

PAPER • OPEN ACCESS

## Neural net clustering in the study of electrical grids failures in relation to geomagnetic storms

To cite this article: A Siluszyk *et al* 2019 *J. Phys.: Conf. Ser.* **1391** 012107

View the [article online](#) for updates and enhancements.

You may also like

- [Dynamic Responses of Radiation Belt Electron Fluxes to Magnetic Storms and their Correlations with Magnetospheric Plasma Wave Activities](#)  
Xudong Gu, Shanjiao Xia, Song Fu et al.
- [Alfvénicity-related Long Recovery Phases of Geomagnetic Storms: A Space Weather Perspective](#)  
Daniele Telloni, Raffaella D'Amicis, Roberto Bruno et al.
- [A model of the telluric current eddy: a case from eastern China](#)  
Xin Zhang, Jun Liu and Xuebin Du



**ECS**  
The  
Electrochemical  
Society  
Advancing solid state &  
electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research

# Neural net clustering in the study of electrical grids failures in relation to geomagnetic storms

A Siluszyk<sup>1</sup>, A Gil<sup>1</sup>, R Modzelewska<sup>1</sup>, Sz Moskwa<sup>2</sup>, M Siluszyk<sup>1</sup> and A Wawrzynczak<sup>3</sup>

<sup>1</sup> Institute of Mathematics and Physics, Faculty of Sciences, Siedlce University, Konarskiego Str. 2, Poland

<sup>2</sup> Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology in Krakow, al. Mickiewicza 30, Poland

<sup>3</sup> Institute of Computer Science, Faculty of Sciences, Siedlce University, Konarskiego Str. 2, Poland

E-mail: [agnieszka.siluszyk@uph.edu.pl](mailto:agnieszka.siluszyk@uph.edu.pl)

**Abstract.** Nowadays, the space weather issues are of a great importance, especially, for all satellites operators or ground-based electrical and electronic systems caused by geomagnetic storms. In spite of many studies on the changeable Sun and the fluctuations in the interplanetary space triggered by solar-driven disturbances, the question, which of the components of geomagnetic storms influence the strongest the electrical and electronic systems is still open. Here, we consider the data of electrical grids breakdowns having unidentified reasons, as well as failures connected to the aging of the infrastructure elements and breakdowns of electronic devices, which occurred during the periods of an increased geomagnetic activity. Mathematical and statistical methods, among them neural net clustering, demonstrate that there exist connections between the groups of the coefficients characterizing the state of Earths vicinity during magnetic storms and the number of failures in electrical network.

## 1. Introduction

Our home - Earth is not located in a quiet region of the Universe. It is continuously exposed to a quasi-electrically neutral stream of charged particles, which is constantly emitted by the Sun, in the form of solar wind. Solar wind, moving with an average speed of 400 km/s, carries a frozen in magnetic field [12]. When solar wind, with frozen magnetic field, reaches the vicinity of the Earth, the Earth's magnetic field prevents its from penetrating deep into the geosphere [13]. However, the Earth's magnetosphere does not protect everywhere with the same power. In the equatorial area, this shield is the strongest, but around the poles, the Earth's magnetic field is the weakest and there are particles of solar wind penetrating the ionosphere, which occur as spectacular Aurora Borealis or Aurora Australis at high latitudes. There are, however, moments when, in addition to solar wind, the coronal mass ejecta (CME) leave the Sun [11]. And enormous amount of matter and energy causes the Earth's magnetic field to be disturbed and we are dealing with the phenomenon of a magnetic storm. The faster CME and more oriented towards the Earth, the stronger is the impact. At that time, numerous effects are induced on



Earth and its atmosphere, such as degradation or blocking of high-frequency radio waves used in radio communication, by the induction of extra currents in the ground. Furthermore, we can observe that the power transmission grids can be degraded and the signals from radio navigation systems (GPS and GNSS) can be modified causing lower accuracy [1, 15, 2]. Also, auroras are observed at lower latitudes [16].

The parameters that depict the behavior of the solar wind are its speed, density and temperature, as well as the strength of the heliospheric magnetic field, measured by the probes in the interplanetary space. The state of the Earth's magnetic field is defined by the values of the strength and components of the Earth's magnetic field measured in numerous observatories located all over the Earth. Geomagnetic indices such as Dst, Kp, Ap, and AE are used to characterize the strength of disturbances of the Earth's magnetic field.

A huge quantity of natural physical data that we obtain from stations conducting observations of the Earth's magnetic field (e.g., the Earth Observatory and climate and environmental science at NASA, The Institute of Geophysics in Polish Academy of Sciences in Belsk, GFZ German Research Centre for Geosciences in Potsdam, etc.) in conjunction with statistical techniques and methods are used in geomagnetic storms' analysis processes. The complexity of these multidimensional analyzes leads to a variety of statistics and coefficients. Understanding the importance and significance of these statistics and coefficients gives "in times of data" enormous possibilities. Correlation analysis, regression analysis, analysis of variance (ANOVA/MANOVA), canonical analysis, discriminant analysis, principal components and classification analysis, neural net clustering and many, many other give a wide spectrum of knowledge in physics, in mathematics, but also in medicine, the computer science, energetics, etc., [14, 9, 4]. Nowadays, one of the most popular method to study big data is the neural network.

The aim of this paper is to analyze the problem of space weather effects and the indices of magnetic storms which influence the most the ground-based electrical systems. We study whether, and, in what extent, the Polish energy infrastructure is affected by the space weather outcomes. We present that there exist connections between the groups of the coefficients characterizing the state of Earth's vicinity during magnetic storms and the number of failures in electrical network in southern Poland. Such analysis is presented in the next sections of this paper. This article is organized as follows: in Section 1 we shortly point the problem of space weather, the changeable Sun and the fluctuations in the interplanetary space triggered by solar-driven disturbances. Section 2 characterizes electrical grids in the southern Poland and describes common causes of their failures. Here, we introduce data analyzed in this paper. In Section 3 we present applied methods and discuss our results.

## **2. Characteristics of the electrical grids in the southern Poland and the most common causes of their failures**

### *2.1. Solar storms-a short overview*

The unique phenomenon of the Sun's activity is the solar wind - an extension of the solar corona into interplanetary space [12], changing in time and space. Due to this extension, the powerful events taking place on the Sun: e.g. solar flares (SFs), CMEs, etc. cause the disturbances in the interplanetary space. The term space weather is defined as the set of the solar-driven effects affecting the Earth's environment. Interactions of Sun-induced-phenomena with the Earth's magnetic field can lead to geomagnetic disturbances. Geomagnetic storms are classified by the geomagnetic indices: Dst, Kp, Ap, and AE. Strong magnetic storm affects the normal operation of ground located electrical systems and causes damages of satellites and its equipment, which impacts satellite phones, GPS systems, etc.

### 2.2. Electrical grids failures

We consider data of the electrical grid failures (EGF) from the Distribution System Operator (DSO) Tauron, which has the largest share on the electricity sales market in Poland. This DSO supplies  $\sim 25\%$  electricity of Poland, therefore there is important to understand the nature of these failures, which could be linked with geomagnetic storms. Our data concern two periods: year 2010 and period at January-July 2014. In 2010, DSO noticed the number 25616 of minor failures, whilst in January-July 2014 it was 30155. All these failures we have grouped into six more general clusters according to possible causes (see, A-F causes).

**Table 1.** Electrical grids disruptions causes in southern Poland energy distribution network in 2010 and in period of January-July 2014.

Cluster	Description of the cluster	Number of failures in 2010	% of failures	Number of failures in January-July 2014	% of failures
A	Meteorological effects	3653	14.26	9004	29.8
B	Operational shutdowns	16614	64.86	9703	32.0
C	Vandalism	824	3.22	792	2.6
D	Aging	1917	7.5	6209	20.6
E	Electronics devices	32	0.1	1181	3.91
F	Unidentified reasons	2576	10.1	3266	10.8

We can see that the clusters A-C in Table 1 can be treated as objective causes, whereas the groups D-F can be associated to space weather effects. It gives 4525 failures in 2010 and 10656 in the first seven months of 2014, which might have solar origin, and only these failures are considered in the further analysis.

### 2.3. Space weather event in April 2010

During considered time, we have investigated a few space weather events. One of them is the event observed in April, 5-7, 2010 (some details of behavior of solar geomagnetic indices data for this period we can see in the Table 2). At the beginning of April 2010 during of enhanced solar and geomagnetic activity we could observed the change of the parameters. Then Kp-index reached almost 8, HMF strength increased almost three and a half times during only five hours, reaching value 18.8 nT at noon on 5<sup>th</sup> of April. Bz component was very changeable during this time interval, varying from -7 nT up to 11.5 nT. Solar wind speed grew almost twice up to 814 km/s in early afternoon on 5<sup>th</sup> of April. Even more pronounced effect was visible in solar wind temperature which increased from less than 200000 K up to more than 1000000 K during few hours. Solar wind density grew three times up to 12 n/cc at noon on 5<sup>th</sup> of April. Exactly at the same time electric field dropped to -8.6 mV/m and was very variable during this period. Earlier, on 3<sup>rd</sup> of April at 10:33 UT the Halo CME occurred with preceding solar flare having the onset at 9:04 UT. CME apparent speed was 668 km/s and space speed 939 km/s (<https://cdaw.gsfc.nasa.gov>). All of the above mentioned effects had their reflections in the geomagnetic indices alterations. Ap-index increased from around 15 nT up to almost 180 nT in

the morning of 5<sup>th</sup> of April. At the same time AE-index exceeded -1400 nT. One day later, in the early afternoon of 6<sup>th</sup> of April Dst-index dropped around four times to -81 nT.

**Table 2.** Behavior of the values of some parameters of solar wind and geomagnetic indices for event in April, 5-7, 2010

Kp index	B [nT]	Bz [nT]	Speed of wind [km/s]	Temp. of solar wind [K]	Density of solar wind [n/cc]	Electric field [mV/m]	Ap [nT]	AE [nT]	Dst [nT]
↗ 8	↗ 18,8	-7÷ 11	↗ 814	$2 \cdot 10^5 \div 10^6$	↗ 12	↓ -8,6	↗ 180	↗ 1400	↓ -81
	3.5×	big variability	2×		3×	big variability			4×

#### 2.4. Effects of space weather on electrical transmission network

Solar storms are a phenomenon that has a comprehensive collection of different consequences on industrial systems like power transmission systems, oil and gas pipelines, telecommunication cables, and railway equipment. During severe solar storms, a geomagnetically induced current (GIC) starts to flow through long conductors, such as power lines and pipelines. GIC is caused by the interaction between the intensive movements in the interplanetary space and the Earth's magnetic field due to, e.g., coronal mass ejecta transferring its energy to the magnetosphere. As a violent stream of particles penetrates the Earth's magnetosphere (often referred to as a geomagnetic storm) it can result in a high current electrojet in the ionosphere. Changes in the current of the electrojet cause fluctuations in the geomagnetic field, which under Faraday's Law of induction, induces an electric field [17]. This electrojet can reach several million amperes during geomagnetic storms.

The power grids are today's systems the most vulnerable to GIC [6], and a great deal of research has been carried out about GIC impact on power systems. These research underline the need to decrease the GIC influence on power systems. It can be done by two ways 1) block the flow of GIC and design equipment to be GIC-insensitive, but this is expensive and not simple, and 2) forecasting GIC to allow systems operators take actions to minimize problems in their systems. In these contexts, important is to fully understand all relationships between measurable parameters describing the state of the interplanetary space and failures of the power systems occurring at a similar time.

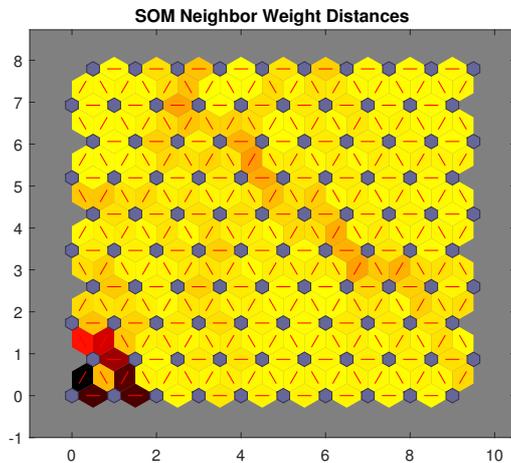
### 3. Methods applied in analysis

In this work we study a set of 16 parameters of solar wind and geomagnetic field with 72 three-hours observations of each parameter, which might be essential for explanation of causes of the failures in the electrical network. Here, we present three methods, i.e., neural net clustering, discriminant analysis and a method of principal components and classification analysis in order to show that there exist connections between the groups of the coefficients characterizing the state of Earth's vicinity during magnetic storms and the number of failures in electrical network

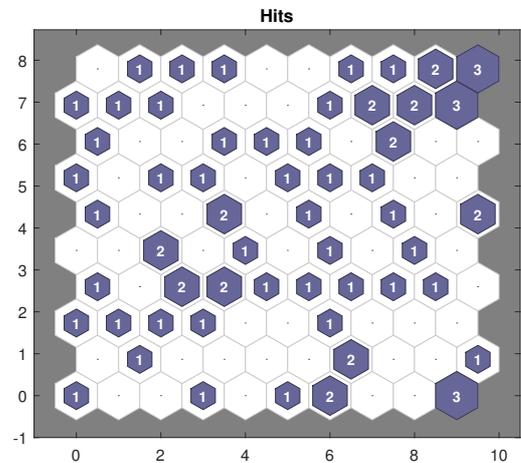
for southern Poland. Moreover by using above methods we demonstrate that the question about uniqueness of the solar and geomagnetic components which influence the strongest on the electrical systems is still under consideration.

### 3.1. Method of neural net clustering

Now, we deal with the method of neural net clustering, i.e., we study Self Organizing Maps (SOM) of Kohonen [8, 7]. SOM are used to calculate and analyze signals, moreover, they realize mappings of the input variables to output ones. The basic unit is a neuron. It is an approximate mathematical description of the human neuron. In real neural networks exist many neurons and their migration process is very large. The migration process leads to grouping of similar objects in compact region. Additionally, we can study the network itself or the connection of neurons only. Model of the neuron used in the Kohonen network works in this way that on the input is calculated the distance between the input vector  $x_i$  and the weight vector  $w_i$ , ( $i = 1, \dots, n$ ), i.e.,  $q = \sum_{i=1}^n (x_i - w_i)^2$ , then on the output we observe that the signal  $y$  is the bigger if their distance  $q$  is smaller, i.e.,  $y = f(q) = \frac{1}{\sqrt{q+\epsilon}}$ ,  $\epsilon$  means some error. Self organization leads to the fact that neurons which surround similar input objects, in the topological layer of the network are as neighboring neurons. Due to the influence of the neighborhood, the topological layer neurons form clusters. Kohonen's network strives to create the optimal map showing relations between the input. In laboratory conditions, it is easy to obtain the Kohonen network, which will correctly differentiate individual situations occurring in the multidimensional data space. Unfortunately, this is not so easy when we analyze the actual multidimensional set of data. In

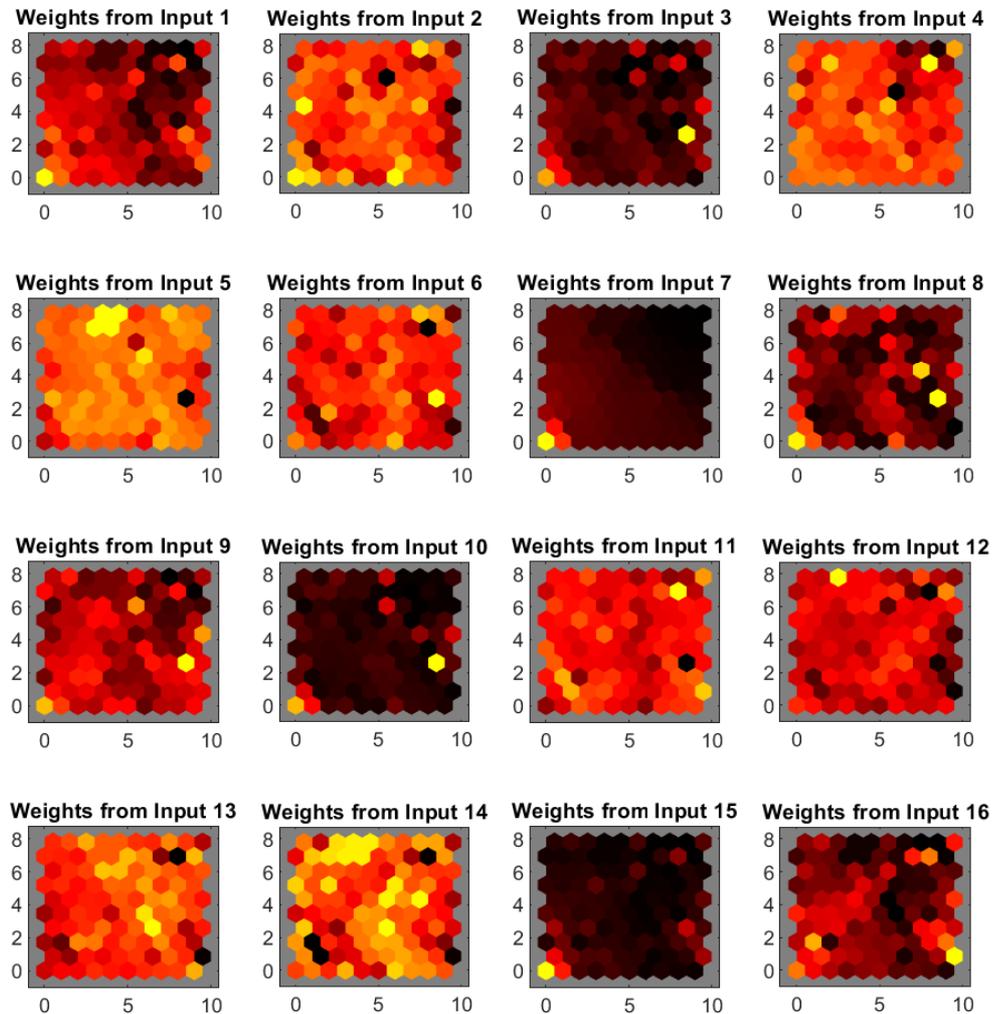


**Figure 1.** SOM neighbor weight distances for the considered solar and geomagnetic parameters for event in April, 5-7, 2010



**Figure 2.** SOM sample hits for the considered solar and geomagnetic parameters for event in April, 5-7, 2010

the Fig.1 we see that darker colors represent longer distances whilst lighter colors mean smaller distances. The belt of darker segments passes in the further middle area, which may suggest that the SOM focused the solar and geomagnetic parameters on certain groups. The default SOM topology is hexadecimal. Among sixteen considered parameters, only three are distinguished and among them the number 3 is the most important (see, Fig.2). We have received the SOM network for weights, i.e. we have received the weight planes for each element of the input vector. As before, darker colors represent larger weights. In case the connection patterns of the two



**Figure 3.** SOM input planes for the considered solar and geomagnetic parameters for event in April, 5-7, 2010. Here input 1 is Kp-index, input 2 is GCR, input 3-B,input 4-Bx,input 5-By,input 6-Bz,input 7-Proton $T_{emp}$ ,input 8-Proton $Density$ ,input 9-Flow Speed, input 10-Flow Pressure,input 11-Ey,input 12-Plasma Beta,input 13-A.M.Number,input 14-Dst-index,input 15-Ap-index,input 16-AE-index

inputs were very similar, it can be assumed that the input data is strongly correlated. Observing the Figure 3, it can be seen that the planes 3 and 10, as well as 7 and 15; 1 and 16 are similar to each other.

### 3.2. Method of discriminant analysis

Discriminant analysis is the method based on multidimensional statistics and is used to resolve which variables distinguish two or more naturally emerging groups [4]. With this method, we can determine the rules of conduct that assign multidimensional objects to one of many populations with known parameters with possibly minimal classification errors. The main idea of the discriminant analysis is the answer a question "whether the groups of observations for some variable differ in terms of the average or not", after that discriminant analysis uses this variable to predict adherence to a group. In the case of a single variable, the answer to a question whether the variable discriminates groups, i.e., two or more groups differ significantly from each other of the average, we obtain by the F-test, as in Analysis of Variance (ANOVA) [4].

F-test is obtained as the ratio of the between-groups variability called *Mean Square Effect* ( $MS_{effect}$ ) to the within-group variability called *Mean Square Error* ( $MS_{error}$ ). If between-groups variance is really large, there must be significant differences between the means. In the case when we consider many variables, the matter is the more computationally complicated. We consider a matrix of total variances and covariances; in addition, we have a matrix of the within-group variances and covariances. To resolve whether there exist any significant differences (relating to all variables) between groups or not, we compare these two matrices applying multidimensional F-test (like in multivariate analysis of variance (MANOVA)) [4].

Below we present results of discriminant analysis for the failures of electrical grids caused by the aging that we obtain for observations on the Earth during the geomagnetic storm in April, 5-7, 2010 (see Tables 3-5). Here, we have divided type of failures into 3 groups, i.e., "gr.1" means 0 failures connected with aging of electricity infrastructure's elements, "gr.2" means the number of the failures connected with aging in the interval (0, 10) and "gr.3" means the number of failures  $\geq 10$ .

Here, we have chosen progressive step analysis. With this choice, we have introduced variables to the model one by one, always choosing the variable that makes the most significant contribution to discrimination. With this way we can build the best prediction of the failures, knowing the variables selected for this model.

**Table 3.** Discriminant Function Analysis Summary in April, 5-7, 2010. There are 7 variables in the model and 3 group for Aging (grouping variable). Wilk's Lambda: 0.62054 approx.  $F(14, 126) = 2.425$ ,  $p < 0.0048$

Variables	Wilk's Lambda	Partial Lambda	F-remove (2, 63)	p-value	Toler.	1-Toler. (R-square)
Dst	0.66703	0.93031	2.35976	0.10274	0.35013	0.64987
Ey	0.68487	0.90607	3.26568	0.04472	0.36863	0.63137
B	0.67675	0.91694	2.85344	0.06512	0.24293	0.75707
Kp	0.70557	0.87949	4.31616	0.01751	0.18252	0.81748
AE-index	0.65179	0.95204	1.58670	0.21267	0.12418	0.87582
Bx	0.64116	0.96784	1.04682	0.35708	0.69818	0.30182
A.M.Number	0.64062	0.96865	1.01936	0.36670	0.35679	0.64321

**Table 4.** Discriminant Function Analysis Summary in April, 5-7, 2010. Canonical variables for EGF caused by the aging of infrastructure's elements.

Variables	Root 1	Root 2
Dst	0.00306	0.064744
Ey	0.63589	0.678404
B	0.43381	0.266928
Kp	-1.02001	-0.267892
AE-Index	0.00464	0.001090
Bx	0.01299	-0.221817
A.M.Number	0.21505	0.198882
Constant	-3.34625	-0.893031
Cum.Prop	0.60125	1.00000

**Table 5.** Discriminant Function Analysis Summary in April, 5-7, 2010.  $\chi^2$ -tests for 7 variables in the model

Roots removed	Eigenvalue	Canonical R	Wilk's Lambda	$\chi^2$	df	p-value
Root 1	0.32543	0.49551	0.62054	31.49289	14	0.00473
Root 2	0.21583	0.42133	0.82248	12.89814	6	0.04468

**Table 6.** Discriminant Function Analysis Summary in April, 5-7, 2010. Means of canonical variables for groups EGF caused by the aging of infrastructure's elements.

Group for Aging	Root 1	Root 2
gr.1	-0.02750	0.12221
gr.2	-1.33352	-1.89153
gr.3	2.92142	-1.25659

Discriminant function analysis shows that instead of sixteen parameters we can deal only with seven ones, i.e., *Dst*, *Ey*, *B*, *Kp*, *AE*, *Bx* and *A.M.Number* which are the most significant in our model, despite that there are only two components statistically significant (see, Table 3). The partial value of Wilk's Lambda indicates that the Kp-index has the largest (the Lambda fraction is the smallest) contribution to general discrimination of our analysis and the next variables are Ey, B and Dst. The variable which has the smallest, but still highly significant contribution, is the A.M.Number. In the Table 3, the column before last presents the value of redundancy for given variable. The tolerance is the smallest for AE-index, it means that the variable's

		Correlations															
		Marked correlations are significant at $p < .05000$															
		N=72 (Casewise deletion of missing data)															
Variable	Kp	K(Asm)-Belsk	GCR	B	Bx	By	Bz	Proton_Temperature	Proton_Density	Flow_Speed	Flow_Pressure	Ey	Plasma_Beta	Alfa_Mach_Number	Dst	Ap	AE_Index
Kp	1.000	0.892	-0.397	0.722	0.463	-0.337	-0.462	0.487	0.163	0.743	0.492	0.440	-0.526	-0.576	-0.568	0.827	0.842
K(Asm)-Belsk	0.892	1.000	-0.387	0.638	0.390	-0.364	-0.407	0.449	0.171	0.707	0.487	0.384	-0.451	-0.467	-0.521	0.737	0.763
GCR	-0.397	-0.387	1.000	-0.185	-0.265	0.119	0.427	0.306	0.303	-0.457	0.009	-0.444	0.498	0.276	0.707	-0.181	-0.490
B	0.722	0.638	-0.185	1.000	0.350	-0.639	-0.166	0.358	0.580	0.605	0.802	0.131	-0.538	-0.624	-0.253	0.757	0.617
Bx	0.463	0.390	-0.265	0.350	1.000	-0.068	-0.447	0.212	0.096	0.294	0.128	0.449	-0.286	-0.381	-0.289	0.283	0.423
By	-0.337	-0.364	0.119	-0.639	-0.068	1.000	-0.134	-0.186	-0.492	-0.397	-0.646	0.184	0.244	0.247	0.096	-0.381	-0.279
Bz	-0.462	-0.407	0.427	-0.166	-0.447	-0.134	1.000	0.089	0.329	-0.122	0.254	-0.991	0.446	0.574	0.587	-0.218	-0.651
Proton_Temperature	0.487	0.449	0.306	0.358	0.212	-0.186	0.089	1.000	0.347	0.354	0.446	-0.120	0.141	-0.080	0.063	0.670	0.212
Proton_Density	0.163	0.171	0.303	0.580	0.096	-0.492	0.329	0.347	1.000	0.130	0.816	-0.361	0.192	0.090	0.486	0.328	-0.083
Flow_Speed	0.743	0.707	-0.457	0.605	0.294	-0.397	-0.122	0.354	0.130	1.000	0.599	0.114	-0.505	-0.212	-0.506	0.627	0.582
Flow_Pressure	0.492	0.487	0.009	0.802	0.128	-0.646	0.254	0.446	0.816	0.599	1.000	-0.294	-0.133	-0.066	0.093	0.647	0.234
Ey	0.440	0.384	-0.444	0.131	0.449	0.184	-0.991	-0.120	-0.361	0.114	-0.294	1.000	-0.453	-0.557	-0.604	0.198	0.641
Plasma_Beta	-0.526	-0.451	0.498	-0.538	-0.286	0.244	0.446	0.141	0.192	-0.505	-0.133	-0.453	1.000	0.792	0.635	-0.335	-0.650
Alfa_Mach_Number	-0.576	-0.467	0.276	-0.624	-0.381	0.247	0.574	-0.080	0.090	-0.212	-0.066	-0.557	0.792	1.000	0.535	-0.452	-0.717
Dst	-0.568	-0.521	0.707	-0.253	-0.289	0.096	0.587	0.063	0.486	-0.506	0.093	-0.604	0.635	0.535	1.000	-0.400	-0.774
Ap	0.827	0.737	-0.181	0.757	0.283	-0.381	-0.218	0.670	0.328	0.627	0.647	0.198	-0.335	-0.452	-0.400	1.000	0.664
AE_Index	0.842	0.763	-0.490	0.617	0.423	-0.279	-0.651	0.212	-0.083	0.582	0.234	0.641	-0.650	-0.717	-0.774	0.664	1.000

**Figure 4.** Correlation matrix for the considered solar and geomagnetic parameters for event in April, 5-7, 2010

contribution to discrimination of this variable is minimal in the comparison with the contribution of the others. Moreover, from Table 5 we see that the weights of the canonical variables like Kp-index, Ey and B (Root1 in the Table 4) are statistically significant with  $p = 0,004726$ . If these parameters rise, then the most probable is that the number of failures will grow in the 3<sup>rd</sup> group, see Table 6. Unfortunately, above components are not the same for other events, we have discriminated 14 from 16 parameters (AE-index, Bx, By, GCR, Dst-index, Bz, Ey, Proton Density, B, A.M.Number, Plasma Beta, Proton Temperature and Flow Speed), whereas in other event only 8 from 16 (Bx, A.M.Number, Dst-index, Bz, Kp-index, B, Flow Speed and Plasma Beta) which had the most significant contribution to discrimination. In summary: the parameters Kp-index, Ey, B and Dst-index the most discriminate the 3<sup>rd</sup> group.

### 3.3. Method of principal components and classification analysis

Very real problem which has to be taken into account is a relation between investigated parameters (see, Fig.4). Unfortunately, discriminant analysis demands that the correlation between variables was as small as possible. Therefore, as an additional method we will use the principal components analysis.

The principal components analysis is the set of the methods and procedures that focus on determining completely new variables (principal components)  $Z_i$ ,  $i = 1, \dots, k$  being a linear combination of observed (primary) variables  $X_1, X_2, \dots, X_n$  [10]

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{in}X_n, \quad (1)$$

where  $a_{i1}, a_{i2}, \dots, a_{in}$  are coefficients of  $i^{\text{th}}$  main component. Without loss of generality we obtain a new mathematical model that contains reduced primary variables. The principal components analysis allows to indicate these primary variables that have a large impact on the appearance of individual major components, i.e., only those such form a homogeneous group. The main component is then the representative of this group and the next components are mutually orthogonal (uncorrelated) and their number  $k$  is less than or equal to the number of

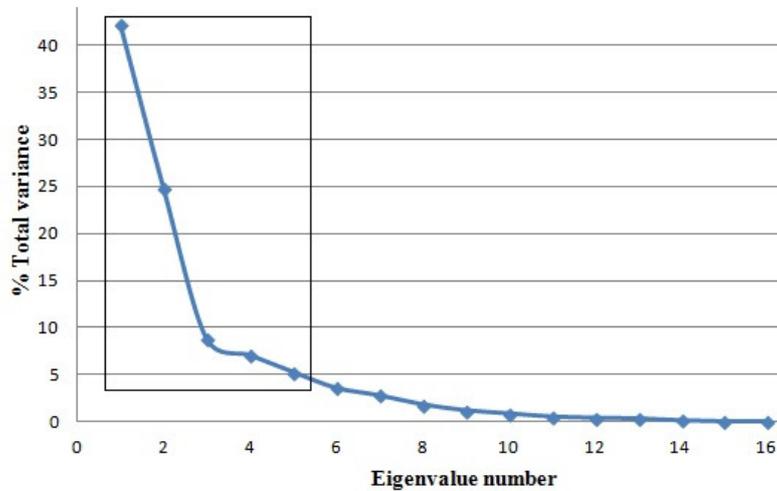
original variables  $n$ . Each of the principal component is described by eigenvalue, eigenvector, factor coordinates of the variables, variable contribution and communalities [10]. The eigenvalue gives information about which part of the total variability is explained by the given principal component. The eigenvector expresses the influence of the individual primary variables on a given principal component. It contains the coefficients  $a_{i1}, a_{i2}, \dots, a_{in}$  and the sign of those coefficients indicates the direction of influence. Factor coordinates the variables, like the coefficients included in the eigenvector, show the influence of particular variables on a given principal component. These are values that present which part of the variance of a given component originates from primary variables. When the analysis is based on a correlation matrix, these values are interpreted as correlation coefficients between the original variables and the given principal component [10]. Variable contribution is based on the value of determination coefficients between the original variables and the given principal component. It indicates what percentage of variability of a given principal component can be explained by the variability of individual primary variables. Communalities are based on the determination coefficients. They indicate what percentage of the variability of a given primary variable can be explained by the variability of the first few principal components. For example, the result for the second variable contained in the fourth column (Table 9) for the principal component tells us what percentage of the variability of the second variable can be explained by the variability of the first four principal components.

**Table 7.** Bartlett test and the KMO coefficient of considered solar and geomagnetic parameters for event in April, 5-7, 2010

	Summary
Number of analyzed variables	16
Number of analyzed cases	72
$\alpha$	0.05
Bartlett test	
$H_0$ :	$M = I$
$H_1$ :	$M \neq I$
$\chi^2_{empir.}$ value	1666.13098
$\chi^2_{crit.}$ value	146.56736
number of degrees of freedom	120
p	0.000000
Kaiser-Mayer-Olkin coefficient	
KMO	0.99

The first step, which we perform, before proceeding analysis of components is checking the advisable of conducting this analysis. In our analysis we start from verification by Bartlett test and the fixing the Kaiser-Mayer-Olkin (KMO) coefficient [10]. Bartlett test checks whether there are significant differences between the correlation matrix  $M$  and the unit matrix  $I$ . If we rejected hypothesis  $H_0$  then there exists justifiability of the principal components analysis. The coefficient  $KMO \in [0, 1]$ , if  $KMO < 0.5$  then we don't have bases to do the principal components analysis.

From the Table 7 we can see that  $p < \alpha$  ( $p=0.000000$ ) in Bartlett test, so there exists the



**Figure 5.** The scree test of considered solar and geomagnetic parameters for event in April, 5-7, 2010

difference between  $M$  and  $I$  which is statistically significant. KMO coefficient equals to 0,99 so assumptions to principal components analysis are fulfilled.

**Table 8.** Eigenvalues of correlation matrix of considered solar and geomagnetic parameters for event in April, 5-7, 2010

Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Comulative %
1	6.74623	42.16393	6.74623	42.16393
2	3.97283	24.83021	10.71906	66.99414
3	1.40853	8.80333	12.12760	75.79747
4	1.13010	7.06313	13.25770	82.8606
5	0.84738	5.29611	14.10507	88.15671
6	0.57547	3.59669	14.68054	91.7534
7	0.44936	2.80850	15.12990	94.5619
8	0.29524	1.84522	15.42514	96.40712
⋮	⋯	⋯	⋯	⋯
16	0.00272	0.01699	16.00000	100.0000

Presented eigenvalues (Table 8) indicate that five main components will supply us with the same knowledge that earlier 16 components. The eigenvalues for the first four components are more than 1 and the percentage of the variance explained by them is 82,86%.

Only factor with eigenvalue greater than 1 can be considered, whilst factor which extracts less than one original variable is dropped. This criterion was proposed by Kaiser in 1960. In our case we also present fifth eigenvalue according to Jolliffe criterion [5] which is similar to criterion of Kaiser but here the critical value for eigenvalue is  $\geq 0.7$ . The fifth component explains much less variance (5.30%) and its eigenvalue is 0.847377.

The scree test is a popular graphical method to study the number of canonical components first proposed by Cattell (1966) [3]. In our study we plot the eigenvalues as a simple line plot (Fig.5). From the scree test we conclude that after the fifth eigenvalue the line tends into the horizontal line. The total percentage of the variance explained by five canonical components is 88.16%.

**Table 9.** Communalities of considered solar and geomagnetic parameters for event in April, 5-7, 2010

Variable	From 1 factor	From 2 factor	From 3 factor	From 4 factor	From 5 factor
Kp-index	0.83339	0.85601	0.87892	0.91500	0.91509
GCR	0.28831	0.45775	0.69038	0.73827	0.86859
B	0.61965	0.87553	0.88105	0.96069	0.96219
Bx	0.28051	0.28881	0.42688	0.44807	0.76181
By	0.17098	0.47734	0.63750	0.69941	0.69941
Bz	0.35074	0.74136	0.84454	0.87615	0.89127
Proton Temp.	0.09092	0.40171	0.73504	0.86843	0.90912
Proton Density	0.00668	0.72980	0.73816	0.85465	0.92011
Flow Speed	0.52183	0.59710	0.66245	0.85587	0.87316
Flow Pressure	0.19258	0.90676	0.92592	0.92691	0.94723
Ey	0.33081	0.77003	0.86703	0.89136	0.90870
Plasma Beta	0.53894	0.60705	0.72895	0.78182	0.83113
A.M.Number	0.56049	0.60998	0.61324	0.80910	0.93805
Dst-index	0.51837	0.75908	0.80751	0.88753	0.89349
Ap-index	0.59890	0.76819	0.81477	0.86740	0.89764
AE-index	0.84308	0.87250	0.87517	0.87697	0.88801

Table 9 presents in the column entitled "From 5 factor" that for the first five components, the variance of each variable is represented by using these components in at least 69.94%. Additionally the communalities are at similar very high level.

Applying eigenvectors, the factor coordinates of the variables and the variable contributions we obtain explanation which variables are the main components. In our case it is:

- The eigenvector for the first component is the largest for AE-index and is equal -0.3535, as well as for Kp-index is -0.3515 (see Table 10). The factor coordinates of the variables for both AE- and Kp-index inform that the correlation between the first main component and AE-Index as well the first main component and Kp-index is very high and amounts -0.91819 and -0.91290, respectively (see Table 11). It gives around 12.5% for AE- and 12.35% Kp-index of the variable contributions in the first component, see Table 12.
- In the case of the second component the eigenvector is the largest for Proton Density and is equal -0.42663, as well as for Flow Pressure=-0.42399. The factor coordinates of the variables for both Proton Density and Flow Pressure informs that the correlation between the first main component and Proton Density, as well as the first main component and Flow Pressure is very high and amounts to -0.85036 and -0.84509, respectively. It gives around 18.2% for Proton Density and 18% Flow Pressure of the variable contributions in the first component (Tables 10-12).

**Table 10.** Eigenvectors of correlation matrix of considered solar and geomagnetic parameters for event in April, 5-7, 2010

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Kp-index	-0.35147	-0.07546	-0.12753	0.17867	-0.00999
GCR	0.20672	-0.20651	-0.40639	-0.20586	-0.39216
B	-0.30307	-0.25378	0.06260	-0.26547	-0.04209
Bx	-0.20391	0.04569	-0.31309	-0.13692	0.60848
By	0.15920	0.27769	-0.33720	0.23404	-0.00181
Bz	0.22801	-0.31356	0.27066	0.16723	-0.13360
Proton Temp.	-0.11609	-0.27969	-0.48647	0.34356	-0.21911
Proton Dens.	-0.03146	-0.42663	-0.07704	-0.32105	0.27793
Flow Speed	-0.27812	-0.13765	0.21538	0.41371	0.14284
Flow Pressure	-0.16895	-0.42399	0.11664	-0.02956	0.15484
Ey	-0.22144	0.33249	-0.26243	-0.14571	0.14305
Plasma Beta	0.28264	-0.13093	-0.29418	0.21629	0.24124
A.M.Number	0.28824	-0.11160	0.04813	0.41630	0.39010
Dst-index	0.27719	-0.24614	-0.18542	-0.26609	0.08382
Ap-index	-0.29795	-0.20643	-0.18185	0.21578	-0.18893
AE-index	-0.35351	0.08605	-0.04348	0.03995	-0.11412

- In the case of the third component we obtain Proton Temperature, for fourth A.M.Number, whilst for the last component we get HMF Bx component (Tables 10-12).
- We conclude that the first component we can call AE-index, the second Proton Density, the third Proton Temperature, the fourth A.M.Number and the fifth HMF Bx component.

**Table 11.** Factor coordinates of the variables of considered solar and geomagnetic parameters for event in April, 5-7, 2010

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Kp-index	-0.91290	-0.15041	-0.15136	0.18994	-0.00920
GCR	0.53694	-0.41163	-0.48231	-0.21884	-0.36099
B	-0.78718	-0.50584	0.07430	-0.28221	-0.03875
Bx	-0.52963	0.09108	-0.37158	-0.14555	0.56012
By	0.41349	0.55350	-0.40019	0.24880	-0.00167
Bz	0.59223	-0.62499	0.32122	0.17777	-0.12298
Proton Temp.	-0.30153	-0.55748	-0.57735	0.36523	-0.20170
Proton Dens.	-0.08173	-0.85036	-0.09143	-0.34130	0.25584
Flow Speed	-0.72237	-0.27436	0.25562	0.43980	0.13149
Flow Pressure	-0.43884	-0.84509	0.13843	-0.03142	0.14254
Ey	-0.57516	0.66273	-0.31146	-0.15596	0.13168
Plasma Beta	0.73413	-0.26097	-0.34914	0.22993	0.22207
A.M.Number	0.74866	-0.22245	0.05712	0.44255	0.35910
Dst-index	0.71998	-0.49062	-0.22007	-0.28287	0.07716
Ap-index	-0.77388	-0.41145	-0.21583	0.22939	-0.17391
AE-index	-0.91819	0.17152	-0.05160	0.04247	-0.10505

**Table 12.** Variable contributions of considered solar and geomagnetic parameters for event in April, 5-7, 2010

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Kp-index	0.12353	0.00569	0.01626	0.03192	0.00010
GCR	0.04273	0.04264	0.16515	0.04237	0.15378
B	0.09185	0.06440	0.00392	0.07047	0.00177
Bx	0.04158	0.00208	0.09802	0.01874	0.37025
By	0.02534	0.07711	0.11370	0.05477	0.00000
Bz	0.05199	0.09832	0.07325	0.02796	0.01784
Proton Temp.	0.01347	0.07822	0.23665	0.11803	0.04801
Proton Density	0.0099	0.18201	0.00593	0.10307	0.07724
Flow Speed	0.07735	0.01894	0.04639	0.17115	0.02040
Flow Pressure	0.02854	0.17976	0.01360	0.00087	0.02397
Ey	0.04903	0.11055	0.06887	0.02152	0.02046
Plasma Beta	0.07988	0.01714	0.08654	0.04678	0.05820
A.M.Number	0.08308	0.01245	0.00231	0.17330	0.15218
Dst-index	0.07683	0.06058	0.03438	0.07080	0.00702
Ap-index	0.08877	0.04261	0.03307	0.04656	0.03569
AE-index	0.12497	0.00740	0.00189	0.00159	0.01302

### 3.4. Summary

In our article, we used three computational methods, i.e., Self Organizing Maps type neural network method, a discriminant analysis method and a principal component analysis method. Using the above mentioned methods, we show groups of solar and geomagnetic parameters that have the strongest influence on the occurrence of an aging type in electric networks on the Earth. The parameter groups that we have received do not give a unique answer, although we can see a parameter that repeats in each of them. Our results display that electrical grids failures from groups D-F in southern Poland can occur more frequently after the solar-driven effects are generated at the Earth's vicinity.

### 3.5. Acknowledgments

The Institute of Meteorology and Water Management -National Research Institute for the meteorological data. Data of solar wind and geomagnetic indexes are from <https://www.gfz-potsdam.de/>, <http://rtbel.igf.edu.pl/>, <https://omniweb.gsfc.nasa.gov/>, CME from <https://cdaw.gsfc.nasa.gov/> and SSC from <http://www.obsebre.es>. We acknowledge the financial support by the Polish National Science Centre, decision number DEC-2016/22/E/HS5/00406

Sz.M. and M.S. acknowledge the Polish National Science Centre, grant no. 2014/15/B/ST8/02315.

## 4. References

- [1] Beggan C, Wild J and Gibbs M 2018 The ground effects of severe space weather, *Astronomy & Geophysics* **59** 4.36–4.39

- [2] Bothmer V and Daglis J 2007 *Space Weather. Physics and Effects* (Berlin: Springer)
- [3] Cattell R 1978 *Use of Factor Analysis in Behavioral and Life Sciences* (New York: Plenum)
- [4] Chiang Ch 2003 *Statistical Methods of Analysis* (World Scientific Pub. Co. Inc.)
- [5] Jolliffe I 2002 *Principal Component Analysis, Second Edition* (New York: Springer)
- [6] Kappenman J 2004 The evolving vulnerability of electric power grids, *Space Weather*, **21**
- [7] Kaski S 1997 Data Exploration Using Self-Organizing Maps *Acta Polytechnica Scandinavica. Mathematics, Computing and Management in Engineering* **82** 55
- [8] Kohonen T 1982 Self-Organized Formation of Topologically Correct Feature Maps *Biological Cybernetics* **43** (1) 59–69
- [9] Lehmann E and Romano J P 2005 *Testing Statistical Hypotheses* (New York: Springer-Verlag)
- [10] Lever J, Krzywinski M and Altman N 2017 Principal component analysis *Nature Methods* **14** 641
- [11] Owens M and Forsyth R 2013 The heliospheric magnetic field *Living Rev. Sol. Phys.* **10** 5
- [12] Parker E 1958 Dynamics of the Interplanetary Gas and Magnetic Field *Astrophysical Journal* **128** 664
- [13] Richardson I 2018 Solar wind stream interaction regions throughout the heliosphere *Living Rev. Sol. Phys.* **15**
- [14] Shao J 2003 *Mathematical Statistics* (New York: Springer-Verlag)
- [15] Schulte in den Bäumen H, Moran D, Lenzen M, Cairns I and Steenge A 2014 How severe space weather can disrupt global supply chains, *Nat. Hazards Earth Syst. Sci.* **14**, 2749–2759
- [16] Vaquero J, Trigo R and Gallego M 2007 Sporadic aurora from Spain, *Earth Planets Space*, **59**, e49–e51
- [17] Wik M 2008 *The Sun, Space Weather and Effects* (Kiruna, Sweden: Swedish Institute of Space Physics)