RESEARCH

Analysis of Geoeffective Impulsive Events on the Sun During the First Half of Solar Cycle 24

Agnieszka Gil[1,2] · Monika Berendt-Marchel[3] · Renata Modzelewska[1] ·
Agnieszka Siluszyk[1] · Marek Siluszyk[1,4] · Anna Wawrzaszek[2] · Anna Wawrzynczak[3,5]

Abstract

A coronal mass ejection (CME) is an impulsive event that emerges rapidly from the Sun. We observed a quiet Sun without many spectacular episodes during the last decade. Although some fast halo and partial halo CMEs had taken place, among them was the backside CME on 23 July 2012. In this work, we verify the link between the variability of solar-wind, heliospheric and geomagnetic parameters and the transmission grid failures registered in southern Poland during 2010 – 2014 when many geomagnetic storms appeared, caused by halo and partial halo CMEs. We aim to apply three machine learning methods: Principal Components Analysis, Self-Organizing Maps, and Hierarchical Agglomerative Clustering to analyze sources on the Sun and the impacts of the intense geomagnetic storms in the first half of Solar Cycle 24. The conducted analyzes underline the importance of solar-wind proton temperature and point out other solar-wind and geomagnetic parameters independently indicated by all the methods used in this study.

1. Introduction

The Sun's magnetic field leads to many effects that are collectively called solar activity, such as solar flares, solar energetic particle (SEP) events, Coronal Mass Ejecta (CMEs), coronal holes, etc. (e.g., Priest, 2001; Nitta et al., 2021).

✉ A. Gil
  gila@uph.edu.pl

[1] Siedlce University, Faculty of Exact and Natural Sciences, Institute of Mathematics, Konarskiego 2, 08-110 Siedlce, Poland

[2] Space Research Centre, Polish Academy of Sciences, Bartycka 18A, 00-716 Warsaw, Poland

[3] Siedlce University, Faculty of Exact and Natural Sciences, Institute of Computer Science, Konarskiego 2, 08-110 Siedlce, Poland

[4] Polish Air Force University, Dywizjonu 303 35, 08-521 Deblin, Poland

[5] National Centre for Nuclear Research, Soltana 7, 05-400 Otwock, Poland

A CME is a significant release of plasma and accompanying magnetic field from the Sun's corona into the solar wind. CMEs are often associated with solar flares and other forms of solar activity and are the subject of ongoing extensive studies (e.g., Papaioannou et al., 2016; Kilpua et al., 2021; Nitta et al., 2021; Asvestari et al., 2022; Rodriguez et al., 2022).

Hess and Zhang (2017) identified 70 Earth-affecting interplanetary CMEs during Solar Cycle 24 using in situ observations from NASA's Advanced Composition Explorer (ACE). They presented a statistical study of the properties of these events, including the source regions. In addition, the authors examined the characteristics of CMEs that are more likely to be highly geoeffective and examined the effect of the flare strength. They found that Earth-affecting CMEs in the first half of Cycle 24 were more likely to come from the northern hemisphere. After April 2012, this reversed, and these events were more likely to originate in the southern hemisphere, following the observed magnetic asymmetry in the two hemispheres.

Fortunately, the most intense CME of Solar Cycle 24, on 23 July 2012, with only 18.5 h Sun-to-Earth shock transit time, was the backside one. It originated in the active region of S17W141, 11520 NOAA. The shock speed exceeded 2000 $\mathrm{km\,s^{-1}}$, with the highest peak speed around 2600 $\mathrm{km\,s^{-1}}$ (e.g., Gopalswamy et al., 2016; Desai et al., 2020). Based on the plasma properties and measured magnetic field, it was assessed that this eruption would have caused a geomagnetic storm analogous to the Carrington storm (if it had been directed to the Earth). A couple of days before, the same active region was also a source of the relatively powerful halo CME that caused an intense geomagnetic storm (Gopalswamy et al., 2014, 2016; Gil et al., 2020a).

The CME link to geomagnetic storms stems from the southward component of the heliospheric magnetic field (HMF) contained in the CME flux ropes and the sheath between the flux rope and the CME-driven shock. A typical storm caused by the CME is characterized by high speed, sizeable angular width, and solar source location close to the central meridian. For CMEs originating at larger central meridian distances, the storms are mainly caused by the sheath field. Both the magnetic and energy contents of storm-producing CMEs can be traced to the magnetic structure of active regions and the free energy stored in them (Gopalswamy, 2009; Zhang et al., 2021; Palmerio et al., 2022).

CMEs impacting the Earth's magnetosphere very often significantly perturb the geomagnetic field, causing geomagnetic storms (e.g., Gopalswamy, 2009). A comprehensive review of extreme solar events and geomagnetic storms can be found in (Temmer, 2021; Zhang et al., 2021; Cliver et al., 2022). These events, rapidly changing geomagnetic fields and causing the geomagnetic storms, generate intense geomagnetically induced currents (GICs) (e.g., Nikitina, Trichtchenko, and Boteler, 2016; Oliveira and Ngwira, 2017).

GICs affecting the regular operation of electrical systems are one of the space-weather manifestations observed at the Earth's surface. During space-weather events, strong electric currents appear in the ionosphere, influencing the background of Earth's magnetic field. In addition, the triggered currents can emerge in the natural and technical conductors at the Earth's surface (e.g., Trichtchenko and Boteler, 2002; Viljanen et al., 2014). Electric transmission lines and buried pipelines are typical examples of such conductor systems. As a result, GICs can cause problems, such as increased corrosion of pipeline steel and damaged high-voltage power transformers. In addition, GICs can also affect geophysical exploration surveys, as well as oil and gas drilling operations (Oliveira and Ngwira, 2017). Substantial distortions in transmission line operation can be produced by GICs caused by rapid variations of the geomagnetic field, which are reflected by high values of its first derivative. The

detection of transmission line emission at large distances from a three-phase power line indicates its unbalanced operation (Veeramany et al., 2016; Fiedorov, Mazur, and Pilipenko, 2021; Pilipenko, 2021).

It was previously shown that GICs can affect transmission lines in mid-latitude countries, such as the Czech Republic (Svanda et al., 2020; Svanda, Smickova, and Vybostokova, 2021), Austria (Bailey et al., 2018, 2022; Albert et al., 2022), Italy (Tozzi et al., 2019), Greece (Zois, 2013), Spain (Torta et al., 2012), and even in low-latitude regions (Barbosa et al., 2015; Zhang et al., 2016).

In Eroshenko et al. (2010), the authors considered seventeen severe magnetic storms, which occurred in 2000–2005, and showed that anomalies appeared during each of these storms in the operation of the Russian railway system. It was linked to the main phase of the most substantial part of the geomagnetic storm. Schrijver and Mitchell (2013), quantifying impacts of the geomagnetic storms on US electric transmission lines in 1992–2010, found that $\sim 4\%$ of the disturbances in this power grid reported to the US Department of Energy were due to strong geomagnetic activity.

Our work primarily focuses on the search for statistical relationships between solar and heliospheric parameters, illustrating short-term, rapid changes on the Sun and their effects on ground-based technology in Poland during the period 01.01.2010–11.07.2014, when many intense geomagnetic storms appeared (Gil et al., 2021).

The article is organized as follows: In Section 1, we announce the background of our studies. In Section 2, we present the data used in the analysis, describing solar, interplanetary, and geomagnetic conditions. Sections 3 and 4 show the methodology and results obtained, respectively. In Section 5, we summarize our investigations.

2. Data

2.1. Analyzed Parameters

In this work, we analyzed solar, heliospheric, and geomagnetic parameters during the first half of Solar Cycle 24, precisely in the period 01.01.2010–11.07.2014. These time series illustrate the state of the Sun, its activity level, as well as its geoeffectiveness (e.g., Temmer, 2021; Zhang et al., 2021). Our attention was especially drawn to the moments of the halo and partial-halo fast CMEs (quite often being drivers of the large solar particle events, e.g., Gopalswamy et al., 2018) and geomagnetic storms caused by them. We considered heliospheric parameters: the strength of HMF B [nT] and its B_y and B_z [nT] components, electric field E_y [mV m^{-1}], solar wind speed SWs [km s^{-1}], proton density SWd [N cm^{-3}] and temperature SWT [K], as well as geomagnetic indices: Dst-index [nT], ap-index [nT], AE-index [nT], and local K-index from Belsk. All considered time series are averaged to three-hour periods to standardize the data time resolution.

Next, we used the local geomagnetic field data measured with a one-minute resolution in Belsk, the Polish INTERMAGNET observatory. More precisely, we focused on the horizontal components of the geomagnetic field B_X [nT] and B_Y [nT].

We started with the calculation of the first derivative of the geomagnetic field component dB_X/dt. Then, for each consecutive three-hour period, we found the maximal value of this derivative and called this parameter max(dB_X/dt). Values of max(dB_X/dt) allowed us to characterize the most decisive changes in the local magnetic field during each considered storm.

Additionally, we considered values of the geoelectric field at the Earth's surface (Ec [mV m^{-1}]). The estimation of Ec was performed by building a 1D conductivity model

Table 1 Number and percentage of electrical grid failure (EGF) causes for electrical transmission network in southern Poland during the whole period 01.2010 – 07.2014.

Cluster	Cluster name	Number of EGF	% of EGF
A	meteorological effects	142 384	29.0
B	operational shutdowns	205 175	41.7
C	vandalism	14 767	3.0
D	aging	49 937	10.2
E	electronic devices	11 709	2.4
F	unknown	67 730	13.8
A-F	Sum	491 702	100.00

(Ádám, Prácser, and Wesztergom, 2012; Gil et al., 2021), and then it was conjugated with the values of the geomagnetic field (B_X and B_Y). The multiplication of the Fourier transform of the magnetic field by the 1D transfer function was performed in the frequency domain. Then the application of the inverse Fourier transform of the output geomagnetic field allowed a return to the time domain (e.g., Boteler, Pirjola, and Marti, 2019).

Finally, we considered detailed data of electrical grid failures/anomalies (EGFs) in South Poland from 01.01.2010 to 11.07.2014 (details in Gil et al., 2020b). Based on transmission line operator logs, we have aggregated particular causes into six general clusters: associated with (A) meteorological effects, (B) operational shutdowns, (C) vandalism, linked to (D) the aging of infrastructure elements (Ag), (E) connected to the unreliability of electronic devices (Ed), and (F) having unknown reasons (Un). A similar categorization was defined by Zois (2013). The first three clusters (A-C), $\sim 73.7\%$ of all registered failures, might be considered objective reasons. The latter three (D-F) may be treated as failures having a solar origin (to some extent), and only these were further considered in the analysis of the influence of heliospheric parameters and geomagnetic indices separately for each type of failures/anomalies during each of the analyzed geomagnetic storms. The number and percentage of particular clusters of EGF during the whole studied period are shown in Table 1.

2.2. Data Processing

For data preparation, we used cross-correlation to measure similarity (e.g., Box, Jenkins, and Reinsel, 2008) between vectors representing solar-wind, heliospheric, and geomagnetic parameters, and vectors of transmission line failures in the three clusters of failures that could be linked to space-weather effects, i.e., failures caused by the aging of infrastructure elements, connected to the unreliability of electronic devices, and having unknown reasons.

Following Gil et al. (2021), we considered all intense geomagnetic storms during the studied period, that is, satisfying conditions: $B_z < -10$ nT and Dst < -100 nT for more than three hours (Gonzalez and Tsurutani, 1987). In the period considered, 01.2010 – 07.2014, 8 events with Dst < -100 nT with duration longer than three hours, 16 events with $B_z < -10$ nT with duration longer than three hours, and 5 events that met both conditions, that is, $B_z < -10$ nT and Dst < -100 nT during more than three hours (compare with Table 1 in Gil et al., 2021), were registered. It is worth underlining that all these storms were associated with the halo or partial halo CME (from cdaw.gsfc.nasa.gov/CME_list/).

We have performed our calculations for each intense geomagnetic storm, with data including 3 days before and 5 days after the geomagnetic storm, with analyzed sets always containing 64 of 3-hour data. The cross-correlation of vectors representing solar wind, heliospheric, and geomagnetic parameters (denoted in a formula by Y^j, $j \in \{1, \ldots, 13\}$) and
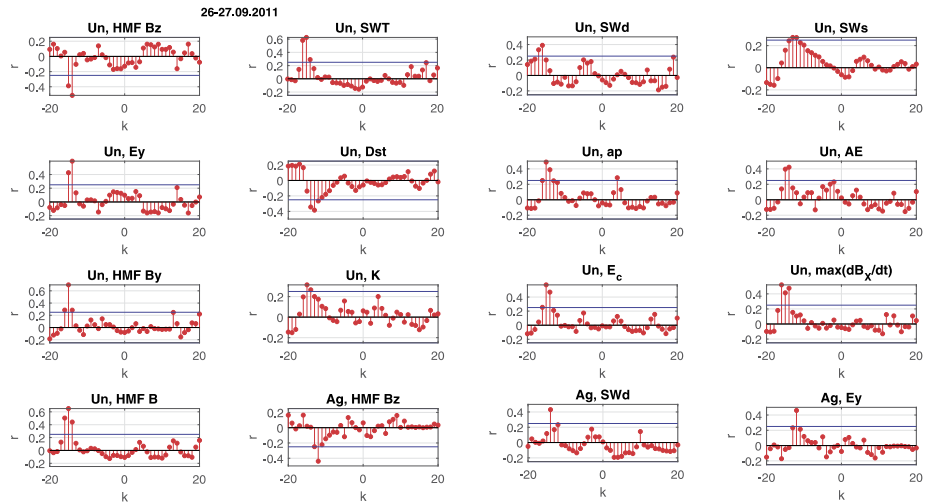
Figure 1 Cross-correlation of electrical grid failures (EGFs) and solar, heliospheric, and geomagnetic parameters as a function of the lag for the geomagnetic storm 26–27.09.2011. On the abscissa, there are lags, k, and on the ordinate, sample cross-correlation values, r. The blue lines denote 95% confidence bounds.

vectors of transmission line failures in the three clusters of failures that could be linked to space weather effects, i.e., failures caused by the aging of infrastructure elements, connected to the unreliability of electronic devices, and having unknown reasons (denoted in a formula by F^i, $i \in \{1, 2, 3\}$), is a function of the lag k. For pairs (F_1^j, Y_1^j), (F_2^j, Y_2^j), ..., (F_T^j, Y_T^j) and lags $k = 0, \pm1, \pm2, \ldots$, we evaluate the lag k cross-covariance as follows (Box, Jenkins, and Reinsel, 2008):

$$c_{F^i Y^j}(k) = \frac{1}{T} \begin{cases} \sum_{t=1}^{T-k}(F_t^i - \bar{F}^i)(Y_{t+k}^j - \bar{Y}^j), & k = 0, 1, 2, \ldots \\ \sum_{t=1}^{T+k}(Y_t^j - \bar{Y}^j)(F_{t-k}^i - \bar{F}^i), & k = -1, -2, \ldots \end{cases} \tag{1}$$

where $\bar{F}^i$, $\bar{Y}^j$ indicate the sample means of the considered time series. The sample standard deviations are $\sigma_{F^i} = \sqrt{c_{F^i F^i}(0)}$ and $\sigma_{Y^j} = \sqrt{c_{Y^j Y^j}(0)}$. Hence, the estimate of the cross-correlation r can be expressed in the following way (Box, Jenkins, and Reinsel, 2008):

$$r_{F^i Y^j} = \frac{c_{F^i Y^j}(k)}{\sigma_{F^i} \sigma_{Y^j}} \tag{2}$$

with $k = 0, \pm1, \pm2, \ldots$. Thus, 39 plots were generated for each of them that present cross-correlations values, r, and lags, k (in units of three hours). The sign of lag k is negative if the variability of failures/anomalies is preceded by changes in heliospheric and geomagnetic parameters, and positive in the opposite situation. Here, as an example, we present results obtained for the intense geomagnetic storm 26–27.09.2011 in Figure 1.

The computed cross-correlations for the 19 geomagnetic storms showed that their values vary between -0.63 during the storm on 17.06.2012, for the parameter E_y and EGF of Ed type, and 0.88 during the geomagnetic storm on 25.10.2011, for parameter $\max(dB_X/dt)$ and EGF of Ed type.

Our analysis revealed that not for all storms and considered time series, cross-correlations give the results with 95% confidence bound. Furthermore, some geomagnetic storms (Sec-
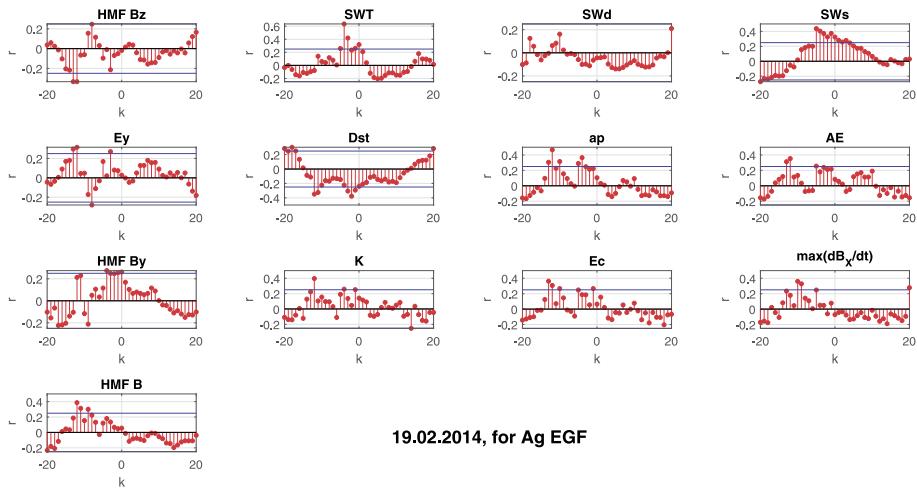
Figure 2 Cross-correlation of electrical grid failures (EGFs) caused by the aging of infrastructure elements (Ag) and solar-wind, heliospheric, and geomagnetic parameters as a function of the delay for the geomagnetic storm 19.02.2014. The blue lines denote 95% confidence bounds.

tion 2) showed that failures/anomalies sometimes occurred ahead of the intense geomagnetic storm (lag k values were positive). Thus, for a more detailed analysis, we took into account only these storms with cross-correlations r that were mostly above 95% confidence: (I) 09.03.2012, (II) 15 – 16.07.2012, (III) 14.11.2012, (IV) 01.06.2013, and (V) 19.02.2014. For each of these storms, most of the lags k between electrical grid failures and solar, heliospheric, and geomagnetic parameters were negative. Detailed results of the cross-correlations for these five geomagnetic storms are presented in Tables 2 and 3. The obtained lag k was selected as the one for which the extreme cross-correlation value appeared, regardless of whether its sign was positive or negative. Hence, it is worth underlining that all the computed lags were taken into further calculations exactly as they were determined. As an illustration, we show results for geomagnetic storm on 19.02.2014 in Figures 2 – 4. Tables 2 and 3 and Figures 2 – 4 display that there are cases when this correlation between EGFs and solar-wind, heliospheric, and geomagnetic parameters is rather small, although mostly reaching the 95% confidence bounds. However, quite often, the absolute value of the correlation is above 0.50. For this particular storm, the highest r is at the level of 0.63 between SWT & EGF of Ag and Ed types. The length of the lags is changeable, both due to successive geomagnetic storms and due to various parameters. There is no regularity in the delay length. However, the lags for the geomagnetic indices are pretty similar. The same situation is found in the case of solar-wind parameters.

Based on the lags listed in Tables 2 and 3, each of the solar, heliospheric, and geomagnetic parameters has been shifted accordingly. Data processed this way has been used in the context of each method described in more detail in the article.

3. Methods

3.1. Principal Component Analysis

The essence of Principal Component Analysis (PCA) is a reduction of the n-dimensional data space of the variables $X \subseteq R^n$ to a m-dimensional one, with uncorrelated factors

Table 2 Results of cross-correlations: lag k (the unit of lag equals three hours) and correlation r for five geomagnetic storms: **I**: 09.03.2012, **II**: 15 – 16.07.2012, **III**: 14.11.2012, **IV**: 01.06.2013, **V**: 19.02.2014 between three groups of transmission line failures: caused by aging of infrastructure elements (*Ag*), connected to the unreliability of electronic devices (*Ed*), and having unknown reasons (*Un*), and all considered in the article heliospheric and geomagnetic parameters (actually, the first seven parameters). The analyzed period is from 01.01.2010 to 11.07.2014. Light gray text indicates positive lags, while results that are not statistically significant are in italics.

| Storm	EGF	B		Bz		By		SWT		SWd		SWs		Ey	
		k	r	k	r	k	r	k	r	k	r	k	r	k	r
I	Ag	−5	−0.26	−18	−0.36	−16	*0.21*	10	0.42	6	*0.44*	−17	0.27	−18	0.43
	Ed	−20	0.39	−18	−0.42	−20	−0.48	−20	0.55	8	*0.46*	−15	*0.23*	−18	0.51
	Un	−19	0.33	−17	−0.43	−19	−0.38	−19	0.41	9	*0.45*	−14	*0.23*	−17	0.48
II	Ag	−20	0.29	−18	−0.31	−13	−0.36	−8	0.42	−16	*−0.17*	−6	*0.23*	−18	0.28
	Ed	−16	0.34	−16	−0.31	−13	0.33	−1	0.34	−18	0.49	−19	0.3	−16	0.36
	Un	−7	0.28	−6	−0.3	−8	0.32	18	0.26	−14	0.33	−18	−0.27	−7	0.31
III	Ag	−20	0.32	−19	−0.55	−17	−0.26	−16	0.62	−6	*−0.19*	−10	0.29	−19	0.56
	Ed	−10	0.31	−10	−0.41	−14	0.41	−7	0.53	−17	0.38	−1	*−0.22*	−8	0.42
	Un	−6	*0.21*	−1	−0.37	−6	0.35	1	0.4	−9	0.34	−1	*0.20*	−1	−0.36
IV	Ag	−18	0.4	−18	−0.6	−13	0.33	−12	0.36	−5	0.36	0	0.25	−18	0.56
	Ed	−5	0.59	−5	−0.56	0	0.48	1	0.5	−6	0.62	−12	−0.30	−5	0.52
	Un	−20	0.27	−19	−0.29	−16	0.29	−15	0.26	−4	*0.15*	−10	*0.14*	−19	0.30
V	Ag	−12	0.39	−12	−0.34	−4	0.27	−4	0.63	−10	*0.16*	−5	0.44	−12	0.31
	Ed	−8	0.38	−9	−0.31	−8	0.29	0	0.63	−6	*0.20*	−1	0.38	−9	0.30
	Un	−8	0.25	−18	0.26	−17	−0.27	0	0.54	−20	0.24	−1	0.28	−4	−0.24
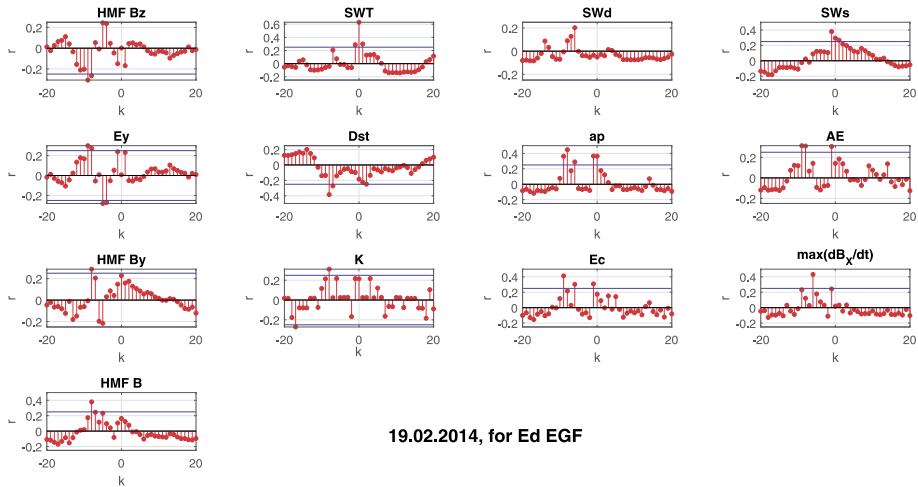
Figure 3 Cross-correlation of electrical grid failures (EGFs) connected to the unreliability of electronic devices (Ed) and solar-wind, heliospheric and geomagnetic parameters as a function of the delay for the geomagnetic storm 19.02.2014. The blue lines denote 95% confidence bounds.
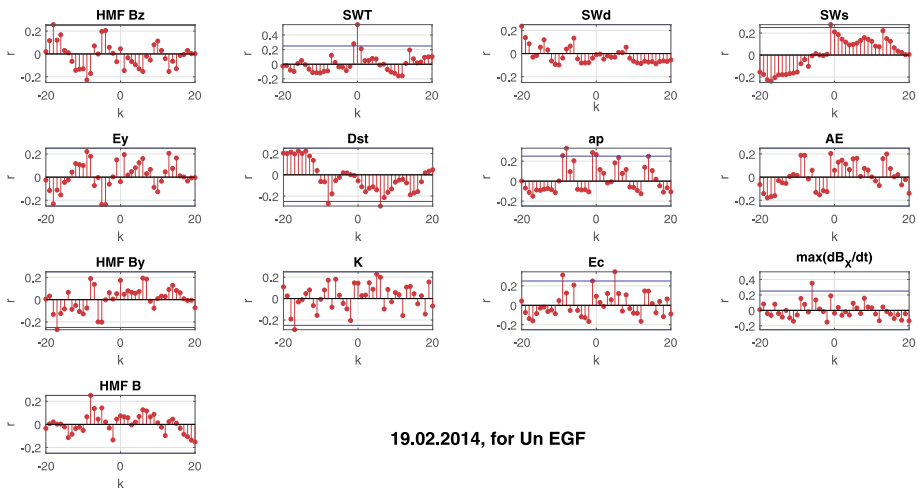

Figure 4 Cross-correlation of electrical grid failures (EGFs) having unknown reasons (Un) and solar-wind, heliospheric and geomagnetic parameters as a function of the lag for the geomagnetic storm 19.02.2014. The blue lines denote 95% confidence bounds.

$Z \subseteq R^m$, where $m < n$. Moreover, using PCA, one can search for possible regularities and dependencies between the investigated variables without prior knowledge of the data. Let $X = (X_1, \ldots, X_n)$ be a vector with known variances for which the correlation or covariance matrix is also known. Using the properties of eigenvectors and eigenvalues, we want to find such new variables Z_i, $(i = 1, \ldots, m)$ that are a linear combination of the old variables X_p, $(p = 1, \ldots, n)$ and will transfer information from these old variables as well as possible (Jolliffe, 2002). In what follows, $\alpha_i \cdot X^T = \sum_{p=1}^{n} a_{ip} \cdot X_p$, where $\alpha_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ and

Table 3 The same as Table 2 for the next six parameters.

Storm	EGF	Dst		Ap		AE		K		Ec		max(dB$_X$/dt)	
		k	r	k	r	k	r	k	r	k	r	k	r
I	Ag	−17	−0.37	−17	0.49	−17	0.37	−9	−0.25	−17	0.45	8	0.52
	Ed	−15	−0.33	−17	0.59	−16	0.41	−15	0.24	−17	0.56	8	0.34
	Un	−16	−0.33	−16	0.53	−16	0.34	−16	0.24	−16	0.48	9	0.45
II	Ag	−17	−0.28	−19	0.33	−9	0.22	−20	0.30	−20	0.46	−8	0.17
	Ed	−14	−0.26	−13	0.41	−2	0.22	−13	0.3	−13	0.47	−20	0.28
	Un	3	−0.32	−2	0.32	−19	−0.26	−2	0.37	6	0.39	−2	0.63
III	Ag	−18	−0.47	−19	0.55	−19	0.53	−19	0.36	−19	0.5	−19	0.36
	Ed	−8	−0.45	−9	0.62	−10	0.55	−10	0.37	−10	0.61	−19	0.66
	Un	0	−0.25	−1	0.44	−2	0.39	7	−0.27	−2	0.47	−11	0.47
IV	Ag	−17	−0.42	−18	0.56	−18	0.47	−16	0.27	−18	0.45	−7	0.42
	Ed	−3	−0.55	−4	0.61	−5	0.49	−4	0.44	−4	0.56	−5	0.45
	Un	−19	−0.31	−19	0.36	−19	0.24	−19	0.22	−19	0.26	−20	0.26
V	Ag	−2	−0.38	−12	0.47	−12	0.35	−12	0.40	−13	0.36	−10	0.36
	Ed	−8	−0.38	−8	0.45	−9	0.31	−8	0.31	−9	0.41	−6	0.43
	Un	−8	−0.27	−8	0.33	−1	0.20	−17	−0.29	−9	0.31	−6	0.35

$\sum_{p=1}^{n} a_{ip}^2 = 1$ for $i = 1, \ldots, m$ thus $Z_i = \alpha_i \cdot X^T$ denotes the ith principal component. In this way, the above problem comes down to the determination of the a_{ip} values of the α_i vector such that the variance Var(Z_i) will be as large as possible. The coefficients α_i can be obtained from the matrix equation $(S - \lambda I) \alpha_i = 0$, where S and I are the covariance matrix and identity matrix, respectively. Therefore, the solution $\alpha_i \neq 0$ is the eigenvector corresponding to the eigenvalue λ_i. Moreover, $\lambda_i = \text{Var}(Z_i)$. When 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. The principle of creating components is that the next components are uncorrelated with each other and maximize the variability not explained by the previous component (Jolliffe, 2002). For this purpose, the PCA method orders the vectors according to the eigenvalues from the largest to the smallest. Thus it eliminates the variables with the smallest eigenvalues. This is known as the scree test (Cattell criterion), which has a geometric interpretation for the eigenvalues and the percentage of clarifying variables (Jolliffe, 2002). If the eigenvalue is very small, the variance is also very small, so the data are focused around the straight line which contains the eigenvector. In conclusion, the feature adds little information to the new set of components. In other words, these eigenvalues from maximal to minimal give us the order of the principal components, while the sum of variances of the variables Z_i is equal to the sum of the variances of the original variables X_p. Therefore, the transformation of the variables does not lead to the loss of information about the processes studied. Before using the PCA method, we check two assumptions: Bartlett's test and Kaiser-Mayer-Olkin (KMO) coefficient (Jolliffe, 2002). Bartlett's test is applied to prove the hypothesis $H_0 : R = 1$, which means that correlation coefficients $r_{ps} = 0$ for $p \neq s$ whereas $r_{ps} = 1$ for $p = s$, $p, s = 1, \ldots, n$. The

KMO coefficient is given by the formula

$$\text{KMO} = \frac{\sum_{p \neq s} \sum_{p \neq s} r_{ps}^2}{\sum_{p \neq s} \sum_{p \neq s} r_{ps}^2 + \sum_{p \neq s} \sum_{p \neq s} \hat{r}_{ps}^2}, \tag{3}$$

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

$$\hat{r}_{ps} = -\frac{c_{ps}}{\sqrt{c_{pp} c_{ss}}}. \tag{4}$$

Here c_{ps}, c_{pp}, c_{ss} are the algebraic complement of r_{ps}, r_{pp} and r_{ss} element, respectively. The condition is $0.5 < \text{KMO} < 1$, that is, $\sum_{i \neq j} \sum_{i \neq j} \hat{r}_{ps}^2 < \sum_{p \neq s} \sum_{p \neq s} r_{ps}^2$.

3.2. Self-Organizing Maps

Self-Organizing Maps (SOM) are a type of artificial neural network. In SOM, each input layer neuron is connected to each output layer neuron. This structure allows to project a multidimensional space of input data (a single vector of input data is treated as one dimension) on the most often two-dimensional matrix of neurons, the so-called map of neurons (Kohonen, 1990). Neuronal maps are presented in the form of a hexagonal or rectangular grid, depending on the selected network topology. A hexagonal grid is most often used because it does not favor any direction, and the distance between neighboring neurons is equal. The connections between adjacent neurons are assigned a certain weight. In a single SOM training iteration, not only the weights of the winning neuron but also the weights of its closest neighborhood are updated based on the selected neighborhood function and the size of that neighborhood (Hu et al., 2019). Weights are updated based on Kohonen's rule:

$$w_{ij}(t + 1) = w_{ij}(t) + \beta(t) h_{ci}(t)[x(t) - w_{ij}(t)], \tag{5}$$

where w_{ij} is the weight j of the neuron i, $\beta(t)$ is the learning rate, $h_{ci}(t)$ is the neighborhood function (Caldas et al., 2017) that determines the neighborhood of the winner neuron, and $x(t)$ is the input vector. The distance between the given neurons on the map reflects the similarity between the input elements. This allows us to find correlations between the input data. SOMs are most commonly used for data clustering and classification problems since SOMs aggregate similar data into clusters (Caldas et al., 2017; Lampinen and Oja, 1992). In our case, SOM was used to group selected heliospheric and geomagnetic parameters with electric grid failures.

3.3. Hierarchical Agglomerative Clustering

Cluster analysis is a broad class of data-mining methods in which data sets placed in the same group (or cluster) are more similar to one another than to those objects placed in other groups. Classification algorithms can be divided into two central models: (1) supervised and (2) unsupervised learning. Hierarchical agglomerative clustering (HAC) (e.g., Müllner, 2013) is an unsupervised learning method similar to SOM. This type of method is less biased and can adapt to unique situations because the resultant clusters are based on models that have not been previously trained. HAC methods work by grouping objects from the bottom up. Each data entity starts as its own 'cluster', and clusters are merged based on similarities until a significantly reduced number of groups are presented as a final solution. Ward's method (Ward, 1963) for clustering is among the most popular approaches for HAC. It is

the only method among the agglomerative clustering methods that is based on a classical sum-of-squares criterion, producing groups that minimize within-group dispersion at each binary fusion. In addition, Ward's approach is useful because it looks for clusters in multivariate Euclidean space. That is also the reference space in multivariate ordination methods, particularly in PCA.

Ward's linkage uses the incremental sum of squares, the increase in the total within-cluster sum of squares due to joining two clusters. The within-cluster sum of squares is defined as the sum of the squares of the distances between all objects in the cluster and the cluster's centroid. The sum of squares metric is equivalent to the following distance function:

$$d(r, s) = \sqrt{\frac{2n_r n_s}{(n_r + n_s)}} \|\overline{x_r} - \overline{x_s}\|_2, \tag{6}$$

where

- $\| \cdot \|_2$ is the Euclidean distance,
- $\overline{x_r}$ and $\overline{x_r}$ are the centroids of clusters s and r,
- n_r and n_s are the numbers of elements in clusters r and s.

A HAC result is typically visualized as a dendrogram. A horizontal line represents each merge. The y-coordinate of the horizontal line is the similarity of the merged clusters, where data series are viewed as singleton clusters. This similarity is called the combination similarity of the merged cluster. A dendrogram allows us to reconstruct the history of merges that resulted in the depicted clustering by moving up from the bottom layer to the top node. A fundamental assumption in HAC is that the merging operation is monotonic. Monotonic means that if $s_1, s_2, ..., s_{K-1}$ are the combination similarities of the successive merges of a HAC, then $s_1 \geq s_2 \geq \cdots \geq s_{K-1}$ holds.

4. Results and Discussion

4.1. Principal Component Analysis

We applied the PCA technique described in Section 3.1. We were looking for the family of each type of failure: Ag, Ed, or Un and 13 parameters describing the state of heliosphere near the Earth's vicinity for 5 distinguished geomagnetic storms (Section 2.2) trying to indicate the principal components for each type of failure.

Under the required assumptions, that is, Bartlett's test and the KMO coefficient (see Section 3.1), we present the results of the PCA analysis for failures caused by the aging of infrastructure elements (Ag) (Table 4). During our calculations, we obtained the set of eigenvalues from 3 to 6 when the eigenvalue was more than 1, or its variance was more than 5%. In the case of EGF of Ag type, we get the minimal cumulative percentage of variance equal to 80.91% for 3 eigenvalues obtained for the 14.11.2012 storm and the maximal cumulative percentage of variance equals 86.14% for 6 eigenvalues obtained for the 19.02.2014 storm (not shown here). Calculations for EGF of Ed-type got us 76.44% of cumulative variance for 4 eigenvalues for the 19.02.2014 storm, which was the minimum value, and 87.57% for 5 eigenvalues obtained on $15-16.07.2012$ as the maximal cumulative variance. For EGF of Un-type, our results are the following: 84.73% cumulative variance for 5 eigenvalues on 1.06.2013 and cumulative variance equals 85.08% for 5 eigenvalues obtained

Table 4 Results of PCA analysis. The 1st principal component consists of *Ag* and 13 parameters that describe the heliosphere. % of variance contributed to the first principal component during 5 geomagnetic storms **I**: 09.03.2012, **II**: 15 – 16.07.2012, **III**: 14.11.2012, **IV**: 01.06.2013, **V**: 19.02.2014.

1st p.c. I	Var. %	1st p.c. II	Var. %	1st p.c. III	Var. %	1st p.c. IV	Var. %	1st p.c. V	Var. %
ap	85.6	Dst	90.6	AE	92.4	ap	92.3	ap	78.2
AE	73.4	B	84.2	ap	86.4	Ec	84.1	AE	73.2
Ec	70.6	Bz	81.8	Ec	85.4	AE	79.6	K	59.2
Ey	62.0	Ey	77.5	SWT	80.4	B	76.1	SWs	55.4
Dst	58.4	Ec	75.1	K	72.3	Bz	69.9	Ec	47.0
Bz	49.0	ap	74.4	Dst	68.1	Dst	68.8	Ag	45.7
Ag	38.8	K	66.2	Bz	62.7	Ey	61.3	B	40.3
SWs	31.3	AE	26.8	B	59.8	K	52.8	Bz	34.6
dBx	14.7	Ag	19.7	Ey	59.7	By	48.5	Ey	34.5
SWd	13.7	By	17.4	dBx	47.9	Ag	32.2	Dst	25.7
SWT	11.5	dBx	15.3	Ag	40.7	SWT	27.4	SWT	23.7
By	10.5	SWd	15.0	SWs	39.5	SWs	17.5	By	18.6
K	6.1	SWT	12.0	SWd	21.1	SWd	6.3	SWd	10.6
B	3.4	SWs	0.0	By	1.5	dBx	2.1	dBx	8.3

on 14.11.2012. For other cases, we have obtained the cumulative percentage of variance between these values. In Table 4, we present the 1st principal components containing *Ag* and 13 parameters describing the state of the heliosphere. Furthermore, we show that their percentage of variance contributed to the first principal component during 5 geomagnetic storms. For example, during the geomagnetic storm on 9.03.2012, we can see that 85.6% of variance of ap, 73.4% of variance of AE, and 70.6% of variance of *Ec*, etc., are explained by the 1st principal component (see, the first and second columns of the Table 4). Continuing, 89.5% of the variance of ap, 81.6% of AE, and 78.9% of *Ec*, etc., is explained by the 2nd principal component obtained during the same geomagnetic storm. To be more precise, we need to take the 3rd, 4th, and 5th principal components so that the contribution in the transfer of information from each parameter is at least 80%. Similar results we obtained for other geomagnetic storms except for $15 – 16.07.2012$, when we got set of Dst-index, B, B_z and E_y (see Table 4). On the other hand, PCA analysis revealed that ap, AE, and *Ec* had the largest values of the factors coordinates of variables, and it means that ap, AE, and *Ec* are the parameters grouped together in the first principal components of considered geomagnetic storms, except $15 – 16.07.2012$. We present the arrangement of parameters considered in the PCA diagrams for the aging type of failures (*Ag*), connected to the unreliability of electronic devices (*Ed*), and having unknown reasons (*Un*) (Figures 5 – 7).

Figures 5 – 7 present only two principal components that have the largest cumulative variance (the projection on the OZ_1Z_2 plane). However, it is worth underlying that usually two principal components are enough to determine the space described by new variables (more than 50%); otherwise, it does not disqualify the other components. The analysis of variance during geomagnetic storms on 9.03.2012 gave us for *Ag* and the other heliospheric and geomagnetic parameters 5 principal components with cumulative variance 83%, on $15 – 16.07.2012$ 5 eigenvalues and 85.4%, 14.11.2012 3 eigenvalues with 80.9%, 1.06.2013 4 eigenvalues with 82.6%, and on 19.02.2014 gave us up to 6 principal components and 81.1% of the total variance. The application of PCA for the other geomagnetic storms re-

Figure 5 The projection of variables in the plane of two factors with the highest variance and with the EGFs of aging-type failure (Ag) on 19.02.2014.
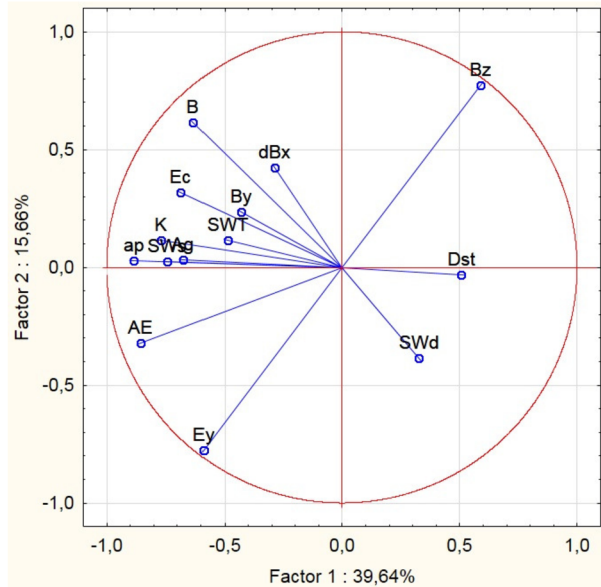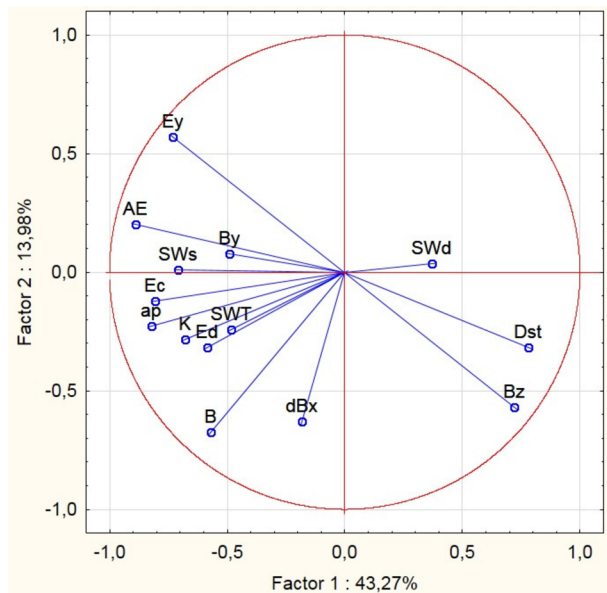

Figure 6 The projection of variables in the plane of two factors with the highest variance and with the EGFs of electronic devices type of failure (Ed) on 19.02.2014.

vealed a change of the basic set of components. In the projection of Ag, Ed, Un, and the heliospheric and geomagnetic parameters in the plane of two factors with the highest variance (Figures 5 – 7), the observations concern the factor coordinates of the variables corresponding to the correlations between these parameters and the variables of the main components. It suffices to show that the longer the eigenvector of a particular variable is, the higher is the correlation of a given variable with a given main component. Taking into account the first component and EGF of Ag type, mentioned eigenvector of Ag is long, medium,

Figure 7 The projection of variables in the plane of two factors with the highest variance and with the failure of unidentified type of failure (Un) on 19.02.2014.
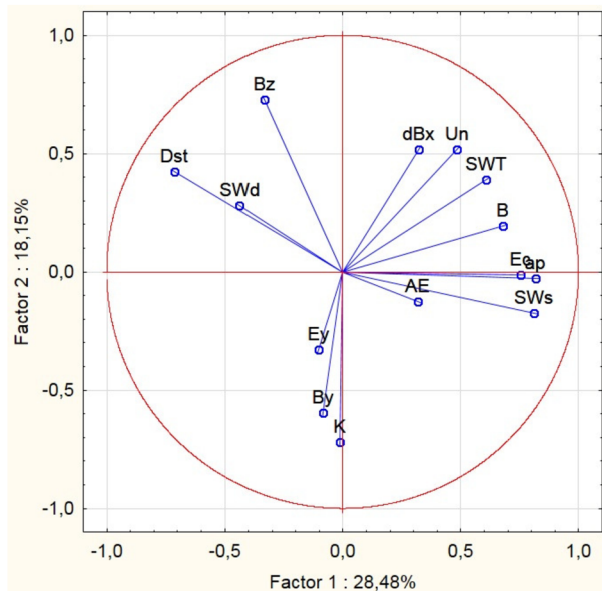
medium, long, medium for geomagnetic storms on 9.03.2012, 15 − 16.07.2012, 14.11.2012, 1.06.2013, and 19.02.2014 (see Figure 5), respectively. The eigenvectors for Ed and Un are not so long for storm on 14.11.2012 (see, Figures 5 − 7), nor do they satisfy KMO for storms 9.03.2012 (Ed), 15 − 16.07.2012 (Un), 1.06.2013 (Ed), 19.02.2014 (Un). In Figures 5 − 7, quite compact clusters of geomagnetic storm parameters can be distinguished. Our next objective is to note that both Ag and the eigenvectors of the other parameters are situated in the second or third quadrant of the coordinate system, contrary to B_z, Dst-index, SWd. For Ed, we see a similar arrangement as for Ag. Figures 5 − 7 show that on the PCA map all types of failures, Ag (Figure 5), Ed (Figure 6), and Un (Figure 7), are closely connected to solar wind protons temperature.

4.2. Self-Organizing Maps Analysis

Models built based on neural networks are statistical and nondeterministic. They approximate the conditional mean value of the modeled input variables on the outputs. Moreover, the approximation process depends on several random elements, e.g., selecting weights of connections between neurons. So, in Self-Organizing Maps, maps with the same topology, the same training model parameters as the neighborhood function, or the number of iterations with the exact input data representation may show different results. However, in many cases, the conclusions of these seemingly different maps may still be similar (Lippmann, 1987; Wehrens, 2009). For this reason, the SOM analysis was performed 50 times for each combination of a given geomagnetic storm and a given type of electric grid failure to check the percentage of the number of results in which the heliospheric and geomagnetic parameters are grouped with a given failure in a given storm. Data sets with a 3-hour resolution were standardized from 01.01.2010 to 11.07.2014. The model parameters were a 4×4 neuron grid, 100 training iterations, and a link distance function as a neighborhood distance function.

The SOM analysis showed that different heliospheric and geomagnetic data are grouped into the same cluster with a given type of failure in each considered storm. However, despite

Figure 8 SOM neighbor weight distances for the geomagnetic storm, 19.02.2014 for electric grid failures caused by the aging of infrastructure elements (Ag). The blue hexagons represent neurons, and the red lines show which particular neurons are connected. The colors from black to yellow display the weight values of the connection between neighboring neurons.
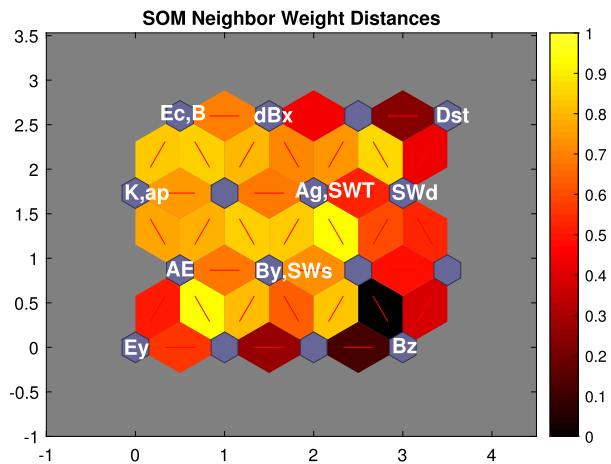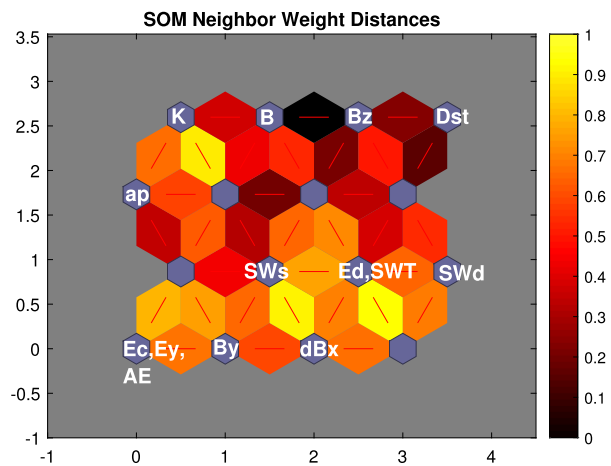

Figure 9 SOM neighbor weight distances for the geomagnetic storm, 19.02.2014 for electric grid failures connected to the unreliability of electronic devices (Ed). The blue hexagons represent neurons, and the red lines show which particular neurons are connected. The colors from black to yellow display the weight values of the connection between neighboring neurons.

obtaining different results, it is possible to notice a group of parameters that correlate with a given type of failure in most analyzed cases.

Example results of our analysis are presented in Figures 8 – 10 for the geomagnetic storm on 19.02.2014, for the three groups of electric grid failures: caused by the aging of infrastructure elements (Ag), connected to the unreliability of electronic devices (Ed), and having unknown reasons (Un). Figures 8 – 10 show the SOM neighbor weight distances, which describe the weight assigned to the connections between neighboring neurons, labeled blue hexagons. The red lines show how the neurons are connected. The colors in the areas with the red lines represent the connection's weight between neurons. The brighter the color, the shorter the distance between neurons is. Yellow represents a weight equal to 1, which means the shortest distance between neighboring neurons and a grouping into the cluster of the parameters from the studied space. The darker the color (black means weight equal to 0), the longer the distance between neurons. This means that the parameters grouped in such neurons are not correlated. Our analysis showed that the Ag type of failure in all analyzed geomagnetic storms is often grouped – over 60% of the results, with the SWd proton density. For example, in the storm on 09.03.2012 and 15 – 16.07.2012, the SWd is grouped

Figure 10 SOM neighbor weight distances for the geomagnetic storm, 19.02.2014 for electric grid failures having unknown reasons (Un). The blue hexagons represent neurons, and the red lines show which particular neurons are connected. The colors from black to yellow display the weight values of the connection between neighboring neurons.
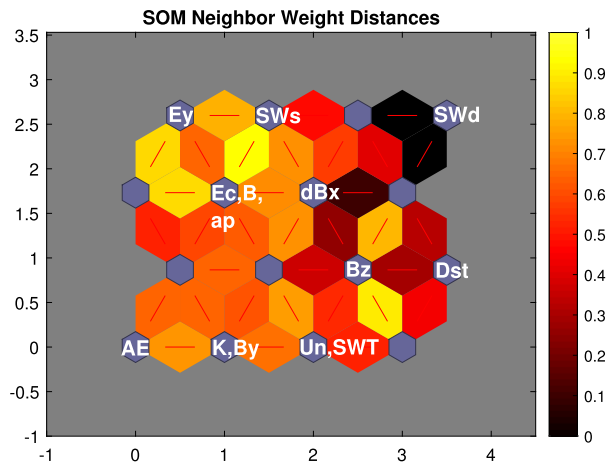
with an Ag type of failure in more than 95% of the obtained results. Also, other solar-wind parameters, such as the speed SWs and the temperature SWT of protons, are grouped into more than 80% of the results with failure related to aging of infrastructure elements in storms 15.07.2012, 19.02.2014, and 14.11.2012. A similar tendency is also visible in all storms, excluding the storm on 14.11.2012 in the B_y component of HMF (in a cluster with a Ag type of failure in more than 80% of the obtained results). Furthermore, the following parameters (often more than 70% of obtained results) correlate with the Ag-type failure: K, Ec, ap and AE.

The same situation can be observed for the solar-wind parameters: SWd, SWs, and SWT can be noticed in the case of the failure Un. Also, with this type of failure, in more than 60% of obtained results, the HMF component B_y is grouped with a failure, except for the storm on 19.02.2014 and on 09.03.2012. In more than 60% of the results, in at least four analyzed storms in the same cluster with a failure having an unknown reason, a computed geoelectric field Ec, ap- and AE-index are also grouped.

Slightly different dependencies can be observed in the analyzed storms between the solar-wind parameters and EGFs connected to the unreliability of electronic devices. However, these parameters are grouped into the same cluster in more than 65% of the results in the storms on $15 – 16.07.2012$, 19.02.2014, and 14.11.2012. Proton density SWd and B_y components of HMF are also grouped with this failure in more than 60% of the results in all considered storms.

The results of the SOM analysis revealed that the highest impact on each of the considered electric grid failures (Ag, Ed, Un), regardless of the geomagnetic storm, seems to have one of the solar-wind parameters, the proton density SWd. In the case of storms on $15 – 16.07.2012$, 14.11.2012, and 19.02.2014, it is also possible to indicate the correlation between all analyzed failures and other SW parameters – temperature SWT and speed SWs. Moreover, depending on the storm with a given type of failure, different heliospheric parameters and geomagnetic indices may group in the same cluster, for example, in the storm on 19.02.2014, together with the Ag type of failure parameters: K, Ec, B, B_y, ap and AE are grouped in more than 70% of the results, while in the storm on 09.03.2012, only the parameter B_y correlates with the failure of the Ag type – more than 70% of the results. All obtained results are given in Table 5. Hence, it is difficult to generalize and evidently list a specific set of parameters that could affect the occurrence of a given type of electric grid failure during more than one storm.

Table 5 Results in percentages obtained from all the SOM analyses for all considered geomagnetic storms and failures caused by the aging of infrastructure elements (Ag), connected to the reliability of electronic devices (Ed), and having unknown reasons (Un). Green denotes combinations in which the given heliospheric or geomagnetic parameter was grouped with the given electric grid failures in more than 60% of conducted experiments.

EGF	Storm	K	Ec	dBx	B	By	Bz	SWT	SWd	SWs	Ey	Dst	ap	AE
Ag	I	68	68	52	64	84	56	42	98	76	54	20	60	68
	II	58	54	30	54	98	16	98	96	94	48	14	52	28
	III	82	82	84	42	22	28	82	52	88	82	40	82	82
	IV	70	74	60	74	80	44	28	62	34	74	38	72	74
	V	80	78	56	72	82	28	96	58	86	52	28	78	76
Ed	I	54	50	20	52	62	48	52	66	64	40	30	42	50
	II	48	42	36	40	96	6	98	98	100	40	6	42	36
	III	68	68	96	46	56	0	68	90	66	62	0	68	68
	IV	66	66	66	66	64	78	66	64	4	66	68	66	66
	V	62	76	48	60	76	32	94	74	84	68	14	60	76
Un	I	60	60	42	60	58	48	52	90	54	52	26	56	60
	II	62	58	58	30	92	22	82	94	70	30	12	52	30
	III	34	64	90	14	60	32	78	56	90	60	32	64	64
	IV	58	76	80	80	92	66	28	62	44	82	52	76	76
	V	52	72	66	72	54	48	82	38	72	54	26	70	68

4.3. HAC with Ward's Linkage Results

We performed the HAC with Ward linkage of the solar, heliospheric and geomagnetic parameters described in Section 2.1 in detail, merged with the three types of failure (Ag, Ed, Un) during five geomagnetic storms. The HAC performs optimally if all variables are independent and can be well described by a normal distribution (Norusis, 2011). Thus, we have used the standardized data shifted for the cross-correlation factors as described in Section 2.2. The results are presented in the form of the hierarchical binary cluster tree called the dendrogram in Figure 11.

One way to determine the natural cluster divisions in a data set is to compare the height of each link in a cluster tree with the heights of the neighboring links below it in the tree. A link that is approximately the same height as the links below indicates that there are no distinct divisions between the objects that are joined at this hierarchy level. These links exhibit a high level of consistency because the distance between the data being bound is approximately the same as the distances between the objects they contain. On the other hand, a link whose height differs noticeably from the height of the links below indicates that the data bound at this level in the cluster tree are much farther apart from each other than their components were when they were joined. Taking the above into account and analyzing the dendrogram presented in Figure 11, we see that the considered data were divided into two main clusters. Unique colors are assigned to each group of nodes in the dendrogram whose linkage is less than 0.7 of the maximum link.

We are most interested in what heliospheric and geomagnetic parameters are within a small distance from the network failures (Ag, Ed, Un) during geomagnetic storms. The smaller the link height between transmission line failures (Ag, Ed, Un), the stronger the relationship with the considered solar activity parameter is. Figure 11 presents the results
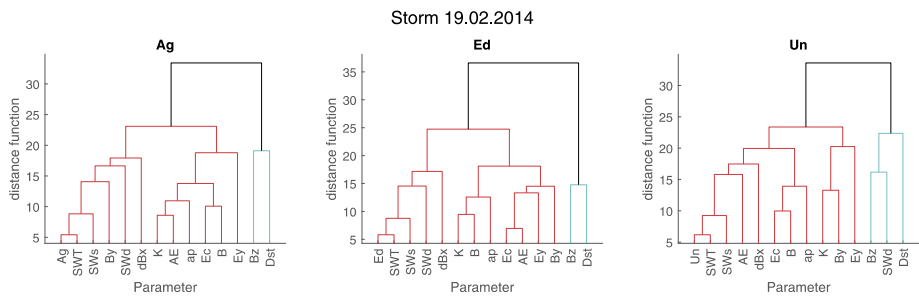
Storm 19.02.2014

Figure 11 Dendrograms presenting the Ward's linkage of the solar–activity parameters and three types of failure (Ag, Ed, Un) during the geomagnetic storm on 19.02.2014.

for the storm that took place on 19.02.2014. One can see that the parameters SWT and SWs are closest to all types of failures/anomalies. In the same central cluster are grouped the parameters By, SWd, and max(dB$_X$/dt) for the Ag failures; the parameters SWd and max(dB$_X$/dt) for the Ed failures and SWd and max(dB$_X$/dt) for the Ag failures AE and max(dB$_X$/dt) for the Un type of failures. Very similar clusters were recorded for the storm on $15 - 16.07.2012$. During this storm, the parameters of the same cluster with a distance below 20 are grouped: SWd, SWs, SWT and B$_y$, for both Ag and Ed types of failures and SWd, B$_y$, K, SWT, SWs for Un failures. The HAC analysis for the storm on 14.11.2012 revealed the cluster containing the SWs, SWd parameters with a distance below 20 for Ag failures; max(dB$_X$/dt), SWs, SWd, and E$_y$ for Ed failures, while for Un failures were listed SWs, max(dB$_X$/dt), and K parameters. During the storm on 9.03.2012, the list of parameters linked with distance below 20 is shorter and contains for Ag failures the parameters: SWd, B$_y$, while for Ed and Un failures only the parameter SWd. Slightly different results were obtained for the storm that occurred on 1.06.2013. The closest to the Ag and Ed failures are the B$_y$, E$_y$ parameters, and the B$_y$, max(dB$_X$/dt) and B parameters for Un failures.

The results obtained by the HAC support the results of the SOM analysis. A careful examination of the dendrograms revealed no one-to-one correspondence in the cluster structure between the storms. Nevertheless, in most of the considered storms, the solar-wind parameters (SWT, SWd, SWs) are at a small distance from the failures. Furthermore, the HMF components B$_y$ and variation max(dB$_X$/dt) are also interchangeably grouped with the failures for all storms, while the electric field component E$_y$ is clustered with the failures during the storm on 01.06.2013.

Consequently, we can conclude that the HAC method suggests that, as a proxy for transmission line failures, the solar-wind parameters, HMF B$_y$ component or max(dB$_X$/dt), and the electric field component E$_y$ can be considered.

5. Summary

In this work, we have considered sources on the Sun (halo and partial halo CMEs) and the impacts of the five intense geomagnetic storms that occurred during the first half of Solar Cycle 24. Applying three machine learning methods: Principal Components Analysis, Self-Organizing Maps, and Hierarchical Agglomerative Clustering (PCA, SOM, HAC), we have tried to discover the hidden dependencies between the solar-wind and heliospheric parameters (B, B$_y$, B$_z$, SWs, SWd, SWT, E$_y$), as well as geomagnetic characteristics (Dst, ap,

AE, K, max(dB_X/dt)), and computed geoelectric field Ec with three types of transmission grid failures/anomalies registered in southern Poland. The main findings of this work can be summarized in the following way:

i) Cross-correlation analysis showed that for some geomagnetic storms, the correlation between EGFs and solar-wind, heliospheric and geomagnetic parameters is relatively small, although statistically significant (fairly often, the absolute value of the cross-correlation coefficient is above 0.50). Moreover, the lag length is changeable without apparent uniformity due to consecutive geomagnetic storms and different parameters. However, the lags for the geomagnetic indices are fairly comparable. A similar situation holds for solar-wind parameters.

ii) The PCA analysis revealed that in the first principal component, the parameters ap, AE, and Ec occur together. The PCA map has shown that in the first principal component, SWT, Ey, and ap-index accompany failures caused by the aging of infrastructure elements, Ag, for all considered geomagnetic storms.

iii) The SOM analysis showed that the solar-wind parameters have the most substantial impact on all the considered types of electric–grid failures in southern Poland in all the storms under study. These parameters are usually grouped in the same cluster with failures in more than 80% of the results.

iv) The HAC analysis pointed out that the predictive factor for the transmission line failures should be searched among the solar-wind parameters, HMF B$_y$ component, and the electric field component E$_y$, as well as max(dB_X/dt).

In summary, the analyses conducted for the five chosen geomagnetic storms preceded by halo and partial halo CMEs showed that selected solar-wind and geomagnetic parameters can be treated as the primary/crucial for the analysis and an attempt to explain grid failures appearance. In the future, we plan to extend our analyses to a further period and attempt to establish common features of rapid changes on the Sun and link them to the geomagnetic storms for which there would be increased malfunctioning of energy infrastructure elements. We will also include registrations of the increased occurrence of higher harmonics of alternating current in the network as a possible repercussion of forceful solar variability in our studies.

Declarations

References

Ádám, A., Prácser, E., Wesztergom, V.: 2012, Estimation of the electric resistivity distribution (Eurhom) in the European lithosphere in the frame of the eurisgic wp2 project. *Acta Geod. Geophys. Hung.* **47**, 377. DOI.

Albert, D., Schachinger, P., Bailey, R.L., Renner, H., Achleitner, G.: 2022, Analysis of long-term GIC measurements in transformers in Austria. *Space Weather* **20**, e2021SW002912. DOI.

Asvestari, E., Rindlisbacher, T., Pomoell, J., Kilpua, E.K.J.: 2022, The spheromak tilting and how it affects modeling coronal mass ejections. *Astrophys. J.* **926**, A87. DOI.

Bailey, R.L., Halbedl, T.S., Schattauer, I., Achleitner, G., Leonhardt, R.: 2018, Validating GIC models with measurements in Austria: evaluation of accuracy and sensitivity to input parameters. *Space Weather* **16**(7), 887. DOI.

Bailey, R.L., Leonhardt, R., Möstl, C., Beggan, C., Reiss, M.A., Bhaskar, A., Weiss, A.J.: 2022, Forecasting GICs and geoelectric fields from solar wind data using LSTMs: application in Austria. Space Weather 2021SW002907. DOI.

Barbosa, C., Alves, L., Caraballo, R., Hartmann, G.A., Papa, A.R.R., Pirjola, R.J.: 2015, Analysis of geomagnetically induced currents at a low-latitude region over the solar cycles 23 and 24: comparison between measurements and calculations. *J. Space Weather Space Clim.* **5**, A35. DOI.

Boteler, D.H., Pirjola, R.J., Marti, L.: 2019, Analytic calculation of geoelectric fields due to geomagnetic disturbances: a test case. *IEEE Access* **7**, 147029.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: 2008, *Time Series Analysis: Forecasting and Control*, Wiley, New York.

Caldas, R., Hu, Y., de Lima Neto, F.B., Markert, B.: 2017, Self-organizing maps and fuzzy C-means algorithms on gait analysis based on inertial sensors data. *Adv. Intell. Syst. Comput.* **557**, 197. DOI.

Cliver, E.W., Schrijver, C.J., Shibata, K., Usoskin, I.G.: 2022, Extreme solar events. *Living Rev. Solar Phys.* **19**, 2. DOI.

Desai, R.T., Zhang, H., Davies, E.E., Stawarz, J.E., Mico-Gomez, J., et al.: 2020, Three-dimensional simulations of solar wind preconditioning and the 23 July 2012 interplanetary coronal mass ejection. *Solar Phys.* **295**, 130. DOI.

Eroshenko, E.A., Belov, A.V., Boteler, D., Gaidash, S.P., Lobkov, S.L., Pirjola, R., Trichtchenko, L.: 2010, Effects of strong geomagnetic storms on Northern railways in Russia. *Adv. Space Res.* **46**, 1102. DOI.

Fiedorov, E.N., Mazur, N.G., Pilipenko, V.A.: 2021, Electromagnetic response of the mid-latitude ionosphere to power transmission lines. *J. Geophys. Res. Space Phys.* **126**, 10. DOI.

Gil, A., Modzelewska, R., Moskwa, S., Siluszyk, A., Siluszyk, M., Wawrzynczak, A., Pozoga, M., Tomasik, L.: 2020a, The solar event of 14 – 15 July 2012 and its geoeffectiveness. *Solar Phys.* **295**, 135. DOI.

Gil, A., Modzelewska, R., Moskwa, S., Siluszyk, A., Siluszyk, M., Wawrzynczak, A., Pozoga, M., Domijanski, S.: 2020b, Transmission lines in Poland and space weather effects. *Energies* **13**(9), 2359. DOI.

Gil, A., Berendt-Marchel, M., Modzelewska, R., Moskwa, S., Siluszyk, A., Siluszyk, M., Tomasik, L., Wawrzaszek, A., Wawrzynczak, A.: 2021, 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. DOI.

Gonzalez, W.D., Tsurutani, B.T.: 1987, Criteria of interplanetary parameters causing intense magnetic storms (Dst < −100 nT). *Planet. Space Sci.* **35**, 1101. DOI.

Gopalswamy, N.: 2009, The CME link to geomagnetic storms. *Proc. Int. Astron. Union* **5**, S264. DOI.

Gopalswamy, N., Xie, H., Akiyama, S., Makela, P., Yashiro, S.: 2014, Major solar eruptions and high-energy particle events during solar cycle 24. *Earth Planets Space* **66**, 104. DOI.

Gopalswamy, N., Yashiro, S., Thakur, N., Makela, P., Xie, H., Akiyama, S.: 2016, The 2012 July 23 backside eruption: an extreme energetic particle event? *Astrophys. J.* **833**, 216. DOI.

Gopalswamy, N., Mäkelä, P., Yashiro, S., Lara, A., Xie, H., Akiyama, S., MacDowall, R.J.: 2018, Interplanetary type II radio bursts from wind/WAVES and sustained gamma-ray emission from Fermi/LAT: evidence for shock source. *Astrophys. J.* **868**(2), L19. DOI.

Hess, P., Zhang, J.: 2017, Earth-affecting solar transients. *Solar Phys.* **292**, 80. DOI.

Hu, H., Hu, W., Gökmen, N., Li, P., Huang, P., Chen, Z.: 2019, High resolution wind speed forecasting based on wavelet decomposed phase space reconstruction and self-organizing map. *Renew. Energy* **140**, 17.

Jolliffe, I.T.: 2002 In: *Principal Component Analysis*, 2nd edn. New York.

Kilpua, E.K.J., Good, S.W., Dresing, N., Vainio, R., Davies, E.E., et al.: 2021, Multi-spacecraft observations of the structure of the sheath of an interplanetary coronal mass ejection and related energetic ion enhancement. *Astron. Astrophys.* **656**, A8. DOI.

Kohonen, T.: 1990, The self-organizing map. *Proc. IEEE* **78**(9), 1464. DOI.

Lampinen, J., Oja, E.: 1992, Clustering properties of hierarchical self-organizing maps. *J. Math. Imaging Vis.* **2**, 261.

Lippmann, R.P.: 1987, An introduction to computing with neural nets. *IEEE ASSP Mag.* **4**(2), 4.

Müllner, D.: 2013, Fast hierarchical, agglomerative clustering routines for R and Python. *J. Stat. Softw.* **53**, 1.

Nikitina, L., Trichtchenko, L., Boteler, D.H.: 2016, Assessment of extreme values in geomagnetic and geoelectric field variations for Canada. *Space Weather* **14**, 481. DOI.

Nitta, N.V., Mulligan, T., Kilpua, E.K.J., et al.: 2021, Geomagnetic storms associated with "stealth" coronal mass ejections. *Space Sci. Rev.* **217**, 82. DOI.

Norusis, M.: 2011, Cluster analysis. In: *IBM SPSS Statistics 19 Guide to Data Analysis*, Norusis & SPSS Inc.

Oliveira, D.M., Ngwira, C.N.: 2017, Geomagnetically induced currents: principles. *Braz. J. Phys.* **47**, 552. DOI.

Palmerio, E., Lee, C.O., Mais, M.L., Luhmann, J.G., Lario, D., et al.: 2022, CMEs and SEPs during November–December 2020: a challenge for real-time space weather forecasting. *Space Weather* **20**, e2021SW002993. DOI.

Papaioannou, A., Sandberg, I., Anastasiadis, A., Kouloumvakos, A., Georgoulis, M.K., Tziotziou, K., Tsiropoula, G., Jiggens, P., Hilgers, A.: 2016, Solar flares, coronal mass ejections and solar energetic particle event characteristics. *J. Space Weather Space Clim.* **6**, A42. DOI.

Pilipenko, V.A.: 2021, Space weather impact on ground-based technological systems. *J. Solar-Terr. Phys.* **7**, 3. DOI.

Priest, E.R.: 2001, The role of the Sun's magnetic field. In: *ASP Conference* **248**, 25.

Rodriguez, L., Barnes, D., Hosteaux, S., Davies, J.A., Willems, S., et al.: 2022, Comparing the heliospheric cataloging, analysis, and techniques service (HELCATS) manual and automatic catalogues of coronal mass ejections using solar terrestrial relations observatory/heliospheric imager (STEREO/HI) data. *Solar Phys.* **297**, A23. DOI.

Schrijver, C.J., Mitchell, S.D.: 2013, Disturbances in the US electric grid associated with geomagnetic activity. *J. Space Weather Space Clim.* **3**, A19. DOI.

Svanda, M., Smickova, A., Vybostokova, T.: 2021, Modelling of geomagnetically induced currents in the Czech transmission grid. *Earth Planets Space* **73**(1), 229. DOI.

Svanda, M., Mourenas, D., Zertova, K., Vybostokova, T.: 2020, 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. DOI.

Temmer, M.: 2021, Space weather: the solar perspective. *Living Rev. Solar Phys.* **18**, 4. DOI.

Torta, J.M., Serrano, L., Regué, J.R., Sánchez, A.M., Roldán, E.: 2012, Geomagnetically induced currents in a power grid of northeastern Spain. *Space Weather* **10**, S06002. DOI.

Tozzi, R., De Michelis, P., Coco, I., Giannattasio, F.: 2019, A preliminary risk assessment of geomagnetically induced currents over the Italian territory. *Space Weather* **17**, 46. DOI.

Trichtchenko, L., Boteler, D.H.: 2002, Modelling of geomagnetic induction in pipelines. *Ann. Geophys.* **20**, 1063. DOI.

Veeramany, A., Unwin, S.D., Coles, G.A., Dagle, J.E., Mil-lard, D.W., Yao, J.: 2016, Framework for modeling high-impact, low-frequency power grid events to support risk-informed decisions. *Int. J. Disaster Risk Reduct.* **18**, 125. DOI.

Viljanen, A., Pirjola, R., Prácser, E., Katkalov, J., Wik, M.: 2014, Geomagnetically induced currents in Europe: modelled occurrence in a continent wide power grid. *J. Space Weather Space Clim.* **4**, A09. DOI.

Ward, J.H.: 1963, Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**(301), 236.

Wehrens, R.: 2009, Data mapping: linear methods versus nonlinear techniques. In: *Comprehensive Chemometrics*, Elsevier, Amsterdam, 619.

Zhang, J.J., Wang, C., Sun, T.R., Liu, Y.D.: 2016, Risk assessment of the extreme interplanetary shock of 23 July 2012 on low-latitude power networks. *Space Weather* **14**, 259. DOI.

Zhang, J., Temmer, M., Gopalswamy, N., et al.: 2021, Earth-affecting solar transients: a review of progresses in solar cycle 24. *Prog. Earth Planet. Sci.* **8**, 56. DOI.

Zois, J.P.: 2013, Solar activity and transformer failures in the Greek national electric grid. *J. Space Weather Space Clim.* **3**, A32. DOI.