



OPEN Reduction of the space dimension of parameters characterizing geomagnetic storms during the Solar Cycle 24

Agnieszka Siluszyk^{1✉}, Agnieszka Gil^{1,2,6}, Renata Modzelewska^{1,6}, Marek Siluszyk^{1,3,6}, Anna Wawrzaszek^{2,6} & Anna Wawrzynczak^{4,5,6}

Space weather phenomena related to solar activity are usually considered a threat mainly at high geomagnetic latitudes, yet recent studies show that countries at lower latitudes are not immune to their effects. In this work, we analyse thirteen geomagnetic storms during Solar Cycle 24 (2010–2021) using a set of twelve heliospheric and geomagnetic parameters. We aim to reduce the dimensionality of this parameter space from \mathbb{R}^{12} to \mathbb{R}^4 (or less) while preserving the essential physical information. To this end, we apply Principal Component Analysis (PCA) and show that the first 3–4 principal components explain at least 80% (typically 82–91%) of the total variance. The first component is consistently dominated by geomagnetic indices (Kp, Ec, Dst, ap, AE) and, in most storms, also by B , B_z and E_y , thus capturing the overall level of geomagnetic disturbance. The second and third components are mainly governed by solar wind properties, with robust pairings such as (SWs, SWt) and (B_z, E_y) , while B_y often forms a separate weakly coupled mode. We further use the leading components to fit simple regression models linking space-weather drivers to power-system loads and demonstrate that PCA can act as a diagnostic of unreliable interpolation in data with gaps. Our results indicate that PCA provides an efficient and physically interpretable reduction of heliogeomagnetic parameter space, facilitating the construction of statistical and machine-learning models for assessing and forecasting the impact of geomagnetic storms on technological infrastructure.

Keywords Data mining, Mining methods and algorithms, Clustering, Classification, Principal Components Analysis

Nowadays, many measurements and observations require the efficient use of computer data processing techniques and the development of new ones. Among many statistical analysis methods, there are well-known analysis of linear and nonlinear models and multivariate exploratory techniques, like cluster analysis, factor analysis, canonical analysis, correspondence analysis, discriminant analysis, multidimensional scaling or principal components, and classification analysis (e.g.^{1,2}). Statistical analysis methods are used to gain knowledge about real-world processes, such as investigating geomagnetic storms or building the most advanced processes or technologies in the computer world (e.g.^{3,4} and references therein). A huge amount of data on the one side is a fascinating field creating new possibilities, but on the other hand, it is a curse of data-space dimension. At least dozen geomagnetic indices and parameters delineating the state of the solar wind and, generally, heliosphere exist, so analyzing geomagnetic storms we deal with multidimensional systems in the vector space sense, where each parameter means a vector of data.

Geomagnetic storms, characterized by significant changes in the geomagnetosphere, are primarily triggered by disturbances in the solar wind (e.g.^{5–7}). These disturbances, such as coronal mass ejections or high-speed solar wind streams, set off a chain of interactions with Earth's magnetosphere, leading to intensified geomagnetic activity (e.g.^{8,9}). One notable change during those kind of periods is the modification of solar wind properties.

¹University of Siedlce, Institute of Mathematics, Siedlce 08-110, Poland. ²Space Research Centre, Polish Academy of Sciences (CBK PAN), Warsaw 00-716, Poland. ³Polish Air Force University, Deblin 08-521, Poland. ⁴University of Siedlce, Institute of Computer Science, Siedlce 08-110, Poland. ⁵National Centre for Nuclear Research, Otwock 05-400, Poland. ⁶Agnieszka Siluszyk, Agnieszka Gil, Renata Modzelewska, Marek Siluszyk, Anna Wawrzaszek and Anna Wawrzynczak contributed equally to this work. ✉email: agnieszka.siluszyk@uws.edu.pl

The solar wind's flow speed (SWs) typically increases, carrying more energetic particles towards Earth. This heightened flow speed contributes to enhanced geomagnetic disturbances and is closely correlated with other solar wind parameters¹⁰. One of the direct and immediate effects of disturbances is the rise in proton density (SWd) within the solar wind⁴. The SWd increases and when the solar wind encounters Earth's magnetosphere, directly and immediately contributes to heightened plasma interactions and more intense geomagnetic disturbances. Simultaneously, the solar wind temperature (SWT) may fluctuate. Although higher temperatures can increase plasma collisions and heating, variations in SWT during storms are influenced by complex interactions between the solar wind and Earth's magnetospheric plasma. The heliospheric magnetic field (HMF) strength (B) changes affecting the orientation and intensity of geomagnetic disturbances. Components of the magnetic field (Bx, By, Bz) vary, with Bz playing a particularly crucial role in inducing geomagnetic disturbances^{5,11}. The heliospheric electric field (Ey), derived from the solar wind velocity and Bz HMF component, reflects induced electric fields within the heliosphere, contributing to geomagnetic activity during storms. Geomagnetic indices such as AE, Dst, Kp, and ap provide quantifiable measures of geomagnetic activity during these time intervals. The interplay of processes and parameters during geomagnetic storms forms a complex network of interactions. This complexity underscores the necessity of understanding these intricacies in order to predict space weather conditions and mitigate the effects of these storms on technological infrastructure (e.g.^{4,12–14}). It further emphasizes the need for a comprehensive approach to studying these phenomena.

For studies of such complicated data sets, dimension reduction is the best and most necessary condition. Why is this necessary? Because in this way, we want to correct the quality of our model, an excess (redundancy) of the features can go to over-fitting, and too many variables in the model can be challenging to explain in practice; if we get fewer features, then the process will be more straightforward. The method of reducing the number of variables taken into account during the analysis should be carried out in such a way as to retain as much relevant information as possible. Reduction of the multidimensional system may consist of the following:

- *feature selection* limits the set of variables according to certain rules, e.g., features being excessively correlated with each other or statistically insignificant features. There are known filter, wrapper, embedded, and hybrid methods to feature selection of variables exploited in pattern recognition, or machine learning;
- *feature extraction* creates new derived features from the initial data set to obtain a smaller set of variables.

The feature selection method generally produces a subset consisting of the initial inputs. In contrast, the feature extraction method creates new composite attributes (e.g.¹⁵). The Principal Components Analysis (PCA) is one of the best techniques for the reduction of data dimensionality, i.e., the indication of the best features that fulfill the basic idea of principal components analysis - singular decomposition or spectral decomposition. Moreover, the procedure of PCA can be applied to detect structures and general regularities in the relationships between variables as well as their verification and to describe and classify the tested objects in new spaces defined by new variables. A multivariate statistical tool, PCA, carries information on variables using statistical variance; the highest variances determine a minimum number of significant components, see¹⁶. The initiators of principal components method were Karl Pearson¹⁷ and Harold Hotelling, who gave assumptions of principal components and developed theory for PCA^{18,19}. The capability of PCA are reported in mathematics, physics, IT, chemistry, and biology, etc.^{15,20}. We can see applications of PCA, such as image processing, text processing, speech recognition, and recommendation engines. Moreover, besides the classical method of PCA, we can get information on Robust PCA and Targeted PCA as the feature selection method. In Robust PCA²¹ the possible reduction of space dimension, is executed keeping the essential features and with minimal effect of outliers. In the Targeted PCA method, we can analyze a data set with dependent features¹⁵. Nowadays, extended of PCA are ICA (Independent Component Analysis)^{22,23}, *t*-Distributed Stochastic Neighbor Embedding²⁴, Uniform Manifold Approximation and Projection²⁵ and Multidimensional Scaling (MDS)^{26–30}.

In our work, we want to show the use of advanced machine learning methods to explore geomagnetic and heliospheric data. It is a non-parametric method and does not require assumptions about the distribution of the studied variables (features). The main idea of PCA is to replace a set of correlated features (if the variables are not correlated, PCA gives no data reduction) with a small number of uncorrelated so-called principal components that together can explain almost all the variability of the data. The result of this replacement is that the first component (new variable) describes the most variability because the components are linear combinations of input parameters. The second component is selected to be uncorrelated with the first one and explains most of the remaining volatility, etc.

The article consists of four sections. In Sect. 2.1, we present characteristics of geomagnetic and heliospheric parameters; in Section 2.2, we describe the method of PCA, and in Sect. 2.3 we present our results obtained with PCA, where we give a classification of main components, and fitting of the theoretical distribution to the empirical data, see subsection 2.4. Moreover, in subsection 2.5, we show verification of the interpolation correctness by using PCA. In Sect. 3 we summarize the presented investigations.

Methods and results

Characteristics of data

The expansion of the solar corona provides the solar wind with an embedded solar magnetic field that develops into the heliospheric magnetic field (HMF) (e.g.⁷). Time variation of the HMF follows a 22-year cycle with a reversal about every ~ 11 years at the time of extreme solar activity-solar maximum. An average value of HMF is ~ 5 nT at Earth's orbit for quiet solar conditions-solar minimum. Solar wind velocity is radially directed from the Sun. During periods of minimum solar activity, this radial flow becomes distinctively latitude dependent, changing from an average of 450 km/s in the equatorial plane to 800 km/s in the polar regions as observed by the Ulysses mission with maximum values observed up to 3000 km/s during extreme coronal mass ejections (e.g.⁹).

Here, we analyze solar wind parameters from OMNI (URL: omniweb.gsfc.nasa.gov): flow speed SWs [km/s], proton density SWd [N/cm^3], temperature SWT[K] and heliospheric magnetic field strength B [nT] and its components Bx[nT], By[nT], and Bz[nT] in GSM system and derived electric field.

Solar activity (SA) level is also revealed in changes in the galactic cosmic ray (GCR) flux, registered by a global network of ground neutron monitors^{31,32}. It is anti-correlated with the state of SA. GCR particles continuously reaching Earth deliver an irreplaceable source of information about the global state of the heliosphere in the Earth's vicinity³³. The neutron monitors (NM) measure the secondary cosmic rays on the ground, particularly the nucleonic part of the atmospheric cascade generated by primary cosmic rays³⁴. NMs registrations of the GCR flux allow solar and heliospheric properties to be traced.

We have also considered geoelectricfield E_c , which we have estimated from one-minute geomagnetic field data [B] in the frequency domain applying a 1D layered conductivity Earth model^{35,36}.

The last group of analyzed data is geomagnetic indices: AE, Dst, Kp, and ap. The Auroral electrojet index, AE, is obtained based on the geomagnetic variations in the horizontal component registered by the twelve selected observatories: Abisko, Amderma, Dixon Island, Tixie Bay, Pebek, Barrow, College, Yellowknife, Fort Churchill, Sanikiluaq, Narsarsuaq and Leirvogur³⁷. Those observatories are located in the northern hemisphere, along the auroral zone. AE-index was introduced to estimate global electrojet activity in the auroral zone. Its enhancements can serve as an indicator of the geomagnetic substorm onset. It continues growth for a few tens of minutes to one hour after substorm onset and vanishes in the order of one hour³⁸. The ring current index, Dst, is based on the measurements of magnetometers from middle latitudes: Kakioka, Honolulu, San Juan, and Hermanus³⁹. It is often used for the geomagnetic storm main phase identification⁴⁰. Local geomagnetic K-index from twelve magnetometers located in the range of 48-63 deg geomagnetic latitudes, both of north and south hemispheres, serves to obtain the standardized mean, i.e., Kp-index⁴¹. Kp-index is determined in a quasi-logarithmic scale in relation to disturbance amplitude. The corresponding linear amplitude geomagnetic index is ap. Kp- and ap-indices measure geomagnetic field variations responding only to disturbance changeability. The National Oceanic and Atmospheric Administration uses Kp-index to formulate the Geomagnetic Storms scale⁴².

All the above-described parameters were collected for the thirteen intensive geomagnetic storms, which appeared during the Solar Cycle 24.

Method of PCA

Principal component analysis (PCA) is a well-known technique that has been successfully applied for the analysis of various data sets, starting from mathematical, physical, medical, etc., to geophysical data (e.g.^{17,43–45} and references therein). First of all, the PCA reduces the dimension of space $X \subseteq R^n \rightarrow Z \subseteq R^m$, where $m < n$ and supports the pattern recognition, looking for possible regularities and dependencies between the investigated variables without prior knowledge of the data. The reduction consists of an indication of the 'most interesting' directions for our data, i.e., such directions that will distort the data the least when projected onto it¹⁷. By the smallest distortion, we mean the minimum value of the sum of the squares of the perpendicular projections in the direction. For a one-dimensional representation of P_i points, $i = 1, \dots, N$, the problem can be written as

$$\sum_{i=1}^N |O' P_i|^2 = \sum_{i=1}^N |O' P_i'|^2 + \sum_{i=1}^N |P_i P_i'|^2, \quad (1)$$

where O' is in a center of new coordinate system, (see, $O' = (\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i)$), while P_i' is a projection of the P_i . According to formula (1) minimum of $\sum_{i=1}^N |P_i P_i'|^2$ is equivalent maximum of $\sum_{i=1}^N |O' P_i|^2$, which is an attempt variance, i.e., $\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$, here $\bar{X} = 0$.

Consider vectors $X_j \in X$, ($j = 1, \dots, n$), $X \in R^n$ and $Z \in R^m$, where $m < n$. Each X_j has the form $X_j = (x_{ij})$, $i = 1, \dots, N$, where N denotes the number of observations. Obviously, both the variances X_j , ($\text{Var} X_j = \frac{1}{N-1} \sum_{i=1}^N (x_{ij} - \bar{X}_j)^2$), $\bar{X}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$ and covariance of matrix X ($\text{cov}(X_j, X_k) = \frac{1}{N-1} \sum_{i=1}^N (x_{ij} - \bar{X}_j)(x_{ik} - \bar{X}_k)$) are known. Instead of the covariance, we can also study the correlation of the matrix X . Now, $\bar{X} = (\bar{X}_1, \dots, \bar{X}_n)^T$ is transposition of the column vector \bar{X} , moreover $x_i^T = (x_{i1}, x_{i2}, \dots, x_{in})$. According to this remark, we have $S = (s_{jk})$

$$s_{jk} = \frac{1}{N-1} \sum_{i=1}^N (x_{ij} - \bar{X}_j)(x_{ik} - \bar{X}_k) \quad (2)$$

$$S = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})(x_i - \bar{X})^T \quad (3)$$

Let a be vector with the length 1. The projection of X on a is $a^T X$. Since $z_i = a^T x_i$, it follows that $\bar{Z} = \frac{1}{N} \sum_{i=1}^N z_i = a^T \bar{X}$. We thus get

$$z_i - \bar{Z} = a^T x_i - a^T \bar{X} = a^T (x_i - \bar{X}), \tag{4}$$

whereas

$$(z_i - \bar{Z})^T = (x_i - \bar{X})^T a. \tag{5}$$

Now from (4) and (5) we calculate

$$(z_i - \bar{Z})^2 = a^T (x_i - \bar{X}) (x_i - \bar{X})^T a, \tag{6}$$

where $(z_i - \bar{Z})^2 = (z_i - \bar{Z}) (z_i - \bar{Z})^T$. This gives

$$Var (a^T X) = a^T \sum X a, \tag{7}$$

because

$$\frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{Z})^2 = a^T \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X}) (x_i - \bar{X})^T a = a^T S a. \tag{8}$$

Consequently, there exists an $a \in R^n$ such that

$$a^T a = 1 \tag{9}$$

and the projection onto a , i.e., $a^T x_1, \dots, a^T x_N$ maximizes the variance among all a satisfying (9). In other words, above problem consists of looking for maximum of $a^T S a$ with condition (9). Using the Lagrange Multiplier method

$$L(\lambda, a) = a^T S a - \lambda (a^T a - 1) \tag{10}$$

we know that (10) holds only in the case

$$S a = \lambda a, \tag{11}$$

i.e., if the vector a is eigenvector of S -matrix and a has to correspond of maximal eigenvalues: $a^T S a = a^T \lambda a = \lambda$. This formula says that we are looking for a_1 which is the normalized eigenvector corresponding to the largest eigenvalue of the matrix S equal to λ_1 . Then we look for the normalized $a_2 \perp a_1$ such that the projections on the direction perpendicular to a_1 have the greatest variance, now $a_2 \perp a_1$ is an eigenvector corresponding to λ_2 , where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ are the eigenvalues. a_3 such that $a_3 \perp a_1, a_3 \perp a_2$ is an eigenvector corresponding to λ_3 . In this way, we obtain a_1, \dots, a_m corresponding to a new frame of coordinates in R^m . There exists i_0 that the main components for $i > i_0$ are omitted as they do not contribute significant information. The variance, which is the lengths of the vectors of projections X_1, \dots, X_n ($= \sum_{j=1}^n \|P X_j\|^2$), into the subspace spanned by a_1, \dots, a_m is equal to $\lambda_1 + \lambda_2 + \dots + \lambda_m$, i.e. $\lambda_j = \text{Var}(Z_j)$. If the correlation analysis is based on the correlation matrix, these values are interpreted as correlation coefficients between the original variables and the given principal component. Geometrically, the subspace spanned at the first m cardinal directions minimizes the sum of the squares of the distances of points among the subspaces m -dimensional. The first method of choosing the properly number principal components is a value $P_m = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{\lambda_1 + \lambda_2 + \dots + \lambda_n}$ which really means a value of variance. For this purpose, the PCA method orders the vectors according to the eigenvalues from the largest to the smallest, and it eliminates the variables with the smallest eigenvalues. The sum of variances of the variables Z_p equals the sum of the original variables X_j . Therefore, the transformation of the variables does not lead to the loss of information about the processes studied. According to the above remark, we are looking for the smallest m such a value $0 < a \leq 100\%$ (for example 80%, 90%, 95%, etc.), that $P_m \geq a$, i.e., a variance value is sufficient. Another way for choosing the smallest m is the scree test (Cattell criterion), i.e., we plot λ_m as a function of m . In practice, we choose the minimum index as m before the graph begins to flatten⁴⁶ for the eigenvalues and their percentage sum. The flat part of the graph corresponds to unstructured noise that we cannot interpret. It is worth pointing out that the principal components method for the original data works sensibly when the $X = (X_1, \dots, X_n)$ components are measured in the same units, moreover, variances of the components are comparable, i.e. $a^T X$ is interpreted in the same units. Otherwise, the directions of the initial principal components will correspond to the coordinates with the highest variance. Thus we change (standardize x_{ij})

$$x_{ij} \longrightarrow \frac{x_{ij} - X_j}{\sqrt{s_{jj}}}. \tag{12}$$

From this we obtain $s_{jk}, j, k = 1, \dots, n$, as a correlation matrix. The principal components method for standardized variables x_{ij} (12) gives different results than for the original variables. If the correlation analysis

is based on the correlation matrix, its values are interpreted as correlation coefficients between the original variables and the given principal component. In general, PCA should be used for measurable, linearly correlated variables, with Pearson coefficients ≥ 0.3 . However, in the literature, we can see the application of PCA to ordinal variables. In this study, before using the PCA method (for measurable or not variables), the Bartlett test has been applied. This test assumes that a correlation matrix for tested variables is identity, i.e., $H_0 : R = I$. This result means that all Pearson's correlation coefficients $r_{ps} = 0$ for $p \neq s$ and $r_{ps} = 1$ for $p = s$, ($p, s = 1, \dots, n$). By statistics $U = - (N - 1 - \frac{2n+5}{6}) \sum_{j=1}^n \ln \lambda_j$ with χ^2 distribution for $n(n - 1)$ degrees of freedom we verify

H_0 . A more complete check we performed using the second condition for PCA, i.e., KMO coefficient, it means Kaiser-Mayer-Olkin one which is given by the formula⁴⁶

$$KMO = \frac{\sum_{p \neq s} \sum_{p \neq s} r_{ps}^2}{\sum_{p \neq s} \sum_{p \neq s} r_{ps}^2 + \sum_{p \neq s} \sum_{p \neq s} \hat{r}_{ps}^2}, \tag{13}$$

where \hat{r}_{ps} means partial coefficient of correlation and has the form

$$\hat{r}_{ps} = - \frac{c_{ps}}{\sqrt{c_{pp}c_{ss}}}. \tag{14}$$

The c_{ps} , c_{pp} , c_{ss} elements are the algebraic complement of r_{ps} , r_{pp} and r_{ss} , respectively. If $0.5 < KMO < 1$ then $\sum_{i \neq j} \sum_{i \neq j} \hat{r}_{ps}^2 < \sum_{p \neq s} \sum_{p \neq s} r_{ps}^2$.

Classification of main components

PCA results were subjected to detailed analysis to classify heliospheric and geomagnetic parameters into main components. The first 3-4 components in all analyzed storms describe over 80% of the total cumulative value. In particular, in Fig. 1, we present the cumulative and explained variance scales by the first, second, third, and fourth PCs as a percentage of total variance for each geomagnetic storm. For the first PC, we start from 41.17 to 59.90% of the variance, for the second from 58.26–76.68%, 72.24–86.20% for the third, and 82.72–91.18% for the fourth PC, i.e., the increasing variance for the following components is smaller and smaller. The magnitude of the corresponding eigenvalue reflects the importance of each variable. The larger these absolute values are, the more specific feature contributes to that principal component. In Figs. 2 and 3 we present projections of twelve variables on the factor-plane 1×2 principal components for geomagnetic storms on March 9, 2012, and on

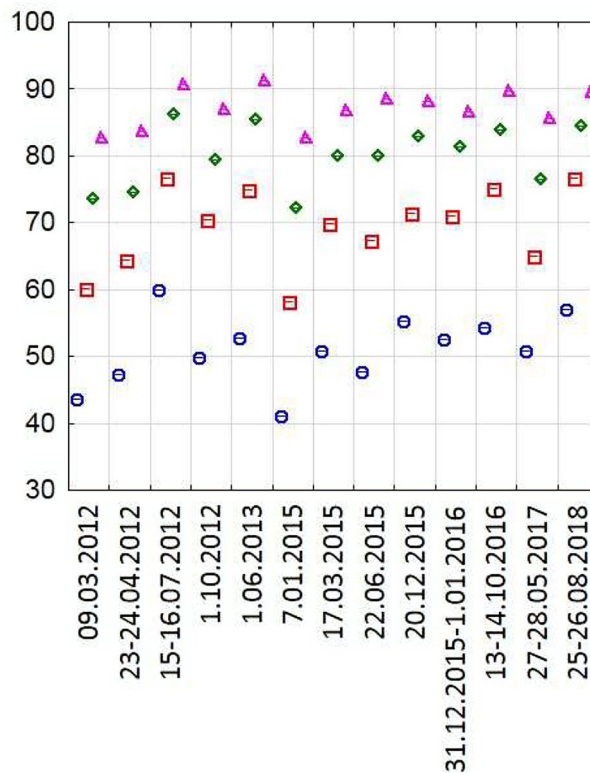


Fig. 1. Range plot of multiple variables of the Cumulative % of Variance of the 1st PC-(blue dot), the 2nd-(red square), the 3rd-(green rhombus), and the 4th PC-(purple triangle) during the geomagnetic storms in Solar Cycle 24.

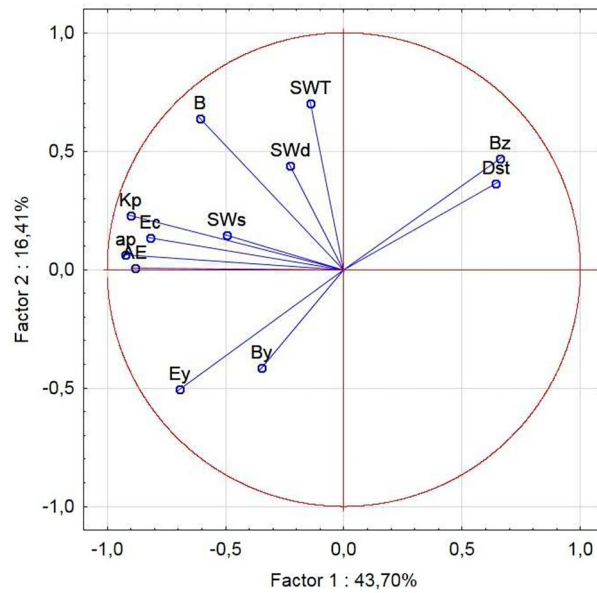


Fig. 2. Projection of the variables on the factor-plane 1 × 2 principal component for geomagnetic storms on March 9, 2012.

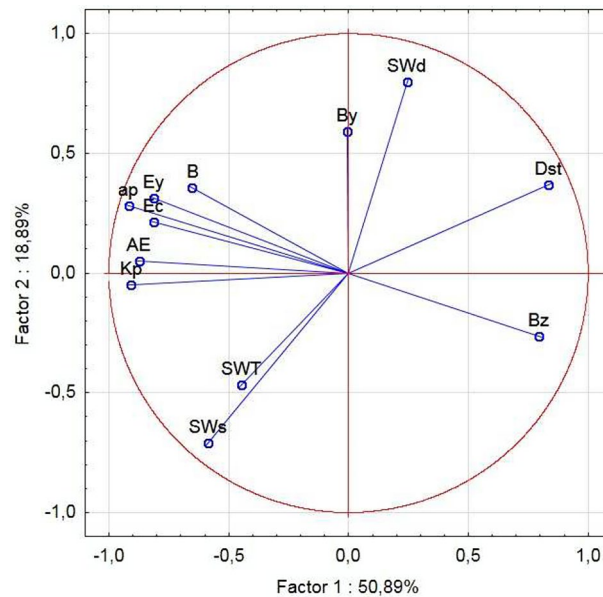


Fig. 3. Projection of the variables on the factor-plane 1 × 2 principal component for geomagnetic storms on St. Patrick Day on March 17, 2015.

St. Patrick's Day, i.e., on March 17, 2015, respectively. All factor coordinate loadings fall within the unit circle. In addition, this circle allows for a graphical assessment of the degree to which the current set of components represents each variable. The further a given variable is placed from the circle's center (the longer the vector), the better the current coordinate system represents it. For example, features ap, Kp, AE, and Ec significantly negatively affect the value of the first component in both storms of 9.03.2012 and 17.03.2015. On the other hand, the influence of Bz and Dst are highly significant but have an opposite effect on the first component than ap, Kp, AE, and Ec. The position of the vector SWT, SWd, Fig. 2 and By, SWd (see Fig. 3) confirms its significant positive impact, but only on the second component. Especially the By interaction, for the first component is negligible. The difficulty in classifying parameters to a given component lies in analyzing many criteria, e.g., factor coordinates, correlation values, and communalities, and then manually assigning parameters to each component. The parameters cannot be assigned in any order. The parameters in each component should be sorted first in descending order by the absolute value of the factor coordinates. Then, the parameters are first classified into the

first principal component. In our case, we assumed that the value of the factor coordinates of a given parameter in the first component cannot be lower than 0.65, and then the remaining parameters are decomposed into the succeeding components, assuming that in each component, they must have the largest factor coordinates values and must be selected one after the other. This often leads to a situation where a given parameter goes to the next component despite the high value of the factor coordinates in the previous component. Usually, there can be only one parameter in a given component. In this work, we adopt a threshold value of 0.65 for the first component in the geomagnetic storm of March 17, 2015. The 12 parameters characterizing this storm are systematically grouped into four main components, reflecting their physical relevance and interdependencies. For other geomagnetic storms, the threshold for assigning parameters to the first component varies accordingly with the storm-specific characteristics, e.g., 0.64 for March 9, 2012; 0.61 for January 7, 2015; 0.65 for June 22, 2015; and 0.62 for August 25–26, 2018. Likewise, the second components exhibit corresponding threshold values that account for variability across events, including 0.64 for March 9, 2012; 0.54 for April 23–24, 2012; 0.57 for June 1, 2013; 0.63 for January 7, 2015; and 0.59 for December 20, 2015. The detailed breakdown of all parameters for the 13 geomagnetic storms considered is presented in Tables 1 and 2. Tables 1 and 2 describe the variables according to their contributions to the first two principal components; for each principal component, we can see which variable constructs most to that component.

After classifying the heliospheric and geomagnetic parameters into four main components, it is noticed that in the first component of all analyzed storms, there are classified geomagnetic indices: Kp, Ec, Dst, ap, and AE. In addition, in the first component of two-thirds of the storms, there are also heliospheric parameters: B, Bz, and Ey with the value of the factor coordinates above 0.75. The second and third components include mainly solar wind or heliospheric parameters not classified in the first component. In the case of solar wind parameters, a strong correlation can be observed between the temperature parameter SWT and solar wind speed SWs. SWT and SWs are having strong input in the main components of more than half of the analyzed storms. We speculate that it might be a response on the CME passage that drives a shock wave and magnetic cloud, accompanied by abrupt changes in solar wind speed and temperature. This might be a consequence of shifting from turbulent sheath to more stable structure of magnetic cloud⁴⁷. Bz and Ey can also be given as a pair of strongly correlated parameters (what is obvious by definition), which always belong to the same component. Noteworthy is the By parameter, which in two-thirds of the considered storms is classified alone, in most cases in the last main component.

The PCA algorithm has saved at least 82% of the variance with just four PCs, giving a significant reduction from the original space's dimension, which is of great importance in practical model building.

9.03.2012	ap	Kp	AE	Ec	Ey	Bz	Dst	-	-
	-0.9218	-0.9026	-0.8815	-0.8167	-0.6931	0.6595	0.6416	-	-
23-24.04.2012	AE	ap	Kp	Ec	-	-	-	-	-
	-0.9132	-0.9039	-0.8944	-0.7985	-	-	-	-	-
15-16.07.2012	ap	B	Kp	Ey	AE	Bz	Ec	Dst	-
	-0.9542	-0.9460	-0.9332	-0.9109	-0.9104	0.9097	-0.8407	0.8154	-
1.10.2012	ap	Kp	AE	B	Ec	Dst	-	-	-
	-0.9278	-0.9217	-0.8765	-0.8483	-0.8020	0.7468	-	-	-
1.06.2013	ap	Kp	AE	B	Ec	Dst	-	-	-
	-0.9359	-0.9242	-0.9128	-0.8915	-0.8775	0.7941	-	-	-
7.01.2015	ap	Kp	AE	Ec	SWd	Dst	-	-	-
	-0.8876	-0.8806	-0.8546	-0.8376	-0.6206	0.6143	-	-	-
17.03.2015	ap	Kp	AE	Dst	Ec	Ey	Bz	-	-
	-0.9171	-0.9079	-0.8730	0.8352	-0.8139	-0.8117	0.7943	-	-
22.06.2015	ap	Kp	Ec	B	AE	SWs	Dst	SWT	-
	-0.9193	-0.9183	-0.8691	-0.8608	-0.8565	-0.7313	0.7008	-0.6455	-
20.12.2015	ap	Bz	Ey	Kp	Ec	AE	B	Dst	-
	-0.9552	0.9138	-0.9069	-0.9035	-0.8680	-0.8614	-0.8109	0.7842	-
31.12.2015-1.01.2016	ap	Kp	AE	Ec	Ey	Bz	B	Dst	-
	-0.9665	-0.9127	-0.9094	-0.8903	-0.8507	0.8474	-0.8463	0.7731	-
13-14.10.2016	ap	AE	Dst	Bz	Ey	Kp	B	Ec	-
	-0.9417	-0.9326	0.9223	0.8715	-0.8629	-0.8481	-0.8112	-0.7796	-
27-28.05.2017	AE	ap	Kp	Ec	Ey	Bz	B	Dst	-
	-0.9660	-0.9551	-0.9400	-0.9057	-0.8604	0.8489	-0.7359	0.6849	-
25-26.08.2018	Kp	ap	AE	Bz	Dst	Ey	Ec	B	SWd
	-0.9204	-0.9193	-0.9079	0.8620	0.8571	-0.8549	-0.8016	-0.8014	-0.6228

Table 1. The composition of the 1st component of the PCA method for chosen geomagnetic storms with factor coordinates of the variables.

9.03.2012	SWT	B			
23-24.04.2012	SWd	SWs	SWT	B	Dst
15-16.07.2012	SWd	SWT	SWs	-	-
1.10.2012	Bz	Ey	SWs	-	-
1.06.2013	SWs	By	SWT	-	-
7.01.2015	Ey	Bz	B	-	-
17.03.2015	SWd	SWs	-	-	-
22.06.2015	Bz	Ey	-	-	-
20.12.2015	By	-	-	-	-
31.12.2015-1.01.2016	SWd	SWT	By	-	-
13-14.10.2016	SWs	SWT	-	-	-
27-28.05.2017	SWT	SWs	-	-	-
25-26.08.2018	SWs	SWT	-	-	-

Table 2. The composition of the 2nd component of the PCA method for chosen storms.

Fitting the theoretical distribution to the empirical data

Modeling actual processes is an attempt to explain in a mathematical and physical sense the operation of unknown mechanisms from the real world and describe these phenomena by formulas. The construction of a mathematical model can be compared to the construction of equipment, a house, or the operation of a particular program. During the fitting of the distribution of a theoretical variable to empirical data, special attention should be paid to the statistical analysis of the values of explained and initial data, as well as to computational experiments and knowledge of the properties of different models. The basis for modeling is data obtained from active or passive observations. In the case of natural phenomena such as the geomagnetic storms described above, the data comes from passive observations. To explain given phenomena, we must prepare a special modeling theory, observations from various sources, and certain assumptions. The Least Squares Method (LSM) is used to construct mathematical models based on time series. This new model and its parameters must fulfill very stringent conditions. In addition, besides the LSM, which can be used for quantitative variables, one can use the Generalized Method of Moments or the Maximum Likelihood Method. A substantive and statistical verification of the model is used to check the quality of the constructed model, i.e., more than twenty criteria are used to check the accuracy of the estimated model parameters and the accuracy of fitting the model to empirical data. Next, the modeling methods will be used not only to estimate model parameters but also to test hypotheses about the parameters and, above all, to prepare forecasts. What did we gain? We started with 12 variables, and after laborious calculations, we obtained 12 variables (components) Z_1, Z_2, \dots, Z_{12} . However, we gained something. First, the new variables (components) are orthogonal to each other, i.e., they are mutually uncorrelated. We can use such components in further studies, e.g., in discriminant analysis or in multiple regression analysis, where the assumption of noncollinearity is required. Second, each subsequent component explains a smaller and smaller part of the variability of the original data-set. At some point we determine the component that explains a negligible part of the primal variability. Let's look at the results obtained for two of the considered storms in more detail. It turns out that two components in the storm on March 17, 2015 explain 69.7824% of the total variability, three components explain 79.9314%, and with the fourth component we have 86.8543% of the explained total variability, whilst for sixth there is 94.2137%. Therefore, the remaining components, which explain 13.1457% or 5.7863%, respectively, are unnecessary. If so, we have reduced the variables by using only the most essential components in further analysis. Instead of a twelve-dimensional space, if we agree on a slight loss of information, a three- or four-dimensional space is enough. Similarly, we solve a problem of reduction of dimension space from R^{12} to R^4 for variables obtained at storm on March 9, 2012. Here, we present a system of equations (15) describing geomagnetic storms on March 17, 2015 in the space of new variables (Z_1, Z_2, Z_3, Z_4) obtained by using PCA method

$$\left\{ \begin{array}{l}
 (\lambda = 6.107), Z_1 = -0.3711ap - 0.3674Kp - 0.3533AE + 0.337957Dst \\
 \quad - 0.32937Ec - 0.3285Ey + 0.3214Bz - 0.2639B \\
 \quad - 0.2377SWs - 0.1809SWT + 0.0986SWd - 0.0018By, \\
 (\lambda = 2.2671), Z_2 = 0.530723SWd - 0.471659SWs + 0.393140By \\
 \quad - 0.309048SWT + 0.243975Dst + 0.236055B \\
 \quad + 0.207727Ey + 0.185705ap - 0.176257Bz \\
 \quad + 0.140535Ec + 0.032600AE - 0.032241Kp \\
 (\lambda = 1.218), Z_3 = -0.508288B - 0.422086SWT - 0.399381SWd \\
 \quad - 0.380799Bz + 0.353132Ey + 0.222948By \\
 \quad - 0.211464Kp - 0.132444Dst - 0.114241ap \\
 \quad + 0.079514AE - 0.045733SWs - 0.029009Ec \\
 (\lambda = 0.831), Z_4 = 0.797355By + 0.451338SWT + 0.238184SWs \\
 \quad - 0.216990Ec + 0.182617Dst - 0.116138SWd \\
 \quad + 0.065360AE - 0.042021Kp - 0.041914Ey \\
 \quad + 0.036341Bz - 0.022561B + 0.0015531ap
 \end{array} \right. \quad (15)$$

As an example Fig. 4 present hourly time variations of the four main components (Z_1, Z_2, Z_3, Z_4) during the geomagnetic storm on March 17, 2015 for standardized initial data, respectively.

The next task consists of building a linear regression model $L(Z_1, Z_2, Z_3, Z_4)$ and determining the coefficients of this equation b_0, b_1, b_2, b_3, b_4 for some real data observed during Patrick’s Day geomagnetic storm on March 17, 2015 obtained from Polish Transmission System Operator (PSE). We tested, for example, mean active powers and mean total active power ($P_i mean$, where $i \in \{1, 2, 3\}$, and $P_{Tot} mean$), at different localizations. The crucial are Fischer-Snedecor’s test, multiple R , and multiple R^2 levels. Moreover, we are examining this estimation’s standard error S_e and verify the model assumptions. The number of estimated parameters is 4, the number of observations is 192. As an example we present the calculations for $P_1 mean$:

$$\begin{aligned}
 P_1 mean = & -0.0000000213757296 - 0.158841028Z_1 - 0.219240228Z_2 \\
 & - 0.114942118Z_3 + 0.291986343Z_4
 \end{aligned} \quad (16)$$

with $F(4, 187) = 24.9681799$, $p < 1.40914479 * 10^{-16}$, multiple $R = 0.590036411$ and multiple $R^2 = 0.348142966$. Moreover, the standard error of estimate is $S_e = 0.818099241$. Values F and p inform us that linear regression equation is significant, as well as each $b_l, l \in \{1, 2, 3, 4\}$ (p -value: 0.00000000347615992, 0.0000000848516137, 0.0334180556, 0.0000121348248) whilst correlation coefficient ($\approx 59\%$) means that there is strong linear connection between $P_1 mean$ and Z_1, Z_2, Z_3, Z_4 . Further thorough verification of the model involves examining homoscedasticity, autocorrelation of residuals, normality of the residuals’ distribution, etc., which is not our goal in this work. These results will be presented for the PSE data in consecutive works.

In Fig. 5, we present the fitting of a polynomial function of 6th degree to the initial (experimental) data for given storms. We observe that all fitting curves to the initial data during storms 7.01.2015, 17.03.2015, 22.06.2015, 20.12.2015, and 31.12.2015-1.01.2016 around maximum of Solar Cycle 24 have the same phase shape, which confirms that parameters of each geomagnetic storm at the studied period are stable. Unfortunately, we can’t say it about initial data of geomagnetic storm parameters for storms 9.03.2012, 23-24.04.2012, 15-16.07.2012, 1.10.2012 and 1.06.2013.

Verification of the interpolation correctness using PCA

During our research, we discovered a new PCA functionality that, to our knowledge, has not been described by anyone before. Quite often, when analyzing actual high-resolution data, there are interruptions in measurements.

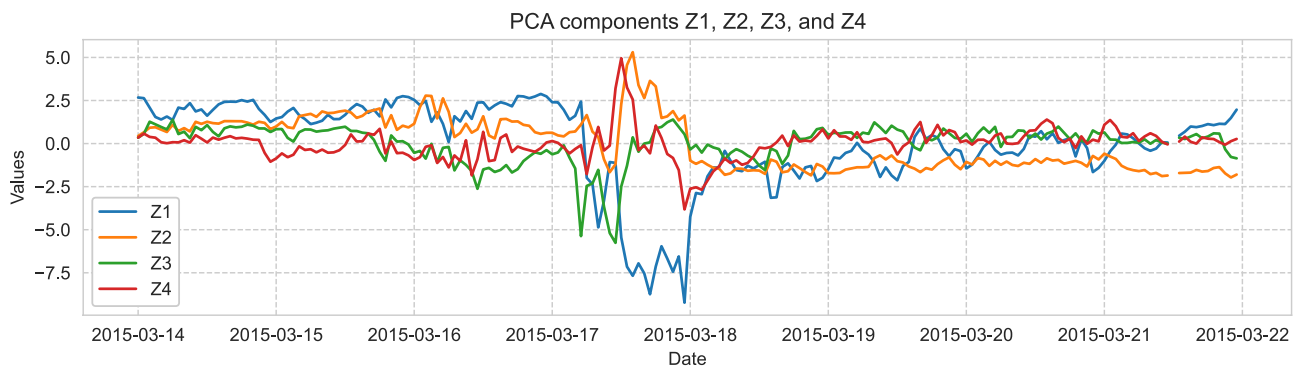


Fig. 4. Time course (hourly) of the main components (Z_1, Z_2, Z_3, Z_4) during the geomagnetic storm on March 17, 2015 for standardized initial data.

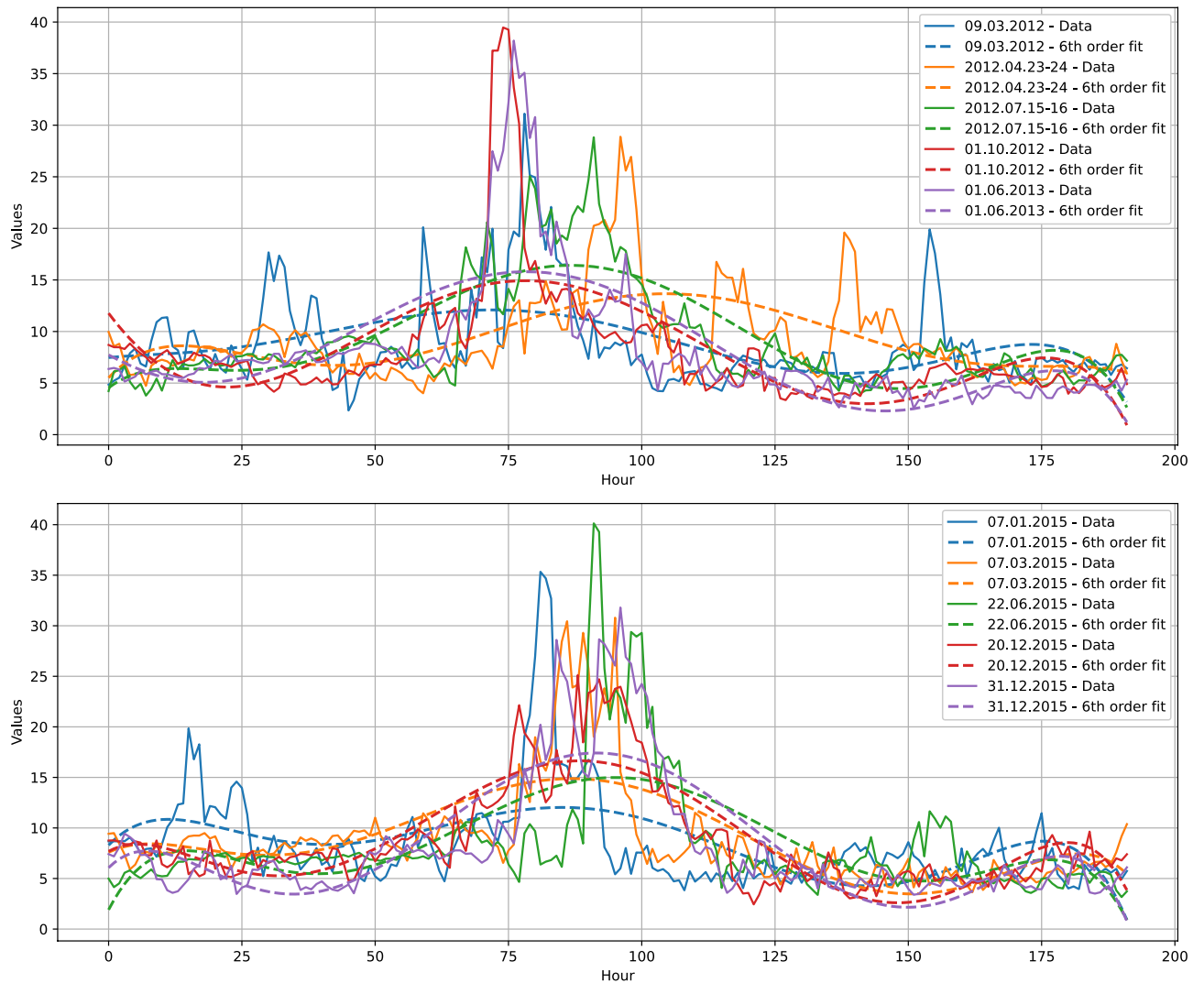


Fig. 5. The graphs of fitting of polynomial function of 6th degree to the initial data for storms 9.03.2012, 23-24.04.2012, 15-16.07.2012, 1.10.2012, 1.06.2013 (upper panel, storms at ascending phase of SC24), and 01.2015, 17.03.2015, 22.06.2015, 20.12.2015 and 31.12.2015-1.01.2016 (lower panel, storms around the maximum of SC24).

This disruption in the data characterizing the solar wind and HMF occurred during the storm in September 2017. The most common approach in this situation is to use interpolation to fill existing gaps in the data sets. However, after interpolating these data and performing PCA analysis, we noticed that the distribution of vectors in the biplot is quite chaotic (Fig. 6) and differs significantly from the distributions observed for other storms (Fig. 3), where there were no data gaps. After conducting various tests, we found that the phenomena occurring on the Sun and in the near-heliosphere during explosive events are too complex and nonlinear to reflect their nature by simple interpolation. Therefore, PCA analysis can be used to verify the effectiveness and correctness of the interpolation performed on real data with gaps.

Discussion

In this work, we focused on the problem of dimensionality reduction in the multidimensional space of real data, performed using the PCA method. More precisely, we systematically selected and considered 13 intense geomagnetic storms observed during Solar Cycle 24. Nature and complexity of the analyzed storms were described by a set of 12 geomagnetic and heliospheric parameters. The application of PCA to the observational data allowed for the reduction of the original data set of 12 variables to 4 main principal components, resulting in a reduced data set represented by the first 4 principal components in a new 4-dimensional space. Moreover, the classification of variables using PCA helped to separate geomagnetic and heliospheric parameters into distinct bins, which leads to greater clarity in observing the interrelationships within these bins. In addition, PCA revealed the incorrectness of applying interpolation in the context of parameters (with data gaps) describing space weather phenomena.

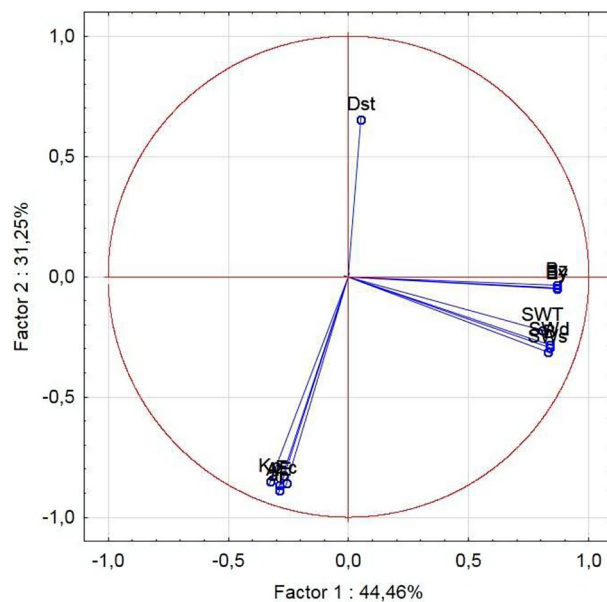


Fig. 6. Projection of the variables on the factor-plane 1×2 principal component for the geomagnetic storm in 7-8.09.2017.

The next stage of the study will involve developing a predictive model based on a neural network to forecast Polish Transmission System' disturbances during geomagnetic storms. The reduction of input parameters from 12 to 4 allows us to focus on the most significant features that influence the behavior of the system while minimizing computational complexity. This reduction in dimensionality is particularly beneficial given the limited availability of training data, as it helps mitigate the risk of overfitting and improves the generalization capabilities of the model.

Using machine learning techniques, the model aims to capture complex, nonlinear relationships within the geophysical and heliospherical parameters space, which are difficult to identify using classical statistical methods. The neural network will be trained on historical geomagnetic storm data, learning patterns that can be used to anticipate disturbances in power network. Such a predictive tool has the potential to improve early warning systems and mitigate the adverse effects of space weather on technological infrastructure, particularly in regions previously considered less vulnerable to space weather phenomena^{4,13,48–51}. Further work will focus on optimizing the neural network architecture, selecting the most suitable activation functions, and fine-tuning hyperparameters to maximize predictive accuracy. Additionally, we will explore the integration of external datasets, such as solar wind parameters and geomagnetic indices, to enhance model performance and robustness.

Data availability

The data sets used during the current study are publicly available: <https://omniweb.gsfc.nasa.gov> and <https://rtbe.ligf.edu.pl>. Data of transmission lines anomalies from the Polish Transmission System Operator are confidential

Received: 4 August 2025; Accepted: 12 February 2026

Published online: 21 February 2026

References

- Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (Wiley, 1958).
- Konishi, S. *Introduction to Multivariate Analysis: Linear and Nonlinear Modeling* 1 edn (Chapman and Hall/CRC, 2014).
- Tian, T. et al. Statistical study on interplanetary drivers behind intense geomagnetic storms and substorms. *Earth Planet. Phys.* **3**, 380–390. <https://doi.org/10.26464/epp2019039> (2019).
- Gil, A. et al. Evaluating the relationship between strong geomagnetic storms and electric grid failures in Poland using the geoelectric field as a gic proxy. *J. Space Weather Space Clim.* **11**, 30. <https://doi.org/10.1051/swsc/2021013> (2021).
- Gonzalez, W. D. et al. What is a geomagnetic storm?. *J. Geophys. Res.* **99**, 5771–5792. <https://doi.org/10.1029/93JA02867> (1994).
- Lakhina, G. S. & Tsurutani, B. T. Geomagnetic storms: Historical perspective to modern view. *Geosci. Lett.* <https://doi.org/10.1186/s40562-016-0037-4> (2016).
- Richardson, I. G. Solar wind stream interaction regions throughout the heliosphere. *Living Rev. Solar Phys.* <https://doi.org/10.1007/s41116-017-0011-z> (2018).
- Richardson, I. G., Cane, H. V. & Cliver, E. W. Sources of geomagnetic activity during nearly three solar cycles (1972–2000). *J. Geophys. Res.: Space Phys.* **107**, 81–813. <https://doi.org/10.1029/2001JA000504> (2002).
- Temmer, M. Space weather: The solar perspective: An update to schwenn (2006). *Living Rev. Solar Phys.* <https://doi.org/10.48550/arXiv.2104.04261> (2021).
- Khabarova, O. V. & Yermolaev, Y. I. Solar wind parameters' behavior before and after magnetic storms. *J. Atmos. Solar-Terrest. Phys.* **70**, 384–390. <https://doi.org/10.1016/j.jastp.2007.08.024> (2008).
- Gonzalez, W. D. & Tsurutani, B. T. Criteria of interplanetary parameters causing intense magnetic storms ($Dst < -100$ nT). *Planet. Space Sci.* **35**, 1101–1109. [https://doi.org/10.1016/0032-0633\(87\)90015-8](https://doi.org/10.1016/0032-0633(87)90015-8) (1987).

12. Výbošťoková, T. & Švanda, M. Statistical analysis of the correlation between anomalies in the Czech electric power grid and geomagnetic activity. *Space Weather* **17**, 1208–1218. <https://doi.org/10.1029/2019SW002181> (2019).
13. Švanda, M., Mourenas, D., Žertová, K. & Výbošťoková, T. Immediate and delayed responses of power lines and transformers in the Czech electric power grid to geomagnetic storms. *J. Space Weather Space Clim.* **10**, 26. <https://doi.org/10.1051/swsc/2020025> (2020).
14. Gil, A. et al. Review of geomagnetically induced current proxies in mid-latitude European countries. *Energies* **16**, 7406 (2023).
15. Rahmat, F. et al. Supervised feature selection using principal component analysis. *Knowl. Inf. Syst.* **66**, 1955–1995. <https://doi.org/10.1007/s10115-023-01993-5> (2024).
16. Taiwo, A. M. Source identification and apportionment of pollution sources to groundwater quality in major cities in southwest Nigeria. *Geofizika* **29**, 157–174 (2012).
17. Pearson, K. Li. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2**, 559–572. <https://doi.org/10.1080/14786440109462720> (1901).
18. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417–441. <https://doi.org/10.1037/h0071325> (1933).
19. Hotelling, H. Analysis of a complex of statistical variables with principal components. *J. Educ. Psy.* **24**, 498–520 (1933).
20. Awomeso, J. A., Ahmad, S. M. & Taiwo, A. M. Multivariate assessment of groundwater quality in the basement rocks of Osun State, Southwest, Nigeria. *Environ. Earth Sci.* **79**, 1–9. <https://doi.org/10.1007/s12665-020-8858-z> (2020).
21. Wang, X. D., Chen, R. C., Zeng, Z. Q., Hong, C. Q. & Yan, F. Robust dimension reduction for clustering with local adaptive learning. *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 657–669 (2018).
22. Herault, J. & Jutten, C. Space or time adaptive signal processing by neural network models. *AIP Conf. Proc.* **151**, 206–211. <https://doi.org/10.1063/1.36258> (1986) (**American Institute of Physics**).
23. Tharwat, A. Independent component analysis: An introduction. *Appl. Comput. Inform.* **17**, 222–249 (2021).
24. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605. <https://doi.org/10.1162/jmlr.2008.08226> (2008).
25. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint [arXiv:1802.03426](https://arxiv.org/abs/1802.03426) (2018).
26. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27. <https://doi.org/10.1007/BF02289565> (1964).
27. Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**, 115–129. <https://doi.org/10.1007/BF02289694> (1964).
28. Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**, 125–140. <https://doi.org/10.1007/BF02289630> (1962).
29. Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* **27**, 219–246. <https://doi.org/10.1007/BF02289621> (1962).
30. Torgerson, W. S. *Theory and Methods of Scaling* (Wiley, 1958).
31. Mishev, A. L. Application of the global neutron monitor network for assessment of spectra and anisotropy and the related terrestrial effects of strong seps. *J. Atmos. Solar-Terrest. Phys.* <https://doi.org/10.1016/j.jastp.2023.106021> (2023).
32. Simpson, J. A. The cosmic ray nucleonic component: the invention and scientific uses of the neutron monitor: keynote lecture. In *Cosmic Rays and Earth: Proceedings of an ISSI Workshop, 21–26 March 1999, Bern, Switzerland* (Springer, 2000). <https://doi.org/10.1023/A:1026567706183>.
33. Usoskin, I. G. A history of solar activity over millennia. *Living Rev. Solar Phys.* <https://doi.org/10.1007/s41116-023-00036-z> (2023).
34. Dorman, L. I. *Cosmic Rays in the Earth's Atmosphere and Underground* (Springer, 2004).
35. Boteler, D. H. & Pirjola, R. J. Numerical calculation of geoelectric fields that affect critical infrastructure. *Int. J. Geosci.* <https://doi.org/10.4236/ijg.2019.1010053> (2019).
36. Boteler, D. H., Pirjola, R. J. & Marti, L. Analytic calculation of geoelectric fields due to geomagnetic disturbances: A test case. *IEEE Access* **7**, 147029–147037. <https://doi.org/10.1109/ACCESS.2019.2945530> (2019).
37. Masahito, N., Masahisa, S., Toyohisa, K. & Toshihiko, I. Ae index. World Data Center Geomagn Kyoto (2017).
38. Nosé, M. et al. High-time resolution geomagnetic indices: Ae, asy, wp, and sym. <https://wcd.kugi.kyoto-u.ac.jp/wcd/pdf/> (2011)
39. Masahito, N., Masahisa, S., Toyohisa, K., Toshihiko, I. & Yukinobu, K. Geomagnetic dst index. Geomagnetic Dst index (2015).
40. Echer, E., Gonzalez, W. D. & Tsurutani, B. T. Interplanetary conditions leading to superintense geomagnetic storms ($Dst \leq -250$ nt) during solar cycle 23. *Geophys. Res. Lett.* <https://doi.org/10.1029/2007GL031755> (2008).
41. Jankowski, J. & Sucksdorff, C. *Guide for Magnetic Measurements and Observatory Practice* (International Association of Geomagnetism and Aeronomy Warsaw, 1996).
42. Matzka, J., Stolle, C., Yamazaki, Y., Bronkalla, O. & Morschhauser, A. The geomagnetic kp index and derived indices of geomagnetic activity. *Space weather* **19**, e2020SW002641. <https://doi.org/10.1029/2020SW002641> (2021).
43. Cousins, E. D. P., Matsuo, T. & Richmond, A. D. Mesoscale and large-scale variability in high-latitude ionospheric convection: Dominant modes and spatial/temporal coherence. *JGR Space Phys.* **118**, 7895–7904. <https://doi.org/10.1002/2013JA019319> (2013).
44. Kim, H. J., Lyons, L. R., Ruohoniemi, J. M., Frisell, N. A. & Baker, J. B. Principal component analysis of polar cap convection. *Geophys. Res. Lett.* <https://doi.org/10.1029/2012GL052083> (2012).
45. Shi, Y., Knipp, D. J., Matsuo, T., Kilcommons, L. & Anderson, B. Modes of (facs) variability and their hemispheric asymmetry revealed by inverse and assimilative analysis of iridium magnetometer data. *J. Geophys. Res.: Space Phys.* **125**, e2019JA027265. <https://doi.org/10.1029/2019JA027265> (2020).
46. Jolliffe, I. T. *Principal Component Analysis for Special Types of Data* (Springer, 2002).
47. Kilpua, E., Koskinen, H. E. J. & Pulkkinen, T. I. Coronal mass ejections and their sheath regions in interplanetary space. *Living Rev. Sol. Phys.* **14**, 5. <https://doi.org/10.1007/s41116-017-0009-6> (2017).
48. Torta, J. M. et al. Expected geomagnetically induced currents in the Spanish islands power transmission grids. *Space Weather* **21**, e2023SW003426. <https://doi.org/10.1029/2023SW003426> (2023).
49. Bailey, R. L. et al. Forecasting gics and geoelectric fields from solar wind data using lstms: Application in Austria. *Space Weather* **20**, e2021SW002907. <https://doi.org/10.1029/2021SW002907> (2022).
50. Výbošťoková, T. & Švanda, M. Correlation of anomaly rates in the slovak electric transmission grid with geomagnetic activity. *Adv. Space Res.* **70**, 3769–3780. <https://doi.org/10.1016/j.asr.2022.08.027> (2022).
51. Tozzi, R., De Michelis, P., Coco, I. & Giannattasio, F. A preliminary risk assessment of geomagnetically induced currents over the Italian territory. *Space Weather* **17**, 46–58. <https://doi.org/10.1029/2018SW002065> (2019).

Acknowledgements

Data of geomagnetic field components are from Belsk Observatory, a part of INTERMAGNET, (URL: <https://rt.bel.igf.edu.pl>). Solar, heliospheric, and geomagnetic data are from OMNI (URL: <https://omniweb.gsfc.nasa.gov>). Data of transmission lines anomalies are from the Polish Transmission System Operator.

Author contributions

All authors wrote the main manuscript text. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026