

Geophysical Research Letters[®]



RESEARCH LETTER

10.1029/2021GL096275

Key Points:

- We use a machine learning technique called Echo State Network to reconstruct the solar wind parameters from geomagnetic activity indices
- Two different types of solar wind profiles are discussed for extreme magnetic storms, that is, sheath-driven and cloud-driven types
- Continuous hourly data of the reconstructed solar wind parameters are obtained since 1957

Supporting Information:

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

Correspondence to:

R. Kataoka,
kataoka.ryuho@nipr.ac.jp

Citation:

Kataoka, R., & Nakano, S. (2021). Reconstructing solar wind profiles associated with extreme magnetic storms: A machine learning approach. *Geophysical Research Letters*, 48, e2021GL096275. <https://doi.org/10.1029/2021GL096275>

Received 21 SEP 2021

Accepted 31 OCT 2021

Author Contributions:

Conceptualization: Ryuho Kataoka

Formal analysis: Ryuho Kataoka

Funding acquisition: Shin'ya Nakano

Investigation: Ryuho Kataoka, Shin'ya Nakano

Methodology: Ryuho Kataoka, Shin'ya Nakano

Project Administration: Ryuho Kataoka

Software: Ryuho Kataoka

Supervision: Ryuho Kataoka

Validation: Ryuho Kataoka

Visualization: Ryuho Kataoka

Writing – original draft: Ryuho Kataoka

Writing – review & editing: Shin'ya Nakano

© 2021. The Authors.

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

Reconstructing Solar Wind Profiles Associated With Extreme Magnetic Storms: A Machine Learning Approach

Ryuho Kataoka^{1,2,3}  and Shin'ya Nakano^{3,4,5} 

¹National Institute of Polar Research, Tachikawa, Japan, ²Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan, ³The Graduate University for Advanced Studies, SOKENDAI, Hayama, Japan, ⁴The Institute of Statistical Mathematics, Tachikawa, Japan, ⁵Center for Data Assimilation Research and Applications, Joint Support Center for Data Science Research, Tachikawa, Japan

Abstract The lack of data on solar wind have prevented a detailed understanding of extreme magnetic storms. To address this issue, we apply a machine learning technique in the form of an Echo State Network (ESN) to reconstruct solar wind data for several extreme magnetic storms for which little or no solar wind data were previously available. Multiple geomagnetic activity indices are used as the input data for the ESN, which produces a continuous time series of solar wind parameters as output. As a result, the solar wind parameters for the largest storm event in March 1989 are obtained, and the minimum Bz is estimated to be $-95 \text{ nT} \pm 10 \text{ nT}$. Two different types of solar wind profiles are discussed for the extreme magnetic storms, a sheath-driven profile and a magnetic cloud-driven profile. The results reported here will be highly useful as input data for future simulation studies modeling extreme magnetic storms.

Plain Language Summary The detailed cause of extreme magnetic storms are unknown because of the very limited solar wind data. A machine learning technique can help to reconstruct the lost solar wind data from the geomagnetic response. New findings from the reconstructed solar wind data includes the unexpectedly large magnetic field strength of the solar wind that caused the largest magnetic storms in space age. A statistical analysis using the reconstructed solar wind data also tells the estimation of one in 100 years amplitude of the solar wind magnetic field for the first time. The method proposed in this study is also useful to obtain the long-term continuous solar wind data since 1957, for the period even before the solar wind observation started.

1. Introduction

Identifying the solar wind profiles of extreme magnetic storms, as roughly defined by the peak Dst index $< -300 \text{ nT}$, is important for mitigating possible space weather hazards driven by coronal mass ejections (CMEs). The body of these CMEs is referred to as a “magnetic cloud”; the shock downstream region ahead of the magnetic cloud is called the “sheath.” The magnetic cloud and sheath region are the only interplanetary structures that can cause extreme magnetic storms (Kilpua et al., 2017). In fact, they are the drivers of large amplitude, geomagnetically induced currents that can destroy power grids (Boteler, 2019; Huttunen et al., 2008; Kataoka & Pulkkinen, 2008).

The time variation of the extreme solar wind parameters at 1 AU is not randomly determined in a slot machine manner; rather, it is determined by the physics of the interaction among CMEs and ambient solar wind (Kataoka et al., 2015). The theoretical examination of the “perfect CME” gives the solar wind parameters of 2,700 km/s, 20/cc, and 127 nT (Tsurutani and Lakhina, 2014), which are also used as input parameter for magnetohydrodynamic (MHD) simulation to investigate the super sudden commencement (Welling et al., 2021). In general, superposed epoch analysis is useful for identifying average solar wind profiles for CME-driven storms and others (e.g., Kataoka & Miyoshi, 2006). However, there are essential problems in attempting to identify the basic solar wind profiles of extreme storms. First, extreme magnetic storms are rare, which makes it difficult to develop an average profile based on multiple observations. Furthermore, extreme magnetic storms tend to occur concurrently with solar energetic particle events, which cause data gaps in the solar-wind observations. As a consequence, it is impossible to establish average solar wind profiles for extreme events using observational data alone.

The purpose of this paper is to present a novel method for reconstructing the set of solar wind parameters from a set of geomagnetic activity indices. The availability of relevant data is briefly discussed in Section 2. Section 3 explains the machine learning technique that will be used to fill the gaps in the data. An example and the metrics

used for evaluation are presented in Section 4, along with the results of a superposed epoch analysis of two different types of extreme magnetic storms. In Section 5, we discuss our findings and their implications. Concluding remarks are presented in Section 6.

2. Data Set

The solar wind parameters were obtained from the OMNI2 data set, which is available at <https://omniweb.gsfc.nasa.gov/ow.html>. OMNI2 gives the long-term time series of the solar wind parameters at the Earth's position. The time cadence is 1 hr, which was calibrated by time shifting and averaging from available spacecraft data. The available OMNI2 solar wind data covers the period from November 28, 1963, although there are frequent data gaps before 1994. In this study, we use the north-south component of the interplanetary magnetic field (B_z) in the geocentric solar magnetosphere (GSM) coordinate system, the number density of protons (N_p), and the solar wind speed (V_{sw}). These three parameters are the most important parameters driving the geomagnetic activities via exciting the magnetospheric and ionospheric current systems (Akasofu, 1981; Borovsky, 2008; Newell et al., 2007).

The most widely used indices of geomagnetic activity are AE, Kp, and Dst, each of which responds differently to the solar wind parameters because of their respective physical meanings: auroral electrojet (AE), subauroral disturbances (Kp), and ring current development (Dst). For example, the primary parameter affecting the Dst index is B_z , while for the AE and Kp indices the primary parameter is V_{sw} . It is expected that the slight difference in the geomagnetic indices includes information related to the solar wind parameters, which, we propose, can be extracted by the machine learning technique developed in this study (and described in detail below). There are many examples of attempts to reproduce the time series of AE, Kp, and/or Dst from solar wind parameters B_z , N_p , and V_{sw} using various methods, including machine learning techniques (Liemohn et al., 2018).

Geomagnetic activity indices have been available since before the space age. For example, values of the Kp index are available at a 3-hr time cadence for all past years since 1932, at GFZ Potsdam (Matzka et al., 2021). In this study, we converted the 3-hr Kp index to consecutive hourly values in the Kp*10 index, following the OMNI2 format, e.g., Kp = [0+, 3-, 1+] becomes Kp*10 = [0, 0, 0, 26.7, 26.7, 26.7, 13.3, 13.3, 13.3]. The Dst index is available at a 1-hr time cadence for the years 1957 through 2014 at WDC Kyoto. A provisional Dst index is available for three subsequent years, 2015, 2016, and 2017. The hourly data for the AE index are also available for the years 1957 through 2014 at WDC Kyoto. However, there are data gaps in the AE index for the period from January 1, 1976 to December 31, 1977 and the period from July 1, 1988 to December 31, 1989 (except for March 1989). A provisional AE index is available for the years 1990 through 2017. After 2018, only quick-look data are available for both the Dst and AE indices. Given the data availability described above, we have limited our analysis to the time period from January 1, 1957 to December 31, 2017.

3. Method of Analysis: Echo State Network

The Echo State Network (ESN) was developed by Jaeger (2001) as a type of recurrent neural network. The architecture has been investigated and used by many authors (Hart et al., 2020; Tanaka et al., 2019). ESN is now known as a novel method that works with a relatively small set of training data, which is especially suitable to the analysis of times series. The first application of ESN to space physics appears in Nakano and Kataoka (2021). In ESN, the input vector $\mathbf{u}$, reservoir state vector $\mathbf{x}$, and output vector $\mathbf{y}$ are defined by an N -point time series where $n = 1, 2, 3, \dots, N$.

$$\mathbf{u}(n) = \begin{pmatrix} u_1(n) \\ \vdots \\ u_{N_u}(n) \end{pmatrix}, \mathbf{x}(n) = \begin{pmatrix} x_1(n) \\ \vdots \\ x_{N_x}(n) \end{pmatrix}, \mathbf{y}(n) = \begin{pmatrix} y_1(n) \\ \vdots \\ y_{N_y}(n) \end{pmatrix} \quad (1)$$

The reservoir state vector $\mathbf{x}$ consists of a large number of nodes, which are updated in time with the input vector $\mathbf{u}$ and the previous state of the nodes as follows:

$$\mathbf{x}(n+1) = f(W^{\text{in}}\mathbf{u}(n+1) + W\mathbf{x}(n)) \quad (n = 0, 1, 2, \dots) \quad (2)$$

We use the hyperbolic tangent function as function f and fix the weight matrices W^n and W . To create the random and sparse node connections of W , we set the number of nodes, N_x , equal to 10^3 , where only 10% of the matrix elements are random values between -1.0 and 1.0 , and the remaining 90% are zero, that is, the network connection density = 0.1. We chose 0.95 as the spectral radius (maximum eigenvalue) of W in order to satisfy the echo state property guaranteeing the independence of the reservoir state from the initial values when the spectral radius is below unity (Jaeger, 2001).

Thus, there are several parameters to change the model structure. In the present problem, selecting different number of nodes and/or the connection density are not sensitive to change the structure, and the selection of spectral radius is the most sensitive. Figure S2 in Supporting Information S1 shows the dependence on the spectral radius from 0.5 to 1.0. In this paper, we select 0.95 as the best value of the spectral radius to minimize the errors of both test data and training data for Bz and Vsw. To minimize the error of Np, however, smaller spectral radius gives better scores for the test data. Such a separate optimization for each parameter can therefore be a future work to improve the model, although it is beyond the scope of this paper.

The output vector $\mathbf{y}$ is calculated using the linear combination of the output weight matrix and the reservoir state vector as follows:

$$\mathbf{y}(n+1) = W^{\text{out}} \mathbf{x}(n+1) \quad (n = 0, 1, 2, \dots), \quad (3)$$

where W^{out} is the output weight matrix. We train only the output weight matrix by the set of T -point time series of input vectors X and desired output vectors D :

$$X = [\mathbf{x}(1), \dots, \mathbf{x}(T)], D = [\mathbf{d}(1), \dots, \mathbf{d}(T)]. \quad (4)$$

The least squares method for minimizing the difference between the outputs $\mathbf{y}$ and $\mathbf{d}$ can be represented by a standard linear regression as follows:

$$W^{\text{out}} = DX^T (XX^T)^{-1}. \quad (5)$$

We use the AE, Kp, and Dst indices as the three-dimensional input vector ($N_x = 3$), and Bz, Np, and Vsw as the three-dimensional output vector ($N_y = 3$). In this study, we take the logarithms of the density Np and speed Vsw to avoid negative outputs. It is noteworthy here that both the density and speed follow log-normal distributions (Burlaga & Lazarus, 2000). Note, too, that it is the solar wind that drives geomagnetic activities and not the other way around. Considering such a cause-and-result relationship, the time stamp is reversed when we apply ESN, that is, all of the N -point time series are sorted by $n = N, N-1, N-2, \dots, 3, 2, 1$. Here, we assumed that OMNI2 data was perfectly time-shifted already to match the geomagnetic indices and solar wind drivers in time.

To prepare the training data set for X and D in Equation 4, we constructed a special time series by selecting and connecting the storm events, which was intended to ensure that the model was reasonable, especially during disturbed periods. In so doing, we first detected the storm days using the 1-day minimum amplitude of the Dst index below -100 nT, with the previous 3 days of Dst > -100 nT. We then discarded the storm days if there were any data gaps in the solar wind parameters during the 7-day time period spanning the storm day, from 3 days before to 3 days after. Finally, the time series of selected storm events were connected in time. The training data set was constructed from the 28 years of data from 1990 to 2017, a period during which the OMNI2 solar wind data has only infrequent data gaps. We identified a total of 74 storm events during the selected period (1990–2017) to create a time series of $74 \times 7 \times 24 = 12,432$ data points. The percentage of how much of storm events had complete data is $\sim 51\%$ as we discarded the other 70 data-gap storm events. The training data set and the event list of storms are provided in Data Set S1.

4. Results

4.1. Reconstruction of the Solar Wind Parameters in March 1989

The largest observed magnetic storm of minimum Dst = -589 nT occurred in March 1989. Figure 1 shows the corresponding geomagnetic activity indices and the solar wind parameters available in the OMNI2 data. As can

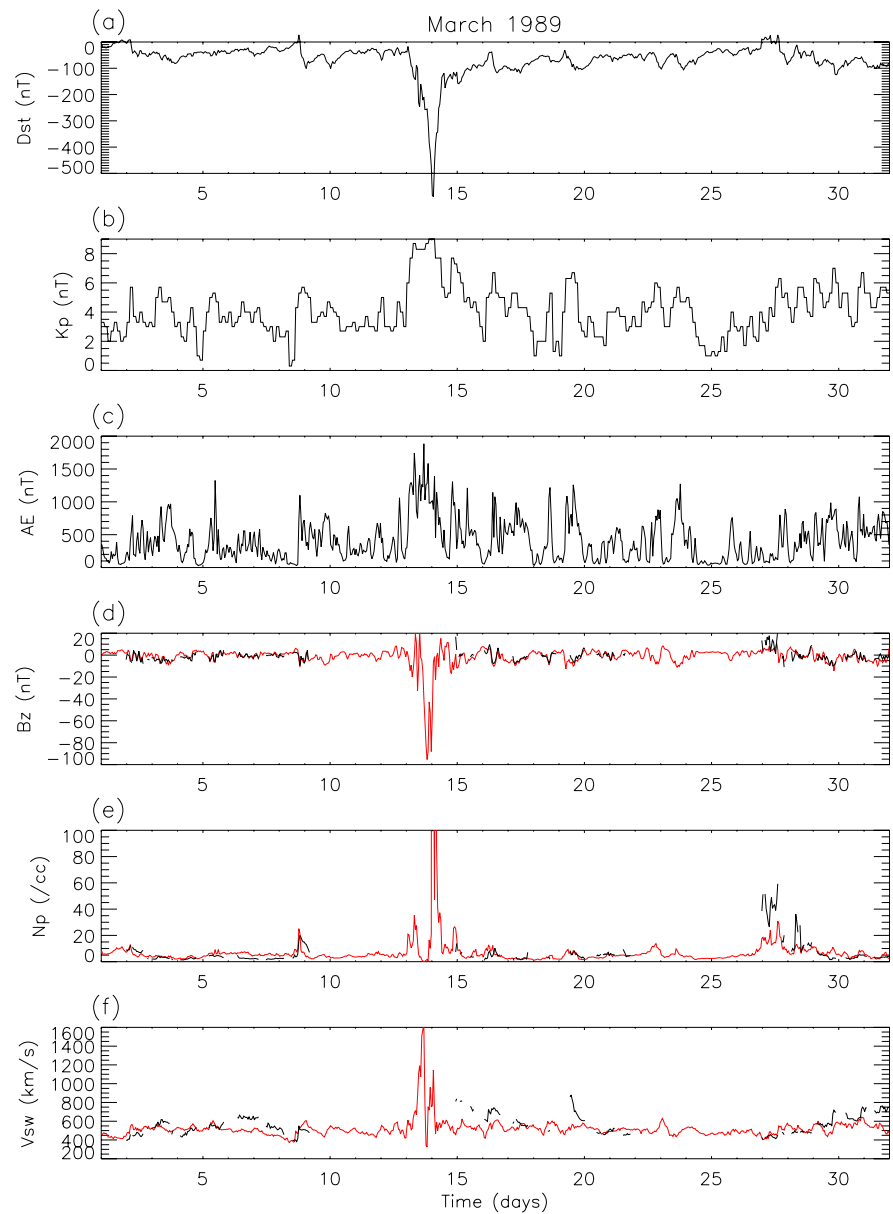


Figure 1. Example of geomagnetic indices and solar wind parameters for March 1989. The top three panels show the respective geomagnetic activity indices, (a) Dst, (b) Kp, and (c) AE. The bottom three panels show the solar wind data for (d) the north-south component of the interplanetary magnetic field Bz, (e) proton density Np, and (f) the solar wind speed Vsw. The black lines indicate observation data; the red lines indicate reconstructed data.

be seen here, there are large data gaps in the solar wind parameters, especially around the magnetic storm on March 13 and 14. The red lines show the reconstructed solar wind parameters using the ESN model. The most notable result is that the modeled minimum Bz reaches -95 nT during the magnetic storm. It is also noteworthy that the ESN model fails to predict the northward Bz (~ 10 nT was predicted to be ~ 5 nT on March 27) and underestimates density (~ 40 /cc was predicted to be ~ 20 /cc on March 27) and speed (>700 km/s was predicted to be ~ 600 km/s on March 19). Although these errors are not small, as described in the next subsection, the example illustrates the first step in reconstructing the entire time series for such large magnetic storms. Further examples of storm-time comparisons are shown in Figures S3-S6.

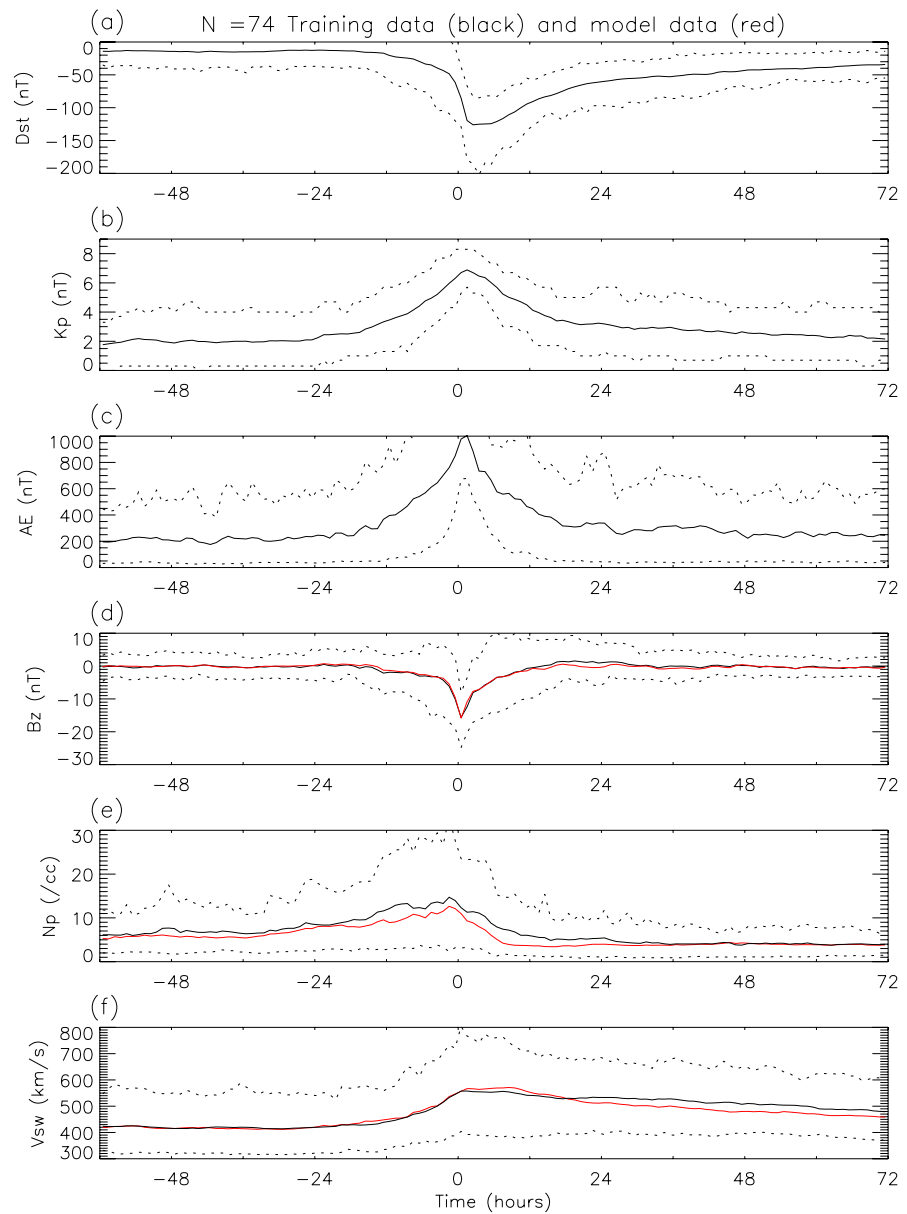


Figure 2. Superposed epoch analysis against the zero epoch of minimum Bz, using the training data set. The top three panels show the geomagnetic activity indices (a) Dst, (b) Kp, (c) AE. The bottom three panels show the solar wind data for (d) the north-south component of the interplanetary magnetic field Bz, (e) proton density Np, and (f) the solar wind speed Vsw. The dotted lines indicate the top 90% and bottom 10% data. The red lines are the superposed time profiles of the solar wind parameters reconstructed by ESN.

4.2. Superposed Epoch Analysis and Metrics

Figure 2 shows the results of our superposed epoch analysis against the zero epoch of minimum Bz. To train the ESN, we used the storm-time solar wind data without data gaps (a total of 74 storm events) for the time period from 1990 to 2017. The black solid curves show the training data set; the red curves show the reconstructed data using the ESN. It was found that the ESN reasonably reconstructed the average time profiles of the solar wind parameters by taking ensemble averages to reduce the errors. It should also be noted that the values of proton density tend to be underestimated during the disturbed time period.

To evaluate the errors in our analysis, the standard metrics were calculated using the storm-time data (a total of 70 storm events) for the time period from 1990 to 2017, in which the gap-less training data were excluded. The

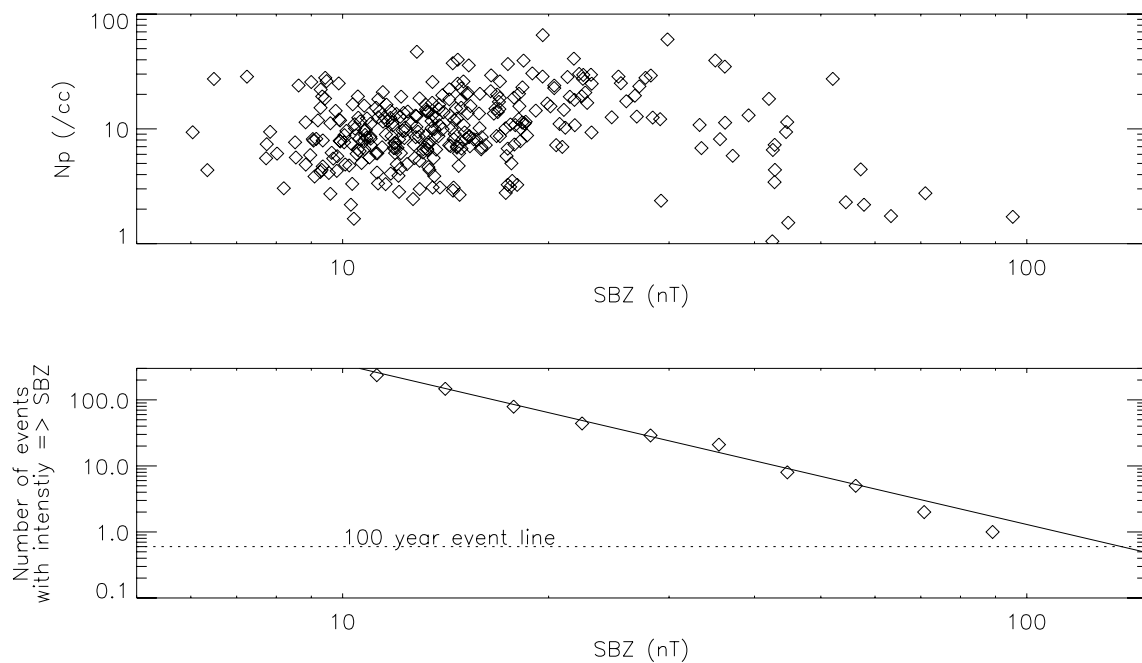


Figure 3. (top) Scatter plot of minimum Bz versus the density at the zero epoch for all storm events of $Dst < -100$ nT. (bottom) Cumulative distribution function of SBZ values. The solid line indicates a power-law fit for $SBZ > 30$ nT; the dotted line indicates a one-in-100-year event.

cross-correlation coefficients were 0.72, 0.42, and 0.61 for Bz, Np, and Vsw, respectively (as documented in Figure S1 in Supporting Information S1). The root mean square errors (RMSE) for Bz, Np, and Vsw were 4.0 nT, 7.2/cc, and 104 km/s, respectively. Normalizing the RMSE values by the respective standard deviations produced normalized root mean square errors (NRMSE) of 0.71, 1.0, and 0.79. The error is large for the large values, for example, ~ 10 nT for $|Bz| > 10$ nT events, as shown in Figure S2 in Supporting Information S1.

4.3. Occurrence Distribution of Extreme Events

To investigate the occurrence distribution of extreme events, the solar wind parameters of all of the storm events from 1957 to 2017 were reconstructed. The absolute values of the minimum Bz (SBZ) and the density at the minimum Bz timing were extracted for each storm event to create the scatter plot in Figure 3.

Figure 3 shows the cumulative distribution function (CDF) of the SBZ values. After extending the CDF curve by fitting the power-law distribution (Kataoka, 2013, 2020), the one-in-100-year-event SBZ is estimated to be 137 nT. It is noteworthy that such a large amplitude of $SBZ \sim 100$ nT was observed at the STEREO-A spacecraft during the July 2012 event (Russell et al., 2013) in which an extreme CME missed the Earth by only a few days.

4.4. Superposed Epoch Analysis for Extreme Magnetic Storms

Based on the density values scattered around ~ 10 /cc as shown in the top panel Figure 3, possible sheath-like (high density > 15 /cc) and cloud-like (low density < 5 /cc) large-amplitude ($Bz < -25$ nT) events were selected for further superposed epoch analysis. Note that this is a biased selection, not based on a cluster analysis, just to discuss possibly different populations. We omitted the medium-density events (5–15/cc) because they can belong either high or low events considering the large error in density estimation of our model.

The left panels in Figure 4 show the result for the sheath-like events. The basic characteristic of the rapid evolution of these storms main phase with spiky AE enhancement (less than ~ 6 hr) is clearly indicated by the superposed epoch analysis for sheath-type selection.

The right panels in Figure 4 show the result for the cloud-like events. The main phase and duration of strong SBZ are several times longer than those of the sheath-like storms. Moreover, the peak amplitude of SBZ is > 50 nT,

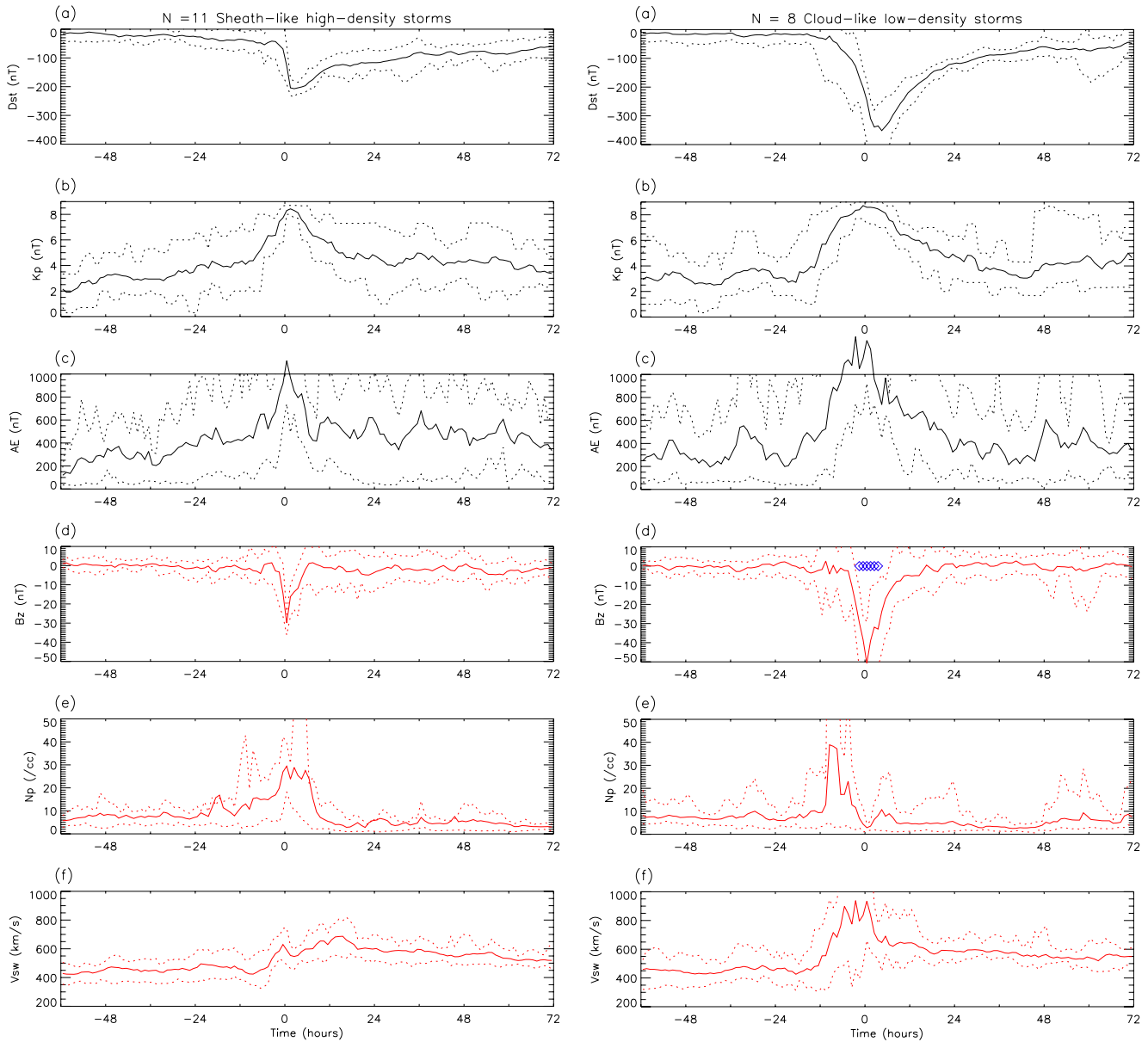


Figure 4. (left) Superposed epoch analysis for high-density ($N_p > 15/\text{cc}$) strong SBZ ($B_z < -25$ nT) and (right) low-density ($N_p < 5/\text{cc}$) strong SBZ ($B_z < -25$ nT) events. The format is the same as Figure 2. The blue diamonds in panel (d) show the Alfvén Mach number < 4 . The event lists are summarized in Tables S1–S2 in Supporting Information S1.

which is also a few times larger than that of the sheath-like storms. The resultant magnetic storms of this population are, therefore, largest. The latest example of this type of storm occurred in November 2003 (Kataoka, Fairfield, et al., 2005). The most prominent difference in these extreme solar wind-driven storms arises from their low-Mach number state, as discussed in the next section.

5. Discussions

The AE and Kp indices in sheath-like events are comparable to those of cloud-like events, as shown in Figure 4, despite of the smaller amplitude of SBZ. This is consistent with the fact that the high-beta sheath regions have higher solar wind-magnetosphere coupling efficiencies than magnetic clouds (Myllys et al., 2016). In fact, the majority of the largest Kp values are related to the sheath regions (Kilpua, Isavnin, et al., 2013). The density effect on the AE index was also recently discussed by Ebihara et al. (2019) and Nakano and Kataoka (2021).

We turn now to a discussion of the most extreme situation of cloud-like events, which becomes especially important for discussing the 100-year event shown in Figure 3. Assuming the fraction of α particles to be a constant value of 5% for simplicity, we can calculate the Alfvén Mach number as follows:

$$M_A = 2.7 \left(\frac{V_{sw}}{600 \text{ km/s}} \right) \left(\frac{B}{25 \text{ nT}} \right)^{-1} \left(1 + 0.2 \left(\frac{f_\alpha}{0.05} \right) \left(\frac{n_p}{5 \text{ /cc}} \right) \right)^{1/2}. \quad (6)$$

The Alfvén Mach number ranges from 7 to 11 in a typical high-speed solar wind, but is less than half that during CMEs or magnetic clouds (Lavraud & Borovsky, 2008). The extreme magnetic storms, therefore, tend to occur under a low Alfvén Mach number.

In the special situation of a low Mach number, the bow shock has been observed far from the magnetopause (Fairfield et al., 2001). The magnetospheric interaction with the low-Mach solar wind can also be different from the usual high-Mach state because of the weak bow shock and resultant low-beta flow in the magnetosheath. Global MHD simulations predict that due to the relatively strong magnetic force, the cross-section of the magnetopause shape and the surrounding magnetosheath flow becomes asymmetric (Lavraud & Borovsky, 2008). In fact, Lavraud et al. (2007) reported in-situ observation data showing that the magnetosheath flow was accelerated beyond 1,000 km/s from the upstream solar wind speed of 650 km/s when the Alfvén Mach number was between 2.0 and 4.4. Other examples of low-Mach events include the inner magnetospheric response (Lugaz et al., 2016). Recent MHD simulation results indicate that the saturation of cross polar cap potential is dependent on the Alfvén Mach number (Lakka et al., 2018).

Even more extreme cases of sub-Alfvénic ($M_A < 1$) solar wind exist. It is suggested that the magnetosphere has a different shape, called Alfvén wings (Chane et al., 2012), rather than the usual comet tail shape. Ridley (2007) showed global MHD simulation results indicating that the northern and southern lobes were separated. The formation of the Alfvén wings is one of the proposed mechanisms to explain the saturation of cross polar cap potential (Kivelson & Ridley, 2008; Ridley, 2007).

Future MHD simulations based on the average solar wind parameters obtained in this study should provide essential information for understanding the energy flow from the solar wind to the magnetosphere and ionosphere during extreme magnetic storms. However, for a detailed modern analysis, solar wind data with a time resolution higher than hourly (for example, 5 min values) will be necessary as input parameters for the simulations. Reproducing such high time resolution data remains a future work; however, there are several physical constraints regarding cloud-like and sheath-like events. For cloud-like events, it is possible to simply interpolate the hourly data into five-minute data, assuming that the magnetic field inside the magnetic cloud rotates smoothly. On the other hand, the fluctuation level of the magnetic field is strong in the sheath region (Kilpua, Hietala, et al., 2013) due to the existence of Alfvénic discontinuities (Kataoka, Watari, et al., 2005). The rapid changes of the disturbed condition embedded in the sheath region will also be important for reproducing the large geomagnetically induced currents.

It should be noted that the estimated minimum value of $B_z = -95$ nT in the March 1989 event (Figure 1) is significantly larger than the value of $B_z = -50$ nT previously estimated by Nagatsuma et al. (2015). They assumed the high-Mach number regime of $M \gg 1$ in which the compression ratio at the bow shock was approximated as 4. The observed strength of ~ 200 nT in the magnetosheath can, therefore, be converted to ~ 50 nT in the solar wind. However, the compression ratio becomes ~ 2 , if $M \sim 2$. Our estimate is thus consistent with the assumption of a low Mach number regime of $M \sim 2$.

Our results provide a continuous data set for the ~ 60 -year period since 1957, well before the OMNI2 data set. The data should prove extremely useful for interdisciplinary studies of the space climate, including efforts such as modeling the possible long-term impact of radiation belts and energetic electron precipitation events to the middle atmosphere via atmospheric chemistry (Seppala et al., 2015). While there are several other space climate studies focused on the reconstruction of long-term solar wind data using the geomagnetic activity indices (e.g., Lockwood, 2013; Svalgaard & Cliver, 2007), the main difference between those studies and ours is the time resolution. Whereas the other studies discuss a time variation of 27-day averaged values, our study focuses on hourly time variation and represents the first step in high-resolution (hourly time cadence) space climate study.

Filling the data gaps is another important aspect of the ESN method applied here. We recognize that there have been a number of excellent prior studies on this topic. For example, using a gap-filling method based on Singular Spectrum Analysis (Kondrashov et al., 2010, Kondrashov et al., 2014) report NRMSE values of 0.77, 0.48, and 0.10 for Bz, Np, and Vsw, respectively, which are better scores than our results, except for Bz. To facilitate further comparisons with other models, the hourly data produced in our study for Bz, Np, and Vsw covering the period from 1957 to 2017 are provided in Data Set S2. The output data from our study gives a starting point for improving the metrics. There are several possible pathways forward. For example, using other machine learning techniques appears promising. Using other datasets, such as the SME index of the SuperMAG project, also holds promise. Finally, we note that the methodology developed in this study can be generally used to complement datasets containing data gaps in various fields of science where there are limited observation data available.

6. Conclusions

With the help of an Echo State Network, a type of recurrent neural network, we discussed sheath-like and cloud-like average patterns in solar wind profiles associated with extreme magnetic storms. We also constructed a continuous solar wind data set for the years 1957 through 2017. Using our approach, the SBZ amplitude during the extreme March 1989 event was estimated to be $95 \text{ nT} \pm 10 \text{ nT}$, while for the one-in-100-year-event, the SBZ amplitude was estimated to be 137 nT .

Data Availability Statement

The OMNI2 data were provided by low-resolution OMNIWeb at NASA/GSFC (<https://omniweb.gsfc.nasa.gov/ow.html>). The Dst and AE indices were provided by Kyoto University (<http://wdc.kugi.kyoto-u.ac.jp/dstae/index.html>). The Kp indices were obtained from GFZ-Potsdam (<https://www.gfz-potsdam.de/en/kp-index/>).

Acknowledgments

R. Kataoka thanks Emile Toubert for accepting the sabbatical visit at OIST. The work of SN was in part supported by JSPS KAKENHI (Grant Number JP17H01704).

References

- Akasofu, S.-I. (1981). Energy coupling between the solar wind and the magnetosphere. *Space Science Reviews*, 28, 121–190. <https://doi.org/10.1007/bf00218810>
- Borovsky, J. E. (2008). The rudiments of a theory of solar wind/magnetosphere coupling derived from first principles. *Journal of Geophysical Research*, 113, A08228. <https://doi.org/10.1029/2007JA012646>
- Boteler, D. H. (2019). A 21st century view of the March 1989 magnetic storm. *Space Weather*, 17, 1427–1441. <https://doi.org/10.1029/2019SW002278>
- Burlaga, L. F., & Lazarus, A. J. (2000). Lognormal distributions and spectra of solar wind plasma fluctuations: Wind 1995–1998. *Journal of Geophysical Research*, 105(A2), 2357–2364. <https://doi.org/10.1029/1999JA900442>
- Chane, E., Saur, J., Neubauer, F. M., Raeder, J., & Poedts, S. (2012). Observational evidence of Alfvén wings at the Earth. *Journal of Geophysical Research*, 117, A09217. <https://doi.org/10.1029/2012JA017628>
- Ebihara, Y., Tanaka, T., & Kamiyoshikawa, N. (2019). New diagnosis for energy flow from solar wind to ionosphere during substorm: Global MHD simulation. *Journal of Geophysical Research: Space Physics*, 124, 360–378. <https://doi.org/10.1029/2018JA026177>
- Fairfield, D. H., Iyer, H. C., Desch, M. D., Szabo, A., Lazarus, A. J., & Aellig, M. R. (2001). The location of low Mach number bow shocks at Earth. *Journal of Geophysical Research*, 106(A11), 25361–25376. <https://doi.org/10.1029/2000JA000252>
- Hart, A., Hook, J., & Dawes, J. (2020). Embedding and approximation theorems for echo state networks. *Neural Networks*, 128, 234–247. <https://doi.org/10.1016/j.neunet.2020.05.013>
- Huttunen, K. E. J., Kilpua, S. P., Pulkkinen, A., Viljanen, A., & Tanskanen, E. (2008). Solar wind drivers of large geomagnetically induced currents during the solar cycle 23. *Space Weather*, 6, S10002. <https://doi.org/10.1029/2007SW000374>
- Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent neural networks* GMD Report 148. GMD - German National Research Institute for Computer Science.
- Kataoka, R. (2013). Probability of occurrence of extreme magnetic storms. *Space Weather*, 11, 214–218. <https://doi.org/10.1002/swe.20044>
- Kataoka, R. (2020). Extreme geomagnetic activities: A statistical study. *Earth, Planets and Space*, 72, 124. <https://doi.org/10.1186/s40623-020-01261-8>
- Kataoka, R., Fairfield, D. H., Sibeck, D. G., Rastaetter, L., Fok, M.-C., Nagatsuma, T., & Ebihara, Y. (2005). Magnetosheath variations during the storm main phase on 20 November 2003: Evidence for solar wind density control of energy transfer to the magnetosphere. *Geophysical Research Letters*, 32, L21108. <https://doi.org/10.1029/2005GL024495>
- Kataoka, R., & Miyoshi, Y. (2006). Flux enhancement of radiation belt electrons during geomagnetic storms driven by coronal mass ejections and corotating interaction regions. *Space Weather*, 4, S09004. <https://doi.org/10.1029/2005SW000211>
- Kataoka, R., & Pulkkinen, A. (2008). Geomagnetically induced currents during intense storms driven by coronal mass ejections and corotating interaction regions. *Journal of Geophysical Research*, 113, A03S12. <https://doi.org/10.1029/2007JA012487>
- Kataoka, R., Shiota, D., Kilpua, E., & Keika, K. (2015). Pileup accident hypothesis of magnetic storm on 2015 March 17. *Geophysical Research Letters*, 42, 5155–5161. <https://doi.org/10.1002/2015GL064816>
- Kataoka, R., Watari, S., Shimada, N., Shimazu, H., & Marubashi, K. (2005). Downstream structures of interplanetary fast shocks associated with coronal mass ejections. *Geophysical Research Letters*, 32, L12103. <https://doi.org/10.1029/2005GL022777>
- Kilpua, E., Koskinen, H. E. J., & Pulkkinen, T. I. (2017). Coronal mass ejections and their sheath regions in interplanetary space. *Living Reviews in Solar Physics*, 14, 5. <https://doi.org/10.1007/s41116-017-0009-6>

- Kilpua, E. K. J., Hietala, H., Koskinen, H. E. J., Fontaine, D., & Turc, L. (2013). Magnetic field and dynamic pressure ULF fluctuations in coronal-mass-ejection-driven sheath regions. *Annales Geophysicae*, 31, 1559–1567. <https://doi.org/10.5194/angeo-31-1559-2013>
- Kilpua, E. K. J., Isavnin, A., Vourlidas, A., Koskinen, H. E. J., & Rodriguez, L. (2013). On the relationship between interplanetary coronal mass ejections and magnetic clouds. *Annales Geophysicae*, 31, 1251–1265. <https://doi.org/10.5194/angeo-31-1251-2013>
- Kivelson, M. G., & Ridley, A. J. (2008). Saturation of the polar cap potential: Inference from Alfvén wing arguments. *Journal of Geophysical Research*, 113, A05214. <https://doi.org/10.1029/2007ja012302>
- Kondrashov, D., Denton, R., Shprits, Y. Y., & Singer, H. J. (2014). Reconstruction of gaps in the past history of solar wind parameters. *Geophysical Research Letters*, 41, 2702–2707. <https://doi.org/10.1002/2014GL059741>
- Kondrashov, D., Shprits, Y., & Ghill, M. (2010). Gap filling of solar wind data by singular spectrum analysis. *Geophysical Research Letters*, 37, L15101. <https://doi.org/10.1029/2010GL044138>
- Lakka, A., Pulkkinen, T. I., Dimmock, A. P., Myllys, M., Honkonen, I., & Palmroth, M. (2018). The cross-polar cap saturation in GUMICS-4 during high solar wind driving. *Journal of Geophysical Research: Space Physics*, 123, 3320–3332. <https://doi.org/10.1002/2017JA025054>
- Lavraud, B., & Borovsky, J. E. (2008). Altered solar wind-magnetosphere interaction at low Mach numbers: Coronal mass ejections. *Journal of Geophysical Research*, 113, A00B08. <https://doi.org/10.1029/2008JA013192>
- Lavraud, B., Borovsky, J. E., Ridley, A. J., Pogue, E. W., Thomsen, M. F., Reme, H., et al. (2007). Strong bulk flow plasma acceleration in Earth's magnetosheath: A magnetic slingshot effect? *Geophysical Research Letters*, 34, L14102. <https://doi.org/10.1029/2007GL030024>
- Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, 16, 2079–2102. <https://doi.org/10.1029/2018SW002067>
- Lockwood, M. (2013). Reconstruction and prediction of variations in the open solar magnetic flux and interplanetary conditions. *Living Reviews in Solar Physics*, 10, 4. <https://doi.org/10.12942/lrsp-2013-4>
- Lugaz, N., Farrugia, C. J., Huang, C.-L., Winslow, R. M., Spence, H. E., & Schwadron, N. A. (2016). Earth's magnetosphere and outer radiation belt under sub-Alfvénic solar wind. *Nature Communications*, 7, 13001. <https://doi.org/10.1038/ncomms13001>
- Matzka, J., Stolle, C., Yamazaki, Y., Bronkalla, O., & Morschhauser, A. (2021). The geomagnetic Kp index and derived indices of geomagnetic activity. *Space Weather*, 19, e2020SW002641. <https://doi.org/10.1029/2020SW002641>
- Myllys, M., Kilpua, E. K. J., Lavraud, B., & Pulkkinen, T. I. (2016). Solar wind-magnetosphere coupling efficiency during ejecta and sheath-driven geomagnetic storms. *Journal of Geophysical Research: Space Physics*, 121, 4378–4396. <https://doi.org/10.1002/2016JA022407>
- Nagatsuma, T., Kataoka, R., & Kunitake, M. (2015). Estimating the solar wind conditions during an extreme geomagnetic storm: A case study of the event that occurred on March 13–14, 1989. *Earth Planets and Space*, 67, 78. <https://doi.org/10.1186/s40623-015-0249-4>
- Nakano, S., & Kataoka, R. (2021). Virtual sounding of solar-wind effects on the AU and AL indices based on an echo state network model. *Annales Geophysicae Discussions*. <https://doi.org/10.5194/angeo-2021-54> [preprint]. in review.
- 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. *Journal of Geophysical Research*, 112, A01206. <https://doi.org/10.1029/2006JA012015>
- Ridley, A. J. (2007). Alfvén wings at Earth's magnetopause under strong interplanetary magnetic fields. *Annales Geophysicae*, 25(2), 533–542. <https://doi.org/10.5194/angeo-25-533-2007>
- Russell, C. T., Mewaldt, R., Luhmann, J., Mason, G., Rosengvinge, T., Cohen, C., et al. (2013). The very unusual interplanetary coronal mass ejections of 2012 July 23: A blast wave mediated by solar energetic particles. *The Astrophysical Journal*, 770, 38. <https://doi.org/10.1088/0004-637x/770/1/38>
- Seppala, A., Clilverd, M. A., Beharrell, M. J., Rodger, C. J., Verronen, P. T., Andersson, M. E., & Newnham, D. A. (2015). Substorm-induced energetic electron precipitation: Impact on atmospheric chemistry. *Geophysical Research Letters*, 42, 8172–8176. <https://doi.org/10.1002/2015GL065523>
- Svalgaard, L., & Cliver, E. W. (2007). Interhourly variability index of geomagnetic activity and its use in deriving the long-term variation of solar wind speed. *Journal of Geophysical Research*, 112, A10111. <https://doi.org/10.1029/2007JA012437>
- Tanaka, G., Yamane, T., Heroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., et al. (2019). Recent advances in physical reservoir computing: A review. *Neural Networks*, 115, 100–123. <https://doi.org/10.1016/j.neunet.2019.03.005>
- Tsurutani, B. T., & Lakhina, G. S. (2014). An extreme coronal mass ejection and consequences for the magnetosphere and Earth. *Geophysical Research Letters*, 41, 287–292. <https://doi.org/10.1002/2013GL058825>
- Welling, D. T., Love, J. J., Rigler, E. J., Oliveira, D. M., Komar, C. M., & Morley, S. K. (2021). Numerical simulations of the geospace response to the arrival of an idealized perfect interplanetary coronal mass ejection. *Space Weather*, 19, e2020SW002489. <https://doi.org/10.1029/2020SW002489>