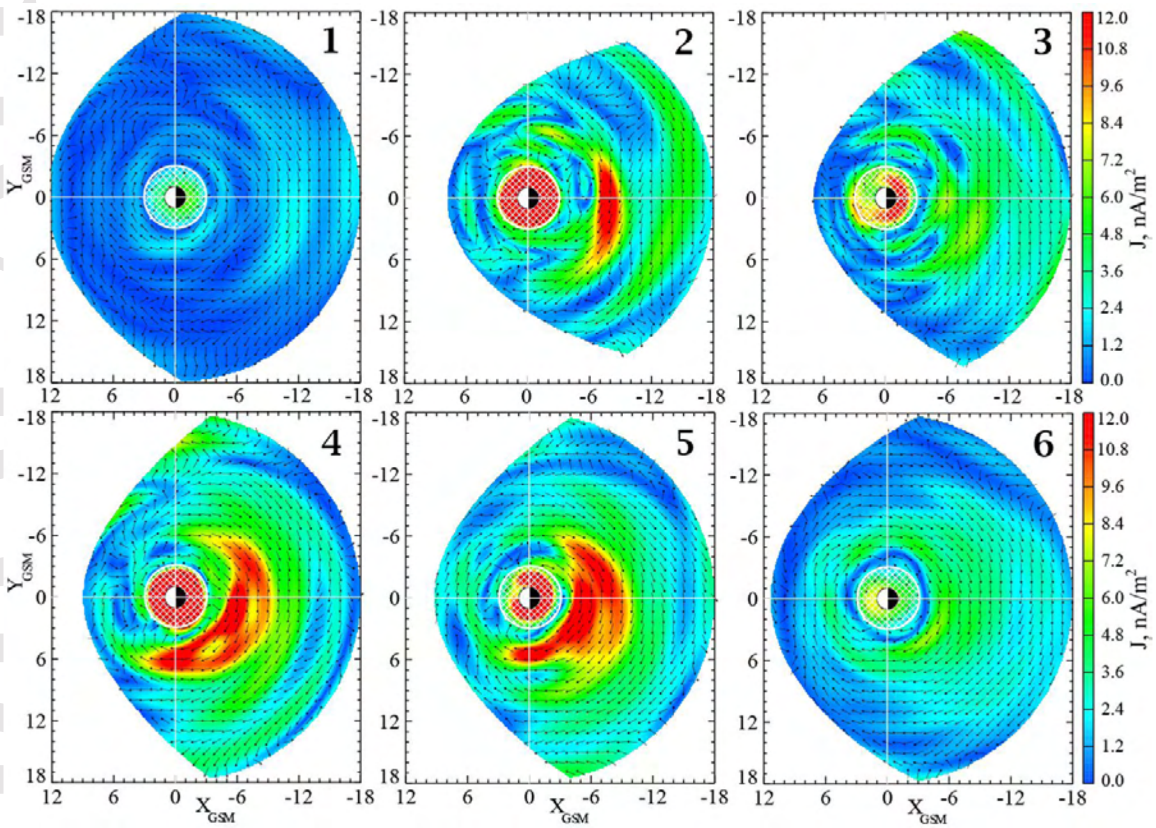

Figure 8. Equatorial plots of the electric current **j** analogous to Figure 6, but corresponding to the model field derived on the basis of the V2 NN subsets, selected using only ground-based indices.

The largest depression intensity ($-160\,$nT) at the time of first SYM-H peak (panel 5) is not as strong as in the V1 reconstruction ($-180\,$nT), and it is also shifted by $\sim$0.5 MLT hours closer to midnight (at 121° of SM longitude, against 114° in V1).

Figure 8 shows the corresponding equatorial distributions of the electric current, analogous in format to Figure 6 above, but based on completely different NN selection criteria that use only the ground disturbance indices. Again, in spite of the fundamental difference between the subset generation procedures, the plots reveal the same sequence of patterns, grossly similar to that in Figure 6: (1) a regular and relatively weak westward current before the storm, (2) a narrow radially localized area of strong current at $R \sim 7-8\,R_E$ in the midnight sector, concurrent with the solar wind pressure pulse and (3) almost disappearing as P_d weakens, (4) a crescent-shaped region of strong westward current, centered at $R \sim 6-7\,R_E$ and spanning a wide range of longitudes between postnoon and pre-dawn hours of MLT, (5) a similar pattern, but with some weakening at dusk and a wider radial extent near midnight. In the last panel (6), one sees a very regular **j** distrubution with a clear almost circular gap at $R \sim 3-4\,R_E$ between the outer westward and inner eastward currents.

In the end of this section, we provide Table 1 which lists some statistical characteristics of the models and NN data subsets. Its horizontal lines correspond to the six time moments of the storm, marked in Fig.4 and discussed above in regard to the equatorial plots of the magnetic field and electric currents (Figures 5–8).

Table 1. Statistics of the NN subsets and models for six time moments during two first days of the storm of May 27–29, 2017 (see Fig.4). The displayed quantities are the r.m.s. values in nT of the observed external field $\langle|\mathbf{B}|\rangle$, weighted residual $Q = \langle|\mathbf{B}-\mathbf{B}_{\mathrm{mod}}|^2\rangle^{1/2}$, ratio $Q/\langle|\mathbf{B}|\rangle$ in percent (parenthesized), and the correlation coefficients between the observed and modeled field components. The results are shown for the variants V1 (left) and V2 (right).

##		Variant 1					Variant 2				
	DOY/UT	$\langle\|\mathbf{B}\|\rangle$	Q, $Q/\langle\|\mathbf{B}\|\rangle$	R_x	R_y	R_z	$\langle\|\mathbf{B}\|\rangle$	Q, $Q/\langle\|\mathbf{B}\|\rangle$	R_x	R_y	R_z
1	147/12:00	15.87	5.71 (36%)	0.94	0.85	0.94	16.90	6.53 (39%)	0.91	0.81	0.94
2	147/19:00	39.07	15.71 (40%)	0.90	0.82	0.92	25.88	9.34 (36%)	0.91	0.85	0.94
3	147/22:00	42.57	19.40 (46%)	0.89	0.80	0.88	25.45	11.05 (43%)	0.90	0.80	0.91
4	148/02:30	65.72	24.75 (38%)	0.91	0.87	0.92	63.68	26.81 (42%)	0.90	0.85	0.90
5	148/07:30	59.53	22.66 (38%)	0.89	0.86	0.92	62.62	22.98 (37%)	0.90	0.86	0.93
6	149/00:00	33.20	13.28 (40%)	0.90	0.82	0.92	39.77	13.13 (33%)	0.92	0.87	0.94

Each line includes r.m.s. magnitudes of the external (i.e., with the IGRF subtracted) observed field $\langle|\mathbf{B}_{\mathrm{obs}}|\rangle$, along with the corresponding residual r.m.s. deviation of the model field from the data $Q = \langle|\mathbf{B}_{\rm obs} - \mathbf{B}_{\rm mod}|^2\rangle^{1/2}$, accompanied with a figure of merit $Q/\langle|\mathbf{B}_{\rm obs}|\rangle$ in parentheses. Additionally, three correlation coefficients are shown for each of three magnetic field components. Left and right parts of the table correspond to V1 and V2 variants, and we draw the reader's attention to a significant difference in the respective values of $\langle|\mathbf{B}_{\rm obs}|\rangle$, especially for the time moments 2 and 3 (storm SC and IMF southward excursion). In spite of that, the obtained figures of merit and the correlations are not that different between the V1 and V2, in line with the already mentioned similarity between the plots in Figures 5 vs 7 and 6 vs 8.

7.3. Validation

A standard procedure in the development of a new model is to test its performance with respect to reproducing independent data, not used in the model's generation. In the present case, a suitable source of such data taken inside the modeling domain is the data by geostationary GOES-13 and -15 satellites. We used 1-minute averaged data from the online source https://satdat.ngdc.noaa.gov/sem/goes/data/avg/2017/05/goes13/csv/ and https://satdat.ngdc.noaa.gov/sem/goes/data/avg/2017/05/goes15/csv/ and compiled two files, in which the observed values of the magnetic field GSM components were represented by the arithmetic mean of those measured by GOES inboard and outboard magnetometers. The model values were calculated for ten time moments during the storm, including the six ones marked in Figure 4 and analyzed in Figures 5–8. The results are shown in Figures 9 and 10 below, corresponding to GOES-13 and -15 data, respectively. The red and blue circles indicate values returned by the V1 and V2 model versions, respectively.

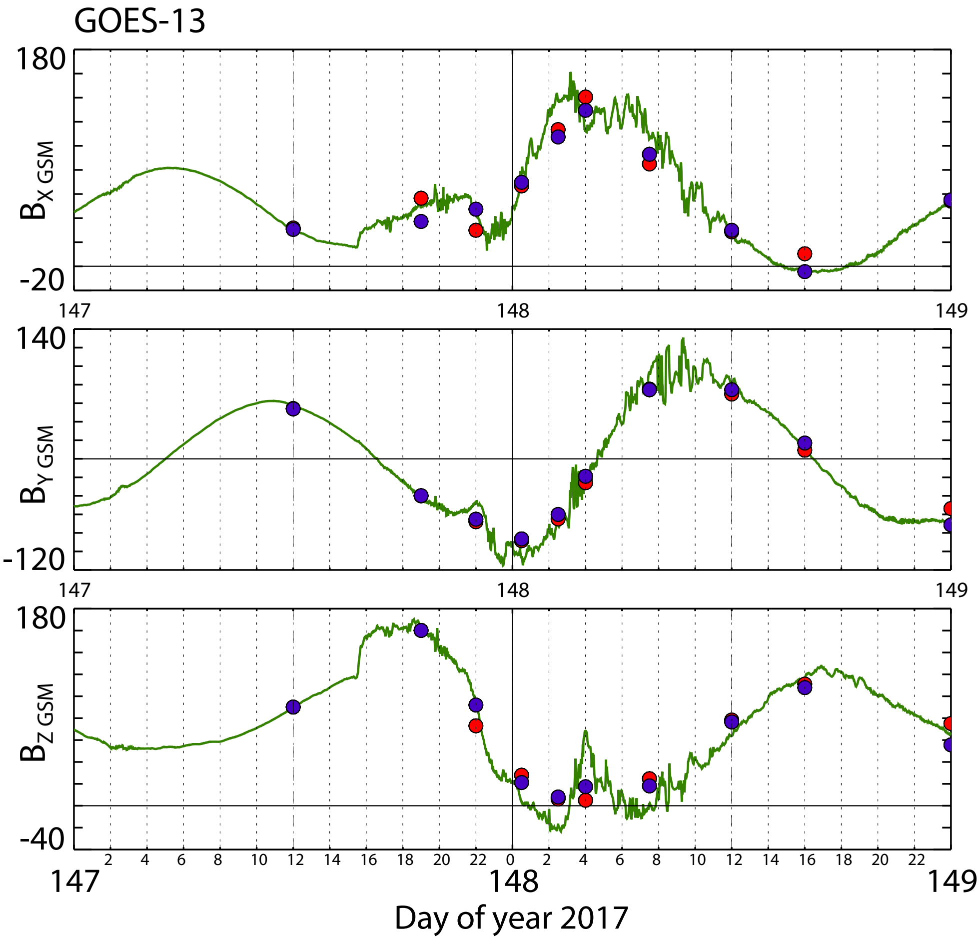

Figure 9. Compares the magnetic field variation observed by GOES-13 magnetometers (green solid curve) with the output of the CBF model (circles) for 10 time moments during DOY 147–148 of May 27–31, 2017, storm. Red and blue circles correspond to the models V1 and V2, respectively.

As expected, the agreement between the data and model is virtually perfect for the quiet-time NN data sample (the leftmost dot, DOY 147, 12:00) and for those corresponding to the early and late recovery phase (DOY 148, 12:00, 16:00, and 24:00). The largest deviations, also in line with expectations, are found during the main phase of the storm, from late UT hours of DOY 147 through the late morning of DOY 148. The most outstanding difference is seen in the GOES-15 plot for B_x (top panel in Fig.10) at the time of the largest negative peak of SYM-H (DOY 148, 07:30). The satellite was located at that moment in the premidnight sector (MLT≈22:20) at $X_{\mathrm{GSM}} = -5.72$, $Y_{\mathrm{GSM}} = 2.68$, $Z_{\mathrm{GSM}} = 1.97$, close to the inner edge of very intense tail current. Judging from the unusually low (even negative, down to $-20\,$nT) values of the total observed B_z (bottom panel) at that period, the spacecraft was inside a dynamic bubble-like magnetic configuration with a neutral point, which makes the obtained disagreement in B_x not surprising at all. One can also note that the largest difference between the V1 and V2 model vectors is found for the end of DOY 147, that is, at the start of the storm active phase. This is in line with the result, discussed in the end of Section 7.2 in relation to the large difference between the V1 and V2 data subsets, corresponding to that period.

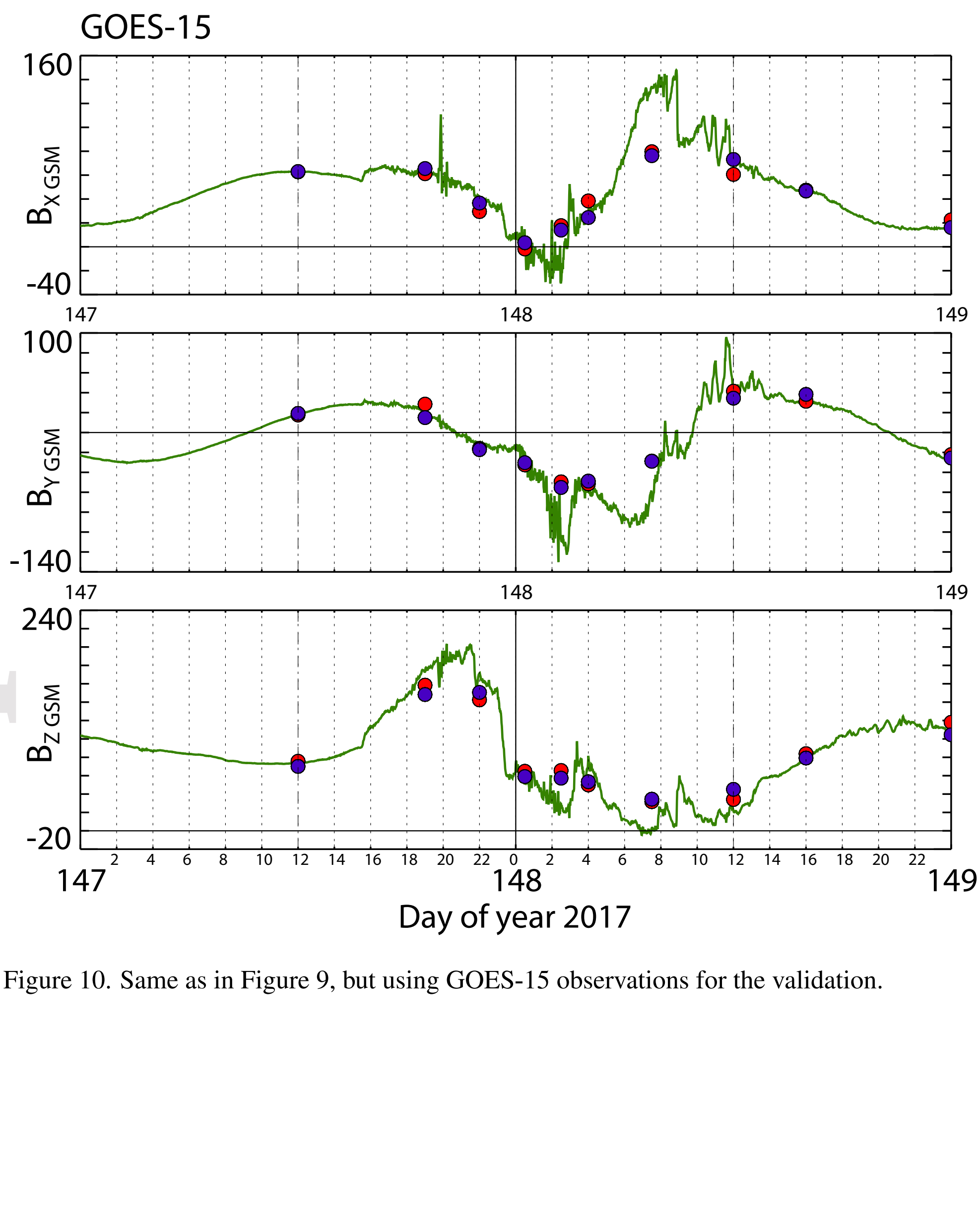

Figure 10. Same as in Figure 9, but using GOES-15 observations for the validation.

8. Discussion: storm-time effects in the field line mapping

The actual degree of the magnetospheric magnetic field distortion during geomagnetic storms has been discussed since long ago in relation to the observed equatorward expansion of the auroral oval, and the empirical modeling based on direct in-situ data plays here a central role. Notwithstanding a well-established classification of storms into a few typical classes, such as the CME-, or CIR-induced storms (e.g., Denton et al., 2006), it should be realized that, even inside the same class, there exists a great variety of different scenarios of external input and its previous history, which is unlikely to be accurately accomodated within a single universal dynamical model, such as, for example, the widely used TS05. In that respect, the NN data-mining approach offers a promising alternative, and its success depends mainly on three factors: (i) the abundance of historical data, (ii) an optimal combination of the NN selection parameters, and (iii) sufficient flexibility and resolution of the mathematical model.

In this section we address the storm-time reconfiguration of the magnetospheric magnetic field lines, as deduced from the CBF model for the same two sets of parameters employed in the NN data selection, for the same phases 1–6 of the May 2017 storm. In order to single out the effects of external input and/or of the magnetospheric disturbance, all field line configurations are reconstructed for zero dipole tilt, even though its actual values varied in the course of the disturbance. Figure 11 shows a sequence of configurations obtained for the V1 model, based on only interplanetary NN selection parameters. The most striking feature, already discussed above in regard to the equatorial depression ΔB (Figs.5 and 7), is the dramatic outward "discharge" of the magnetic flux (panel 2), coincident with the solar wind blow at the time of the storm SC. The initially quasi-dipolar quiet-time field lines that resided deep in the inner magnetosphere become now severely stretched: thus, the line with footpoint latitude $\Lambda_f = 66^\circ$ and quiet-time apex location around the geostationary orbit at $R \approx 6.6R_E$ is swept downtail as far as to $R \sim 16R_E$. At the next phase (panel 3) the pressure subsides, which results in a partial relaxation of the stretch.

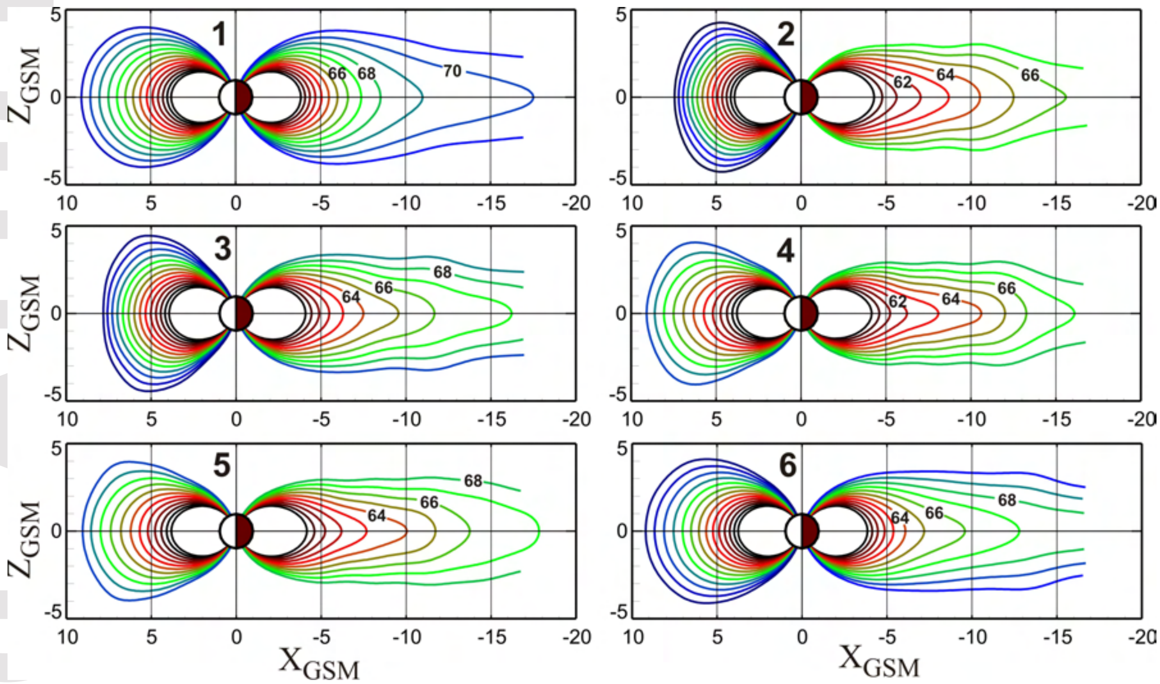

Figure 11. Displays a sequence of field line plots, corresponding to six moments during the storm of May 27–31, 2017, as shown in Figure 4 above. The field line footpoints lie in the noon-midnight meridional plane and follow at $\Delta\Lambda = 1°$ cadence of the SM latitude, starting from $\Lambda_f = 59°$ for the innermost line. To better visualize the changing field deformation, the lines are colored and labeled with footpoint latitudes. The modeling is based on the Variant 1 NN selection (using only interplanetary parameters)

At that time, however, effects of the southward excursion of IMF B_z come into play and result in the next round of the inner field stretching (panel 4, first negative peak of SYM-H), especially clearly visible in the shape of the line with $\Lambda_f = 64°$. By the time of the second Sym-H peak (panel 5) the configuration already gets slightly relaxed due to some decrease of the external driving, as prompted by the IMF B_z plot in Fig.4. Finally, at the recovery phase (panel 6) the configuration is much more relaxed, though still significantly more stretched than the initial quiet-time field in panel 1.

It is interesting to compare the above field line configurations with a similar set of plots, but obtained on the basis of NN subsets selected using only ground-based indices (Variant 2). The result is shown in Figure 12.

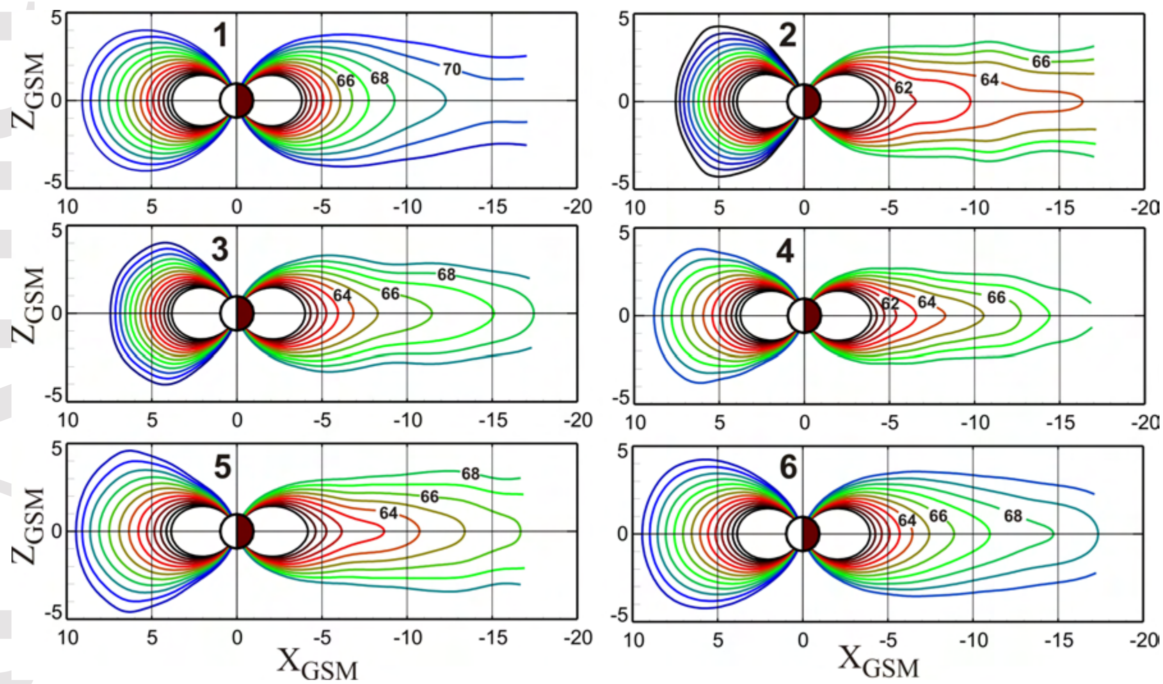

Figure 12. Field line plots, similar in format to Figure 11, but derived from the NN data subsets, generated using the ground-based indices (Variant 2, see Section 6.2 above).

As could already be expected from the comparison of equatorial ΔB distributions, the field line plots in Figs.11 and 12 are quite close, in spite of completely different methods of the NN subset formation. This is a convincing evidence in favor (i) of the general robustness of the magnetic field representation by the CBF expansions, (ii) of the reality of the storm effects derived from the modeling, and (iii) of a reasonable accuracy of their reconstruction. The largest difference between the corresponding plots in Figs.11 and 12 is in the panels 2 for the storm SC: the outward expulsion of the magnetic flux in the latter case is so dramatic that the field line with $\Lambda_f = 64^\circ$ extends to $X \approx -16.5\,R_E$, while the same field line in Figure 11 stretches no further than to $X \approx -11\,R_E$. Such a difference, as well as the somewhat wavy shape of the lines (also reflected in a certain irregularity of the equatorial plots of ΔB in panels 2 of Figures 5 and 7) is a result of the relatively short duration of the storm sudden commencements and, hence, relative rarity of such magnetospheric states. As a consequence, the formation of a corresponding NN subset with sufficiently large number of data points requires a relatively large value of the hypersphere threshold radius G_c. This unavoidably brings in data from significantly different regions of the parametric space with low weight factors, which increases the solution instabilities, rooted in the overfitting phenomenon. At the same time and for the same reason, one may expect that the actual field deformations during a specific event may be even more drastic than those derived from NN subsets that inevitably contain data corresonding to somewhat different magnetospheric states.

The impact on the magnetosphere of external pressure pulses and step-like increases has been extensively discussed in the past literature. Boudouridis et al. (2003) studied the SC effects upon the size of the polar cap and auroral oval size. In a more recent work, Li and Wang (2018) presented results of IMAGE observations of the auroral oval reaction to the external pressure changes. Both cited works, however, were based solely on low-altitude data of polar orbiting spacecraft, focused mainly on the open flux area, and provided no insight whatsoever regarding the distant field configuration, nor its response to P_d jumps. In the present paper, by contrast, we addressed the lower-latitude region, corresponding to the auroral/subauroral zone permeated by closed field lines, and reconstructed the magnetic configurations with unprecedented spatial resolution, not available in any of the previous empirical models.

9. Summary and outlook

The principal goal of this work was to synergistically combine a new flexible CBF model of the magnetic field with the advanced data mining method and, on that basis, demonstrate the high potential of this approach by reconstructing a sequence of magnetospheric configurations during the entire cycle of an intense geomagnetic storm. The model was designed as a generalization of the previously developed RBF representation, which allowed us to optimize the grid by taking into account the actual hierarchy of the magnetospheric scale sizes. Before fitting to actual observations, the new mathematical framework was tested on artificial data, to verify its ability to reproduce storm-time reconfigurations of the real field. The existing archive of historical magnetospheric data was extended to nearly twice its previous size by adding new data taken in the last years; in addition, a large amount of previously unused data has been included in the pool. The new model was fitted to NN subsets, generated by using various combinations of normalized interplanetary parameters, indices of ground geomagnetic acivity, and their temporal trends. In spite of the fundamental difference between the search parameters based on data of upstream solar wind monitors in the first case, and on only ground activity indices in the second, the modeling revealed a surprising similarity in the results. The most interesting finding is the dramatic tailward expulsion, or discharge, of the magnetic flux from the inner magnetosphere at the time of storm sudden commencement, manifested in the enormous stretching of the nightside field lines. A plausible interpretation of this effect can be a sudden compression of the tail lobes, initiating a local transient surge of electric current in the innermost tail current sheet. The following 12-hour period of continuous external driving of the magnetosphere results in the development of an intense symmetric and partial ring currents, accompanied by the formation of a deep magnetic depression in the inner magnetosphere, tightly enveloping Earth over the entire 360° longitude range and strongly asymmetric in the dawn-dusk direction. Upon the relaxation of the external driving, the model magnetosphere recovers to a nearly symmetric post-storm configuration.

In conclusion, some comments are in place. The advantage of the CBF formalism is its ability to locally represent the field in a specified region of interest with a variable resolution. As an attractive challenge for future research, one can envision a dynamical modeling of the nightside sector, aimed at the reconstruction of the substorm currents on the basis of in-situ satellite data. In such studies, the problem of accurate selection of the NN subsets comes to the foreground. As already noted, that problem boils down to the optimal choice of the selection parameters. In regard to the ground-based data, this calls for new advanced indices, providing maximum information on the dynamics, intensity, and localization of the substorm current wedge. A promising candidate could be the recently introduced MPB index (Chu et al., 2015; McPherron and Chu, 2018). In respect to the interplanetary data, a major hurdle is the large distance between Earth and upstream monitors which are often located too far away from the Sun-Earth line. As shown by Vohmyanin et al. (2019), this results in a significant percentage of poor quality OMNI data. A possible remedy can be to use the polar cap (PC) index (Troshichev, 2017, and refs. therein) as a partial substitute for the L1 data. It has been shown that the ground-based data on the polar cap magnetic variations can serve a reliable and continuous source of information on the solar wind electric field. Finally, as already noted above, the most serious challenge in the modeling of the storm-time magnetosphere is the relatively low density of NN data in the vicinity of most interesting time moments on the phase space trajectory. All the above issues offer a very wide agenda for future research.

Acknowledgments. It is a pleasure to acknowledge the teams and PIs of all experiments whose data contributed to this study. In particular, Geotail MGF data were provided by the PIs, S.Kokubun (STEL), and T. Nagai (Tokyo Institute of Technology, Japan) via the NSSDC CDAWEB facility at: https://cdaweb.gsfc.nasa.gov/cdaweb/index.html. Cluster, Van Allen Space Probes, and MMS magnetometer/ephemeris data were obtained from the NSSDC CDAWEB online facility at: https://cdaweb.gsfc.nasa.gov/cdaweb/index.html, originally made available by the PIs: A. Balogh and M. Tatrallyay (Cluster data), C. Kletzing (Van Allen Probes data), J. Burch, C. Russell, and W. Magnus (MMS data). GOES-13 and -15 data (PI H. J. Singer) were obtained from NOAA NCEI online data source (https://satdat.ngdc.noaa.gov/sem/goes/data/avg/). THEMIS team (V. Angelopoulos, K.-H. Glassmeier, U. Auster, and W. Baumjohann) is acknowledged for the use of THEMIS FGM data at: http://themis.igpp.ucla.edu. The data of Polar MGF experiment were made available by the UCLA Polar team led by PI C. Russell (currently at: https://cdaweb.gsfc.nasa.gov/cdaweb/index.html). High-resolution OMNI interplanetary data were obtained from the SPDF OMNIWEB at: https://omniweb.gsfc.nasa.gov (R. McGuire, N. Papitashvili).

All the data sets related to this paper's results are archived on Zenodo (https://doi.org/10.5281/zenodo.4036006).

This work has been funded by the Russian Foundation for Basic Research (RFBR) grant 20-05-00218.

References

Akasofu, S.-I., & Chapman, S. (1961). The ring current, geomagnetic disturbance, and the Van Allen radiation belts. *J. Geophys. Res., 66*, 1321–1350.

Andreeva, V. A., & Tsyganenko, N. A. (2016). Reconstructing the magnetosphere from data using radial basis functions. *J. Geophys. Res. Space Physics, 121*, 2249–2263, https://doi.org/10.1002/2015JA022242.

Andreeva, V. A., & Tsyganenko, N. A. (2018). Empirical modeling of the quiet and storm time geosynchronous magnetic field. *Space Weather, 16*, 16–36. https://doi.org/10.1002/2017SW001684.

Angelopoulos, V. (2008). The THEMIS mission. *Space Sci. Rev., 141,* 5–34.

Apatenkov, S.V., Sergeev, V. A., Amm, O., Baumjohann, W., Nakamura, R., Runov, A., et al. (2008). Conjugate observation of sharp dynamical boundary in the inner magnetosphere by Cluster and DMSP spacecraft and ground network. *Ann. Geophys., 26*, 2771–2780.

Auster, H. U., Glassmeier, K. H., Magnes, W., Aydogar, O., Baumjohann, W., Constantinescu, D., et al. (2008). The THEMIS fluxgate magnetometer. *Space Sci. Rev., 141*, 235–264.

Balogh, A., Dunlop, M. W., Cowley, S. W. H., Southwood, D. J., Thomlinson. J. G., Glassmeier, K. H., et al. (1997). The Cluster magnetic field investigation. *Space Sci. Rev., 79,* 65–92.

Boudouridis, A., Zesta, E., Lyons, L. R., Anderson, P. C., & Lummerzheim, D. (2003), Effect of solar wind pressure pulses on the size and strength of the auroral oval, *J. Geophys. Res., 108*(A4), 8012, https://doi.org/10.1029/2002JA009373.

Boynton, R. J., Balikhin, M. A., Billings, S. A., Wei, H. L., & N. Ganushkina (2011). Using the NARMAX OLS-ERR algorithm to obtain the most influential coupling functions that affect the evolution of the magnetosphere. *J. Geophys. Res., 116,* A05218, https://doi.org/10.1029/2010JA015505.

Chen, W., Wang, X., Tsyganenko, N. A., Andreeva, V. A., & Semenov, V. S. (2019). Reconstruction of local magnetic structures by a modified radial basis function method. *J. Geophys. Res. Space Physics, 124*, https://doi.org/10.1029/2019JA027078.

Chu, X., McPherron, R. L., Hsu, T.-S., & V. Angelopoulos (2015), Solar cycle dependence of substorm occurrence and duration: Implications for onset, *J. Geophys. Res. Space Physics*, 120, 2808–2818, https://doi.org/10.1002/2015JA021104.

Denton, M. H., Borovsky, J. E., Skoug, R. M., Thomsen, M. F., Lavraud, B., Henderson, M. G., et al. (2006). Geomagnetic storms driven by ICME- and CIR-dominated solar wind. *J. Geophys. Res. Space Physics, 111*, A07S07, https://doi.org/10.1029/2005JA011436

Iyemori, T. (1990). Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations. *J. Geomag. Geoelectr., 42*, 1249–1265.

Li, L. Y., & Wang, Z. Q. (2018). The effects of solar wind dynamic pressure changes on the substorm auroras and energetic electron injections on 24 August 2005, *J. Geophys. Res. Space Physics, 123*, 385–399, https://doi.org/10.1002/2017JA024628

Lin, R. L., Zhang, X. X., Liu, S. Q., Wang, Y. L., & Gong, J. C. (2010). A three-dimensional asymmetric magnetopause model. *J. Geophys. Res. Space Physics*, *115*, A04207, https://doi.org/10.1029/2009JA014235.

Liu, J., Angelopoulos, V., Runov, A., & Zhou, X.-Z. (2013). On the current sheets surrounding dipolarizing flux bundles in the magnetotail: The case for wedgelets. *J. Geophys. Res. Space Physics, 118*, 2000–2020, https://doi.org/10.1002/jgra.50092.

Kokubun, S., Yamamoto, T., Acuña, M. H., Hayashi, K., Shiokawa, K., & Kawano, H. (1994). The Geotail magnetic field experiment. *J. Geomagn. Geoelectr., 46*, 7–21.

Kletzing, C. A., Kurth, W. S., Acuña, M. H., MacDowall, R. J., Torbert, R. B., Averkamp, T., et al. (2013). The electric and magnetic field instrument suite and integrated science (EMFISIS) on RBSP. *Space Sci. Rev.,* 179:127–181, https://doi.org/10.1007/s11214-013-9993-6.

Kurihara, Y. (1965). Numerical integration of the primitive equations on a spherical grid. *Mon. Weather Rev., 93*(7), 399–415.

McPherron, R. L., & Chu, X. (2017). The mid-latitude positive bay and the MPB index of substorm activity. *Space Sci. Rev., 206*, 91–122, https://doi.org/10.1007/s11214-016-0316-6.

Mead, G. D., & Fairfield, D. H. (1975). A quantitative magnetospheric model derived from spacecraft magnetometer data. *J. Geophys. Res.*, *80*, 523–534.

Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. J. (2007). A nearly universal solar wind – magnetosphere coupling function inferred from 10 magnetospheric state variables. *J. Geophys. Res. Space Physics, 112,* A01206, https://doi.org/10.1029/2006JA012015.

Nishida, A. (1994). The Geotail mission. *Geophys. Res. Lett., 21,* 2871–2873, https://doi.org/10.1029/94GL01223.

Russell, C. T., Snare R. C., Means, J. D., Pierce, D., Dearborn, D., Larson, M., et al. (1995). The GGS/Polar magnetic fields investigation. *Space Sci. Rev., 71,* 563–582.

Sitnov, M. I., Tsyganenko N. A., Ukhorskiy, A. Y., & Brandt, P. C. (2008). Dynamical data-based modeling of the stormtime geomagnetic field with enhanced spatial resolution. *J. Geophys. Res. Space Physics, 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.

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

Sitnov, M. I., Stephens, G. K., Tsyganenko, N. A., Miyashita, Y., Merkin, V. G., Motoba, T., et al. (2019). Signatures of nonideal plasma evolution during substorms obtained by mining multimission magnetometer data. *J. Geophys. Res. Space Physics, 124,* https://doi.org/10.1029/2019JA027037.

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. *J. Geophys. Res. Space Physics*, *124*, 1085–1110. https://doi.org/10.1029/2018JA025843.

Troshichev, O. A.(2017). *Polar Cap Magnetic Activity (PC index) and Space Weather Monitoring.* Editions Universitaires Europeennes, (pp. 140).

Tsyganenko, N. A., & Usmanov, A. V. (1982). Determination of the magnetospheric current system parameters and development of experimental geomagnetic field models based on data from IMP and HEOS satellites. *Planet. Space Sci., 30*, 985–998.

Tsyganenko, N. A., Singer, H. J., & Kasper, J. C. (2003). Storm-time distortion of the inner magnetosphere: How severe can it get? *J. Geophys. Res.*, 108(A5), 1209, https://doi.org/10.1029/2002JA009808.

Tsyganenko, N. A., & Sitnov, M. I. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *J. Geophys. Res. Space Phys., 110*(A3), A03208, https://doi.org/10.1029/2004JA010798.

Tsyganenko, N. A., & Sitnov, M. I. (2007). Magnetospheric configurations from a high-resolution data-based magnetic field model. *J. Geophys. Res. Space Physics, 112*, A06225, https://doi.org/10.1029/2007JA012260.

Tsyganenko, N. A. (1996). Effects of the solar wind conditions on the global magnetospheric configuration as deduced from data-based field models. Eur. Space Agency Spec. Publ., ESA SP-389, 181.

Tsyganenko, N. A. (2013). Data-based modelling of the Earth's dynamic magnetosphere: a review. *Ann. Geophys. 31,* 1745–1772, https://doi.org/10.5194/angeo-31-1745-2013.

Tsyganenko, N. A., & Andreeva, V. A. (2014). On the "bowl-shaped" deformation of planetary equatorial current sheets. *Geophys. Res. Lett., 41,* 1079–1084, https://doi.org/10.1002/2014GL059295.

Tsyganenko, N. A., & Andreeva, V. A. (2017). A hybrid approach to empirical magnetosphere modeling. *J. Geophys. Res. Space Physics*, *122*, 8198–8213. https://doi.org/10.1002/2017JA024359.

Tsyganenko, N. A. & Andreeva, V. A. (2020). Magnetospheric "penetration" of IMF By viewed through the lens of an empirical RBF modeling. *J. Geophys. Res. Space Physics, 125,* e2019JA027439 https://doi.org/10.1029/2019JA027439.

Verleysen M., & François D. (2005). The curse of dimensionality in data mining and time series prediction. In: Cabestany J., Prieto A., Sandoval F. (eds.) *Computational Intelligence and Bioinspired Systems.* IWANN 2005. Lecture Notes in Computer Science, vol. 3512. Springer, Berlin, Heidelberg.

Vokhmyanin, M. V., Stepanov, N. A., & Sergeev, V. A. (2019). On the evaluation of data quality in the OMNI interplanetary magnetic field database. *Space Weather, 17,* 476–486. https://doi.org/10.1029/2018SW002113.