New Astronomy 86 (2021) 101570

Contents lists available at ScienceDirect

New Astronomy

journal homepage: www.elsevier.com/locate/newast

Statistical analysis on the current capability to predict the Ap Geomagnetic Index

Evangelos Paouris [a,b,*], Maria Abunina [c], Anatoly Belov [c], Helen Mavromichalaki [a]

[a] Faculty of Physics, National and Kapodistrian University of Athens, Athens, Greece
[b] Institute for Astronomy, Astrophysics, Space Applications & Remote Sensing of the National Observatory of Athens, Penteli, Greece
[c] Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation, Russian Academy of Sciences – IZMIRAN, Moscow, Russia

ARTICLE INFO

Keywords:
Magnetosphere
Geomagnetic Disturbances
Forecast
Statistics

ABSTRACT

In this work, forecasting results of the daily Ap geomagnetic index by different Space Weather Prediction Centers (SWPCs) covering the time period from October 2014 up to July 2020, are considered. A three-day forecast of this index obtained from the space weather prediction centers, is analyzed. Standard forecast verification measures, descriptive statistics with correlation coefficients, and error analysis between forecasts and observations have been performed to evaluate the quality or the skill of the predictions. In particular, an error analysis using fit performance metrics, such as the mean average error (MAE) and the root mean squared error (RMSE) among others, as well as the threshold performance metrics, such as the Hansen-Kuipers, Gilbert and Heidke skill scores, the probability of detection (POD), the probability of false detection (POFD) and the area under the curve in receiving the operating characteristic (ROC) plot among others, are calculated. For the Ap geomagnetic index predictions during Day-0, the Pearson's correlation coefficient is ranging between the values of 0.57 and 0.79, while during the Day-2 it is decreased to the range from 0.37 to 0.44. In conclusion, the majority of the used SWPCs perform quite accurately on the conditions of Space Weather in active as well as in quiet periods showing a reliable effort for predicting geomagnetic storms.

Introduction

In the last twenty years, the solar physics community has put a lot of effort to understand, as much as possible, the mechanisms that take place in the interior or in the upper atmosphere of the Sun to predict eruptive events, such as solar flares and coronal mass ejections (CMEs). These events, as well as the high-speed streams of solar wind originating from coronal holes are the events, which change the normal conditions of the interplanetary medium producing a series of effects (Schwenn, 2006; Bothmer and Daglis, 2007; Gopalswamy, 2009; Koskinen, 2011). The impacts from these events to space-borne and ground-based technological systems that affect human life and health are those of the space weather. The term 'Space Weather' reflects the targeted research towards understanding and predicting the heliospheric and geospace environments, which ultimately impact our lives and our technology, mostly due to the solar activity. In general, changes in the physical conditions in space affecting space technologies together with life on Earth in various ways, are the field of the research of the Space Weather. As the community defines Space Weather in multiple contexts, the reviews on the origin of the term Space Weather from Kane (2006) and Cade and Chan-Park (2015) are useful.

Some of the effects of the Space Weather are, but not limited to, satellite drag and sensor degradation, effects of geomagnetically induced currents (GICs) on the power grids and pipelines, radiation threat to polar flight crews and astronauts, as well as high-frequency communication outages in polar regions (Gopalswamy, 2009). Eroshenko et al. (2010) reported anomalies in the operation of the system of signalization, centralization, and blockage (SCB) in some divisions of the high-latitude Russian railways. These anomalies were revealed as false traffic light signals about the occupation of the railways. These abnormal signals occurred during the main phase of 17 severe geomagnetic storms from 2000 to 2005. Recently Vybostokova and Svanda (2019) presented a statistical analysis of the disturbances of the Czech electric power distributors and the geomagnetic activity. They concluded that there is evidence that the mid-latitude power grid may also be affected by the Space Weather events.

The above effects of Space Weather are described by the generic term of "Space Storms". A comprehensive review on these effects is the one by

* Coresponding Author:
E-mail address: evpaouris@phys.uoa.gr (E. Paouris).

https://doi.org/10.1016/j.newast.2021.101570
Received 8 September 2020; Received in revised form 12 January 2021; Accepted 13 January 2021
Available online 19 January 2021

Koskinen (2011), while the potential impacts of Space Storms are described in detail by (Bothmer and Daglis, 2007).

From the above, it is evident that it is of great interest to have early and accurate prediction of a space storm. For that reason, a number of Space Weather Prediction Centers (hereafter SWPCs) have been developed in many countries providing information to the scientific community, as well as to the public about the Space Weather conditions in the days ahead.

The purpose of the current research is to present a statistical analysis of the forecasts of the Ap geomagnetic index by the various SWPCs providing data for the same time interval, as a proxy for the prediction of geomagnetic storms. We are utilizing well-established performance metrics that have been applied in the past in the same kind of verification and validation researches. In particular, Devos et al. (2014) applied a verification analysis to evaluate the performance of the Solar Influences Data analysis Center (SIDC) at the Regional Warning Center of Belgium forecasts of some basic space weather parameters, such as the F10.7 radio flux, the occurrence of a solar flare and the local geomagnetic index (K-index). They used descriptive model statistics, common verification measures, error analysis, and conditional plots related to forecasts and observations. In Cui et al. (2016) verification measures were calculated to assess the quality of solar proton events forecasts by the Space Environment Prediction Center (SEPC) at the National Space Science Center of the Chinese Academy of Sciences. Murray et al. (2017) presented a verification of archived forecasts for solar flares occurrence from an active region over the next 24 hours as well as full-disk forecasts for the next four days at the Met Office Space Weather Operations Center. Kubo et al. (2017) showed a verification study of an operational solar flare forecast in the Regional Warning Center of Japan using many conventional verification measures. Liemohn et al. (2018) presented a series of recommended fit performance metrics and event detection performance to benchmark a new or updated geomagnetic index prediction model. These researches present in general the same metrics for validation and verification of various forecasted quantities, either binary (e.g., the occurrence of a solar flare) or continuous variables (e.g., the value of a geomagnetic index). To the best of our knowledge is the first time to date that a validation analysis on the forecasts of Ap geomagnetic index values utilizing predictions from seven space weather prediction centers is presented here. The space weather prediction center of the National Oceanic and Atmospheric Administration (SWPC/NOAA) has performed a comprehensive forecast verification analysis on their predictions (maximum Kp index values, storm warnings, estimated Ap, etc. – available at https://www.swpc.noaa.gov/content/geomagnetic-activity-forecast-verification). In particular, their estimated Ap forecast verification utilizing Ap geomagnetic index data (observed and predicted) is covering the period July 1986 – December 2013 providing many metrics.

It is noted that in our work we do not focus on the methods used by the various SWPCs to make the forecasts of the Ap geomagnetic index. The goal of the current work is to provide to the community the current capabilities of international forecast centers to predict the daily Ap index as a potential tool for the forecast of geomagnetic storms.

Some information about the Ap geomagnetic index and the SWPCs where their data were used in the current research, are given in Section 2. Statistical analysis with the calculated validation and verification metrics concerning the prediction of the Ap index from the various SWPCs is presented in Section 3. Finally, a discussion and conclusions of this work are summarized in Section 4.

Data Selection and Space Weather Prediction Centers

This current research focuses on the predictions by various SWPCs concerning the geomagnetic index Ap which is calculated from the Kp index. Further analysis, as well as the physics behind the derivation of the Kp index, are actually out of the scope of the current work. An analytical review about the history and the physics behind the K index and Kp derivation one can found at Menvielle et al. (2011). However, a brief description on the derivation of the daily Ap is presented in the next paragraph.

Table 1
One-to-one correspondence transform table from Kp to *ap* (in nT) values.

Quiet conditions				*Disturbed conditions*			
Kp	**ap**	**Kp**	**ap**	**Kp**	**ap**	**Kp**	**ap**
0o	0	2+	9	5-	39	7o	132
0+	2	3-	12	5o	48	7+	154
1-	3	3o	15	5+	56	8-	179
1o	4	3+	18	6-	67	8o	207
1+	5	4-	22	6o	80	8+	236
2-	6	4o	27	6+	94	9-	300
2o	7	4+	32	7-	111	9o	400

The Kp index is a measure of the strength of magnetospheric convection (see e.g. Thomsen, 2004) and we must note that the derivation of Kp values is not a simple one. This process can be split into three steps. The data used for the calculations in these steps are 3-hour intervals of thirteen mid-latitude magnetometer measurements (Bartels et al., 1939; Rostoker, 1972). Firstly, the deviations in magnetic field measurements are used to calculate the local K values for each station. Each observatory has a table to convert these deviations to quasi-logarithmic K values. The table values determined by the geomagnetic latitude of each observatory. K values range from 0 indicating very low activity (very quiet conditions) to 9 for very strong activity (extremely disturbed conditions). Secondly, from K values are calculated the Ks values taking into account the diurnal (near local midnight or not) as well as seasonal (winter, summer and the equinoxes) variations. The result from this process is Ks values as a continuous variable (as opposed to the integral K) ranging between 0.0 and 9.0 and given in thirds of an integer. Finally, the planetary magnetic activity index Kp is simply derived for each 3-hour interval by averaging the Ks indices for the thirteen mid-latitude magnetometer observatories and as Ks index, Kp ranges through twenty-eight grades from 0o to 9o. These values of Kp (e.g. 2- or 4+) are not very easy for arithmetic manipulation and Bartels (1951) introduced the idea of the use an index based on a linear scale rather than on a quasi-logarithmic scale. This index was the “daily equivalent planetary” amplitude Ap. The value of daily Ap geomagnetic index is the average of eight ap values computed for each 3-hour interval and it is expressed in nT. The ap values defined from the known values of Kp using a one-to-one correspondence table (Table 1) and are ranging from 0 (very quiet conditions) up to 400 (extreme geomagnetic storm). Further information on the calculation of K, Kp, ap and Ap indices one could find in Bartels, (1951); Rostoker, (1972); Menvielle et al., (2011) and references therein.

As presented earlier, the effects of Solar Activity on Earth are affecting a broad area of space technology and life on Earth. In particular, it is very important to be able to predict, as accurately as possible the occurrence and the potential strength of a geomagnetic storm. For that reason, we target on the predictions of the daily geomagnetic index Ap covering the period October 2014 up to July 2020 (2131 days) for each one of the seven different SWPCs around the globe. The motivation behind our selection of Ap index instead of Kp as an index of geomagnetic activity is quite simple. As described earlier Ap values obtained by a linear scale on opposition to the quasi-logarithmic scale of Kp values making arithmetic manipulation much easier. It is noted that the examined period from October 2014 up to July 2020 was chosen due to the availability of predictions of the Ap geomagnetic index. In particular, October 2014 was selected as starting point for our analysis due to the availability of Ap forecasts from ASWFC.

The examined period (Oct 2014 – Jul 2020) cover a small period of solar maximum of solar cycle 24 (SC24), the declining phase and finally the minimum between solar cycles 24 and 25. SC24 was one of the quietest solar cycles during the last 100 years and as a result there were

fewer geomagnetic storms on that period in contrast to SC23. Thus, our verification analysis performed utilizing data of Ap index of a quieter period in opposition to previous solar cycles.

The predicted daily values of Ap index were taken from the following SWPCs:

1 Space Weather Prediction Center of the National Oceanic and Atmospheric Administration (USA), through their daily product: "Report and Forecast of Solar and Geophysical Activity" (https://www.swpc.noaa.gov/products/report-and-forecast-solar-and-geophysical-activity). SWPC of NOAA is the pioneer in the field of space weather as they started to provide the report and forecast of solar and geophysical activity since 1966.
2 Athens Space Weather Forecasting Center of the Faculty of Physics of the National and Kapodistrian University of Athens (Greece), through the daily product of "Daily Forecast of Geomagnetic Activity" (http://spaceweather.phys.uoa.gr/Daily_Report.html). ASWFC is providing the Daily Forecast of Geomagnetic Activity in a 7/365 basis from October 2014, and it is the only forecasting center in Greece, which provides daily information to the public about the occurrence of geomagnetic storms based on the daily Ap index.
3 Space Weather Prediction Center of the Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation (Russia), through the daily product "Review and Forecast of Solar and Geomagnetic Activity" (http://spaceweather.izmiran.ru/eng/forecasts.html),
4 Solar Influences Data analysis Center, which is part of the Royal Observatory of Belgium and partner of the Solar Terrestrial Center of Excellence (Belgium), through their daily product "SIDC Ursigram", (http://sidc.oma.be/products/meu/),
5 Space Weather Prediction Center of Kazhakstan (Kazhakstan), through their daily product of Space Weather report, (http://ionos.kz/?q=ru/prognoz),
6 Space Weather Services of the Bureau of Meteorology of the Australian Government (Australia), through their daily product "Daily Report", (http://www.sws.bom.gov.au/Geophysical/3/1), and
7 Space Environment Prediction Center of the Center for Space Science and Applied Research of Chinese Academy of Sciences (China), through their daily product "Daily Reports" (http://eng.sepc.ac.cn/dailyForecast.php).

Hereafter the abbreviation for the above SWPCs is NOAA, ASWFC, IZMIRAN, SIDC, KAZ, SWS, and SEPC, respectively. All the above SWPCs are providing predictions of Ap geomagnetic index for the upcoming three days (the day of the report which was published –Day-0– and the next two days –Days 1 and 2– in our analysis).

At this point we clearly state that the goal of the current work is to provide to the community the current capabilities of international forecast centers to predict the daily Ap index as a potential tool for the forecast of geomagnetic storms. Even if we do not have direct access to the forecast pipeline of each one SWPC used in our work, however we think that some brief information behind the forecasts should be given. All centers are utilizing as many as possible data from various sources (detectors on orbit or on the ground). In particular, observations of the solar disk in many wavelengths (e.g. https://sdo.gsfc.nasa.gov/data/), solar wind plasma and magnetic field conditions (solar wind speed, temperature, density, total magnetic field strength B and the Bz magnetic field component available at https://www.swpc.noaa.gov/products/real-time-solar-wind and https://www.swpc.noaa.gov/products/ace-real-time-solar-wind), coronagraph images and CME data from real-time databases (e.g. CACTUS database: http://www.sidc.oma.be/cactus/out/latestCMEs.html and https://soho.nascom.nasa.gov/data/realtime/mpeg/), geomagnetic conditions through the ap, Ap, Kp and Dst indices (e.g. Potsdam database: http://www-app3.gfz-potsdam.de/kp_index/qlyymm.html and Kyoto database: http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html), etc. The main differences behind the methodologies used by the SWPCs are: a) the degree of the forecast automation, b) the availability of their models and databases and, last but not least, c) the experience and the intuition of the duty forecaster.

Table 2
Correlation coefficient (r) values between observed and predicted values of Ap geomagnetic index by various SWPCs.

Space Weather Prediction Center	Day 0 r	Day 1 r	Day 2 r
ASWFC	0.793	**0.539**	0.398
IZMIRAN	**0.794**	0.490	**0.437**
NOAA	0.583	0.516	0.389
SIDC	0.776	0.426	0.370
KAZ	0.727	0.483	0.384
SWS	0.573	0.470	0.394
SEPC	0.635	0.505	0.397

Furthermore, we provide a brief description of the methodology followed by ASWFC. We utilize a semi-empirical method based on the outcome from an auto-regressive moving average (ARMA) model as well as on the influence from crucial space weather factors such as the existence of coronal holes which are in Earth facing position and Earth directed CMEs. The ARMA method models the next step in the sequence of Ap geomagnetic index as a linear function of the observations (observed Ap values) and residual errors at prior time steps providing the so-called "baseline" forecasts of Ap. In the case of influence of high speed streams of solar wind from coronal holes and/or Earth directed CMEs, then this influence is being taken into account and increases the baseline of Ap values accordingly. For example, when an Earth directed CME is observed then we utilize the Effective Acceleration Model (see Paouris et al., 2021a; Paouris and Mavromichalaki, 2017a; 2017b) to estimate the Time-of-Arrival (ToA) of the associated shock at Earth. The output from EAM model is very important as the arrived CME affects the predictions of the specific day that the CME is expected to reach Earth. An empirical relation between the CME-index (Paouris, 2013) and the Ap index based on archived events is further being used to have an estimation of the potential Ap value according to the CME estimated characteristics (Paouris and Mavromichalaki, 2015). Obviously, at this point is crucial the experience of the duty forecaster or the role of the "*human-in-the-loop*" for the daily forecasts as Devos et al. (2014) and (Riley et al., 2018) also mentioned, to evaluate the bulk of this information and to make interventions accordingly.

The IZMIRAN space weather center follows a similar methodology to ASWFC. In particular, IZMIRAN space weather center combines the automation of collection, preparation, and processing of data with the decisive participation of the duty forecaster. For its forecasts, IZMIRAN relies on extensive local databases that combine all relevant archived data over a large period, as the database of geomagnetic activity and related parameters from 1859 up to date (e.g., daily Ap and 3-hour ap indices, minimum and maximum values of Kp, Ap and Dst indices). Furthermore, the ToA and the efficiency of a CME at Earth and/or high-speed streams of solar wind from coronal holes are estimated utilizing the Forbush-Effects and Interplanetary Disturbances (FEID, http://spaceweather.izmiran.ru/eng/dbs.html) database (Belov et al., 2014).

Forecast verification analysis

A verification analysis for the forecasts of Ap geomagnetic index from various SWPCs covering the period October 2014 – July 2020 is presented. The comparison of the residuals between the forecasted and

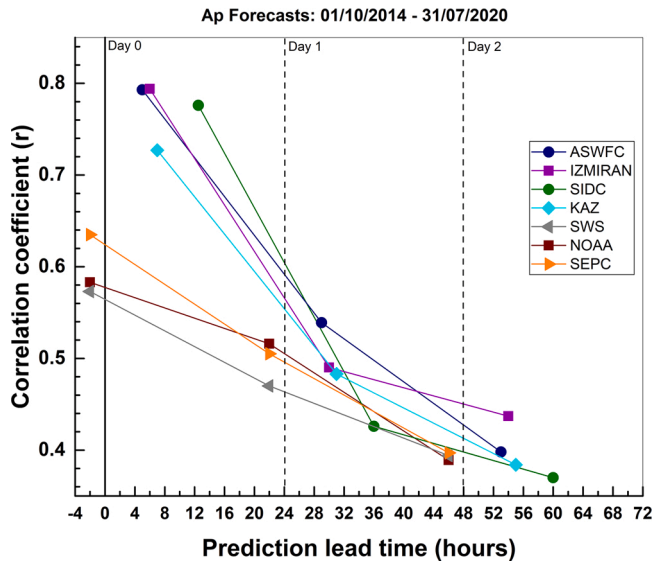

Fig. 1. Correlation coefficients between forecasted and predicted values of Ap geomagnetic index as a function of the prediction lead time, from the day of the report it was published (00:00 of Day-0) up to the next two days (Day 1 and 2).

the observed values is a way to evaluate the quality of the forecast. This can be done for different kinds of variables, such as binary (Yes/No) or continuous variables. For a continuous variable the error analysis as well as the calculation of skill scores can be done utilizing a threshold value.

Firstly, some basic descriptive statistics such as the least-squares method and the Pearson correlation coefficient were applied between the observed values and the predicted ones. The correlation coefficients were calculated for the first day (Day-0 – this is the day when the forecast was made) and for the upcoming next two days (Days 1 and 2). Pearson's correlation coefficients for all SWPCs between observed and forecasted values of the Ap index are presented in Table 2, as well as in Fig. 1. In particular, the horizontal axis represents the prediction lead time in hours and the vertical axis represents the Pearson's correlation coefficient. To avoid any misunderstanding with the sign (plus or minus) of the prediction leading times we explain here the methodology behind this. NOAA, SWS and SEPC providing their forecasts every day at 22:00 UT, 2 hours before the beginning of Day-0, (i.e. 00:00), so their prediction lead time is negative (–2 hours =2 hours before the time 00:00 of Day-0) in Fig. 1. All the other SWPCs providing their reports after the beginning of Day-0 and as a result their lead times are positives. For example, SIDC provide their forecasts around 12:30 UT and that implies the prediction actually covers only the half of Day-0 and the leading time is +12.5 hours. Obviously, the points for each SWPC differ exactly 24 hours. Of course, this "timing issue" could affect the performance of the metrics and we further discuss on that at the "Discussion and Conclusions" section. The best (maximum) correlation coefficient values are noted with bold font in the Table 2. For all SWPCs the Pearson correlation coefficient is decreasing with the maximum value that occurred for the first day of forecasts, while for the third day, the coefficient takes the lower value. The scatter plots of the predicted Ap values (y-axis) in relation to the observed values (x-axis) are presented in Fig. 2. Basic descriptive statistics such as the minimum and maximum values, the lower (25%) and the upper quartile (75%), the 1% and 99% values of each distribution, as well as mean and median values are presented in Table 3.

The histograms for each one class of the observed as well as the forecasted Ap values are presented in Fig. 3. The horizontal axis presents the bins of Ap values, while the vertical axis presents the frequency of appearance for each bin. It is noteworthy that predictions from ASWFC (first panel of the top line of Fig. 3) are almost identical to the actual values of Ap. The observed values of Ap are presented in each histogram as background columns with pattern of crossed lines. This makes easier the visual comparison between the forecast distributions for each SWPC and the observed ones. The next three panels (IZMIRAN, SIDC and KAZ) are different with a more Gaussian distribution and all of them have their maximum counts at the bin [7-9) of the Ap values. At this point we clarify that the limits of the bins e.g. $[x_1, x_2)$ means that we take for this bin, all the available values which are greater or equal to x_1 but less than x_2 Ap values. It is obvious that these three SWPCs are clearly overestimating the Ap values for the first 3 bins, i.e. for the very quiet conditions. The last three plots, of SWS, NOAA and SEPC, presenting different distributions in contrast with the other SWPCs. NOAA's forecasts are biased forcing the predictions equal or above the limit of $Ap \geq 5$. However, this is not correct as there are many daily values of $Ap < 5$, as it is obvious from the distribution of the observed values of Ap (columns with pattern of crossed lines). In particular, 793 days in a sample of 2131 days (37.2%) had Ap less than 5 indicating quiet or very quiet conditions. The same characteristic is also presented in the last panel concerning the SEPC center. NOAA and SEPC centers are overforecasting, especially in the very quiet periods where Ap is less than 5. This Fig. 3 is a very helpful one for the explanation of some of the results presented later in this section.

As it is known, there is no standard set of metrics used by geomagnetic index predictive-model developers to benchmark their models (Liemohn et al., 2018). As a result, we use a set of variables that are used by various researchers, as presented in the introductory section. Furthermore, a very comprehensive analysis of the various metrics can be found in the work of Jolliffe and Stephenson (2012), as well as in the website of Forecast Verification Research of the World Meteorological Organization – WMO (https://www.cawcr.gov.au/projects/verification/).

Fit Performance Metrics – Category I

At this category, the performance metrics are: the slope and the intercept (B and A) from a linear least squares regression fit between the observed and the forecasted values, the mean error (ME), the mean absolute error (MAE), the root mean squared error (RMSE), the skill score (SS) and the bias (BIAS). The Pearson correlation coefficient also belongs to this category, but this parameter has already been calculated (see Table 2 and Fig. 1). The mathematical equations and a brief description for these metrics are given in Appendix I and the results for these metrics (B, A, ME, MAE, RMSE, SS and BIAS) for each one of the SWPCs are presented in Table 4. In the first column of Table 4 is the metric, and the value corresponds to the best performance is inside the brackets. The results which are nearest to the value inside the brackets are noted with bold font.

Threshold Performance Metrics – Category II

The daily equivalent planetary amplitude Ap is the average value of eight "ap" values computed for each 3-hour time interval as described in Section 2. That implies Ap index is a continuous variable. For a continuous variable, it is more complex to create a binary "yes/no" situation. As a result, an index value serves as a "threshold" value (see, e.g., Liemohn et al., 2018) to create this "yes/no" criterion. In the current work a threshold value of Ap geomagnetic index was used starting from 1 up to 70 (an analysis over the whole sample revealed that above 70 there was not any significant difference) with a step of 1 and for every

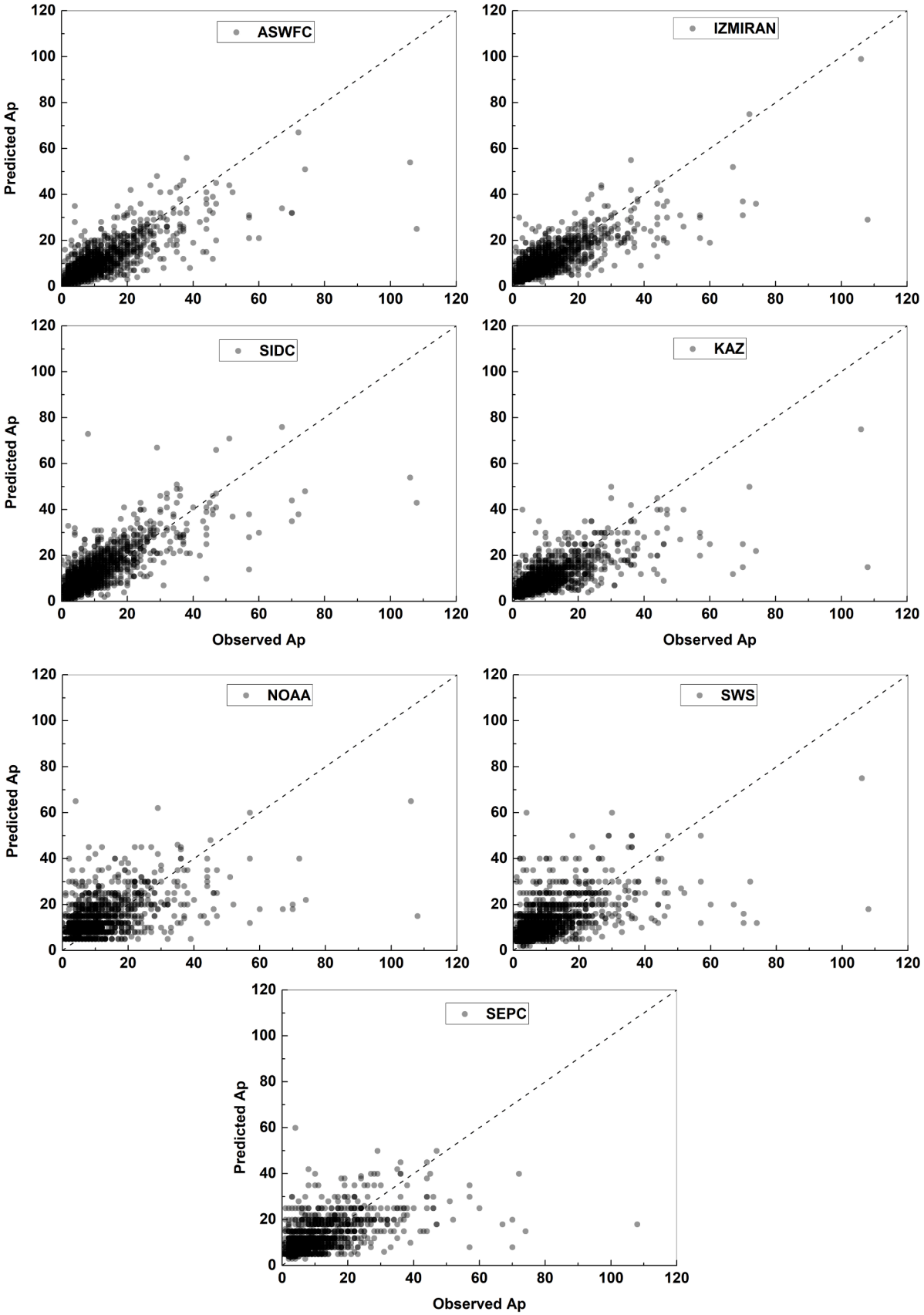

Fig. 2. Scatter plots of the predicted Ap values in relation to the actual ones for each one SWPC. The diagonal dashed line represents the best forecast line and corresponds to the line y = x. The prediction lead times for each SWPC for the Day-0 of the prediction are -2.0 hours for NOAA, SEPC and SWS, +5.0 hours for ASWFC, +6.0 hours for IZMIRAN, +7.0 hours for KAZ and +12.5 hours for SIDC.

Table 3
Basic descriptive statistics for observed Ap geomagnetic index as well as for the forecasted values of Day-0 by the used SWPCs.

Statistics	Min/ Max	1% and 99% of the distribution	25%/75% 1st and 3rd Quartile	Mean	Median
Observed Ap	0/ 108	1 – 44	4/11	8.78±0.19	6
ASWFC	1/67	1 – 36	4/11	8.40±0.15	6
IZMIRAN	1/99	2 – 33	5/12	9.38±0.15	7
NOAA	5/65	5 – 40	5/14	11.17±0.16	8
SIDC	1/76	2 – 41	6/14	11.32±0.18	9
KAZ	1/75	2 – 30	5/10	8.85±0.14	7
SWS	2/75	4 – 40	7/14	11.20±0.16	8
SEPC	3/ 150	4 – 35	5/12	10.47±0.16	8

threshold value a contingency table was created as follows:

- The term A is the number of the correct forecasted events or hits. When the predicted and observed values are less than or equal to the threshold value, this pair is considered as a "hit".
- The term B is the number of false alarms. A false alarm is a forecast of an event, while no event was observed. When a forecasted value is less or equal with the threshold value, while the observed value was higher than the threshold, this pair is considered as a "false alarm".
- The term C is the number of misses. A miss is an event that was not forecasted. In that case, the predicted value is higher than the threshold value, and the observed value is less or equal to the threshold value. This pair is considered as a "miss".
- The term D is the number of true negatives or correct rejections. A correct rejection is a forecast of a non-event, while indeed, no event was observed. In that case, the forecasted and the observed values are greater than the threshold, and this pair is considered as "correct negatives".

The sum of A+B+C+D is equal to the total number of pairs or the sample size (n). Various quantities and metrics are calculated using a contingency table. According to previous works (Jolliffe and Stephenson, 2012; Devos et al., 2014; Liemohn et al., 2018) as well as metrics from WMO (at https://www.cawcr.gov.au/projects/verification/), a set of useful quantities and skill scores, concerning the verification between forecasts and observed values, was created. The threshold verification metrics which were calculated are: the accuracy, the frequency bias, the probability of detection, the false alarm ratio, the probability of false detection, the success ratio, the threat score, the Gilbert skill score, the Hansen and Kuipers discriminant and the Heidke skill score. As stated before, there is no a standard set of verification metrics. We followed previous works on similar verification analysis and we have selected common metrics (see e.g. Devos et al., 2014; Liemohn et al., 2018 and references there in). Information on these metrics and their mathematical equations are presented in Appendix II. The results of these metrics, which are mentioned above, as a function of the threshold value, are presented in Figs. 4 and 5.

The final diagram (Fig. 6), which is also an essential element in the analysis of the event detection assessment, is the receiver operating characteristic (ROC) curve that shows the ability of the forecast to discriminate between events and non-events. The ROC curve plot was created as follows: the probability of detection (POD) set on the y-axis and the probability of false detection (POFD) placed on the x-axis for all the threshold values of Ap geomagnetic index. The ROC curve plot for each SWPC is presented in Fig. 6. The ideal model curve should travel from bottom left to top left of the diagram and then across to the top right of the diagram. As the closer the curve is to the unity slope line represents no skill. The POFD values were set in ascending row in order to calculate the area under the curve for each SWPC, and the results are presented inside the parenthesis in the embedded table of Fig. 6. The area under the curve for a perfect score equals 1.

Discussion and Conclusions

The effects of space weather related phenomena has the potential to affect almost all of the technologies and activities in space as well as on the ground (Eastwood et al., 2017). In the last decade, many SWPCs were developed around the globe in order to provide as accurate as possible predictions of the conditions concerning the Space Weather. In the current work, the predictions of daily Ap geomagnetic index, which were issued by 7 SWPCs, are evaluated.

The predictions of Ap index cover the period of Oct. 2014 – Jul. 2020, i.e. 2131 days. This period consists of a part of the solar maximum of solar cycle 24 (SC24), the declining phase and the minimum between SC24 and SC25. At this point we mention that SC24 is one of the quietest solar cycles of the last 100 years and had fewer strong geomagnetic storms of G3-G4 level and none of severe geomagnetic storms of G5 level in contrast to the previous solar cycle. The maximum observed Ap value for the examined period was 108 nT. For comparison we are referring that since 1970 only three days had Ap greater than 200: a) 13/03/1989 with Ap = 246 nT, b) 29/10/2006 with Ap = 204 nT and 8/2/1986 with Ap = 202 nT. The fact that SC24 was a quieter cycle in the sense of solar activity in contrast to SC23 implies our verification analysis utilized data which are based more on quiet conditions rather than active ones. Thus, it is easier to predict Ap values for a quiet period than for an active one.

It is noted that in this work we do not focus on the methods used by the various SWPCs to make the forecasts of the Ap geomagnetic index. The goal of the current work is to provide to the community the current capabilities of international forecast centers to predict the daily Ap index as a potential tool for the forecast of geomagnetic storms. However, a brief description behind the methodologies followed by SWPCs in general and especially by ASWFC and IZMIRAN were provided to better understand the performance of the SWPCs. To meet this challenge, we performed a verification analysis of the forecasts using several metrics on the fit performance between the forecasted and the observed values, as well as applying a threshold and calculating various skill scores.

According to the results of the first part of our analysis using the relevant performance metrics, the SWPCs are divided into two groups. The first group, including the ASWFC, IZMIRAN, KAZ, and SIDC space weather centers had values closer to the optimal ones and the second group with the NOAA, SEPC, and SWS centers, where their values were not so close (see Tables 2, 3 and 4). In particular, the best correlation coefficient values were for IZMIRAN (r=0.79), ASWFC (r=0.79), SIDC (r=0.78) and KAZ (r=0.73) between the observed and the forecasted Ap values for the Day-0. The correlation coefficient values are decreased when examined the predicted Ap values for days 1 and 2. Furthermore, the best values for the fit performance metrics are: the best slope value (0.719) from SIDC, the best values of the intercept, the mean absolute error, the root mean square error and the skill score values (2.803, 2.981, 5.422, and 0.638 respectively) from ASWFC, the best mean error and bias values (-0.062 and 1.007 respectively) from KAZ.

Useful information about the distribution of the observed and the forecasted Ap values are provided in Figs. 2 and 3. Among the examined SWPCs, it seems that the histogram of ASWFC is the one that is almost identical to the histogram of the observed Ap values. In contrast to NOAA and SEPC where their forecasts are biased with a lower limit of Ap = 5 and Ap = 3 respectively. The maximum frequency for the observed values of Ap occurs at the bin [4-5) and it is obvious that the maximum frequency from ASWFC is coinciding with this bin. In contrary, for the IZMIRAN, SIDC, KAZ, and SWS Centers, the maximum

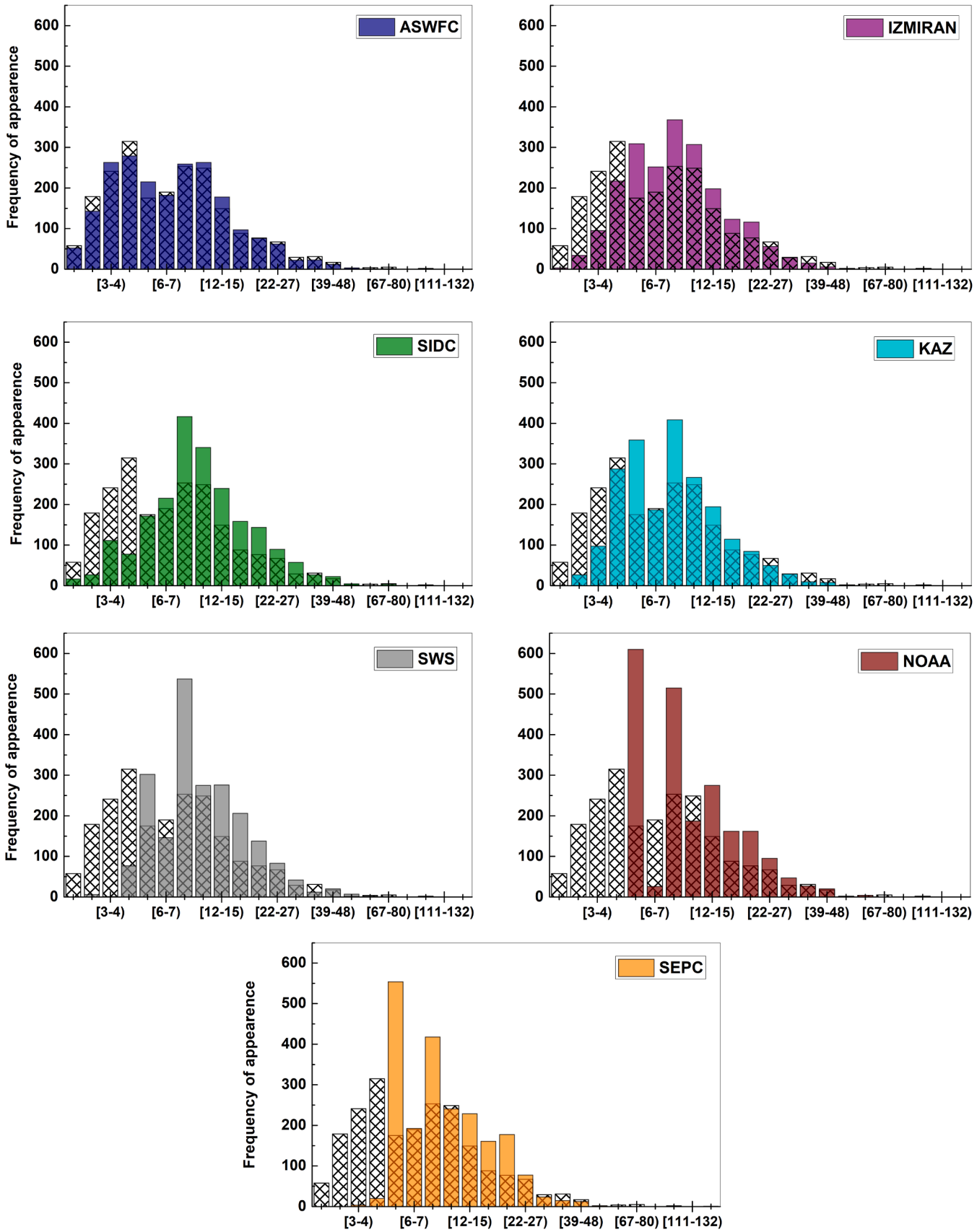

Fig. 3. Histograms of the frequency of appearance for each bin of Ap index for each one of the examined SWPCs. The observed Ap values for each bin are presented as a column with a pattern of crossed lines enabling the easier visual comparison between the observed and forecasted Ap values.

Table 4
Results of fit performance metrics between the forecasts (of Day-0) and the observed Ap values – Categorical Scores I.

Metric [*value for perfect performance*]	ASWFC	IZMIRAN	NOAA	SIDC	KAZ	SWS	SEPC
Slope [1]	0.637	0.607	0.492	**0.719**	0.510	0.473	0.519
Intercept [0]	**2.803**	4.053	6.842	5.003	4.362	7.042	5.910
Mean error [0]	0.386	-0.600	-2.384	-2.538	**-0.062**	-2.413	-1.685
Mean absolute error [0]	**2.981**	3.282	5.153	4.249	3.477	5.216	4.603
Root mean squared error [0]	**5.422**	5.437	7.945	6.299	6.104	7.987	7.248
Skill score [1]	**0.638**	0.636	0.223	0.511	0.541	0.214	0.353
Bias [1]	0.956	1.068	1.271	1.289	**1.007**	1.275	1.192

frequency is presented at the bin [7-9). NOAA and SEPC present a different behavior with the maximum at the bin [5-6). The latter result is an indication of over-forecasting from those two SWPCs in very quiet days where the Ap is ranging at very low levels, between 0 and 4. In particular, there are 793 days in a sample of 2131 days (37.2%) where Ap values were less than 5 indicating quiet or very quiet conditions.

As mentioned above, the Ap geomagnetic index is a continuous variable, and as a result, the building process of a contingency table is based on the use of a threshold value. For various threshold values starting from Ap = 1 up to Ap = 70, several metrics were estimated. Some of those were the accuracy, POD, FAR, POFD, the Gilbert skill score, the Hansen-Kuipers skill score, the Heidke skill score and the area under the curve from ROC plot. From Figs. 4 to 6, a similar result with that obtained from the fit performance metrics is illustrated. It is noted again that the studied SWPCs are also divided into two definite groups, where the ASWFC, IZMIRAN, KAZ, and SIDC Centers are showing a better performance in contrast to the NOAA, SEPC, and SWS ones.

In particular, the ASWFC has presented the best performance between all SWPCs for accuracy, FB, POD, and TS. Especially, a careful look into the four diagrams of Fig. 4 shows that ASWFC, IZMIRAN and KAZ are performing better than the other SWPCs. IZMIRAN presents the best performance for FAR and SR below the threshold of Ap = 5, while SIDC presents the best performance for the same metrics above the threshold of Ap = 5. For POFD, the SIDC has a better result with a line closer to the x-axis in contrast to the others. Excluding SIDC, above the threshold of Ap = 20, ASWFC has the next best performance. For ETS and HSS, it is evident that ASWFC, IZMIRAN, KAZ, and SIDC, showing a very good performance as the other SWPCs have graphs below the lower levels of KAZ. Notably, the maximum values for ETS were 0.503 for IZMIRAN and 0.495 for ASWFC for threshold values of Ap = 15 and Ap = 14, respectively. Above the threshold value of Ap = 25 SIDC has the best performance. ASWFC and SIDC are showing the best performance in the sense of HK score for threshold values less than 10 for the first and above 10 for the latter, respectively. The maximum value for HK was 0.735 for SIDC for a threshold value of Ap = 20. Finally, the maximum values for HSS were 0.67 and for IZMIRAN and ASWFC were 0.662 for threshold values of Ap = 15 and Ap = 14, respectively. Fig. 6 presents the receiving operating characteristic, as well as the calculated area under the curve (AUC) for each SWPC. It is noteworthy that the most significant AUC values were 0.913, 0.908, and 0.896 for SIDC, ASWFC, and IZMIRAN, respectively, while the lowest AUC values were 0.828 and 0.831 for SWS and NOAA respectively.

At this point, we should mention a possible "timing" issue which might explain the performance of the examined SWPCs. These centers are not providing their daily predictions at the same time. This affects the performance of each center crucially as it was presented in this research. There are two distinct categories where the SWPCs in the first one provide their predictions (Ap values for Days-0, 1, and 2) just some hours before the beginning of the first day – "Day-0" (NOAA, SWS, and SEPC). In the second category, the SWPCs provide their predictions some hours later, which means that they already know some of the 3-hour time intervals of the ap index (ASWFC, IZMIRAN, KAZ) or in the middle of "Day-0", which means that the prediction concerns only the rest half-day of Day-0 (SIDC).

In particular, NOAA provides its report daily at 22:00 UTC with predictions for the next three days, and it is published before the others. ASWFC, IZMIRAN and KAZ provide their reports between 05:00 and 08:00 UTC each day containing forecasts for the current day and the next two days, which means that it is already known the space weather conditions of interplanetary medium for the first 3-hour time interval and sometimes for the second 3-hour interval of ap preliminary values. SIDC provides its report around 12:30 UTC of each day. Thus the first four 3-hour time intervals of the ap index are available, making predictions much more manageable as the forecast actually holds for the next half of the current day. SWS and SEPC provide their report in the same period as the NOAA does. That explains in a point the fact that SWPCs which publish their reports at the early beginning (ASWFC, IZMIRAN, KAZ) or much later, in the middle of the day (SIDC) have much better results than other SWPCs which publish their reports at the end of the previous day (NOAA, SWS, SEPC). We do not have access to each SWPC forecasting pipeline, and we do not know which models and data they are actually using, as it is out of the scope of the current work. So, it is reasonable to assume that the time advantage, as we discussed above, is one of the possible explanations for the better performance of some SWPCs in contrast with the others (e.g., SIDC in comparison to NOAA).

Another factor, which affects the results, is the uncertainty of the initial parameters, such as the velocity of a high-speed stream (HSS) of solar wind emanating from coronal holes or the linear speed of a CME. For the periods near the solar minimum, the predictions of Ap based mainly on the HSS of solar wind originating from coronal holes (Tsurutani et al., 2006). In the case of an Earth-facing coronal hole and supposing a mean velocity of the solar wind of about 500 km/s, the expected arrival time of this HSS to Earth is about 80 hours later. If the HSS has a velocity of 700 km/s, the expected arrival time is about 60 hours later, thus 20 hours earlier than the previous scenario. That means the rise in the values of the Ap index will occur almost a day earlier, affecting the forecast crucially. In cases of ICMEs, there is much more uncertainty about the initial parameters. The most important uncertainties are associated with the linear speed of CME, which is a projection on the plane of the sky (see Paouris et al., 2021b and references therein) as well as the interaction of the CME with the ambient solar wind, making difficult the estimation of the ToA of the CME (see Paouris et al., 2021a; Paouris and Mavromichalaki, 2017b). This is crucial for Space Weather forecasting because if the CME arrives earlier (or later), the geomagnetic storm will be triggered at a different time than initially predicted, resulting in an inaccurate prediction of Ap for that day. In the current status, for the majority of SWPCs as well as for outputs from various space weather forecasting models, it seems to be also very important to have the "*human-in-the-loop*" contribution for the daily forecasts (see e.g. Devos et al., 2014; (Riley et al., 2018)). All these uncertainties could describe up to a point, the decreasing forecast skill observed in Table 2, where there is more confidence in the forecasts of the first day than of the third day.

A new effort concerning the forecast of the Ap index using shorter time intervals than daily values (e.g., forecasts for 6 and 12 hours) at the ASWFC is underway. In particular, new research using new advanced Machine Learning techniques for the estimation of the Ap geomagnetic index using various information as input, such as magnetograms, photospheric data, solar flares, and CMEs, will start in the next months.

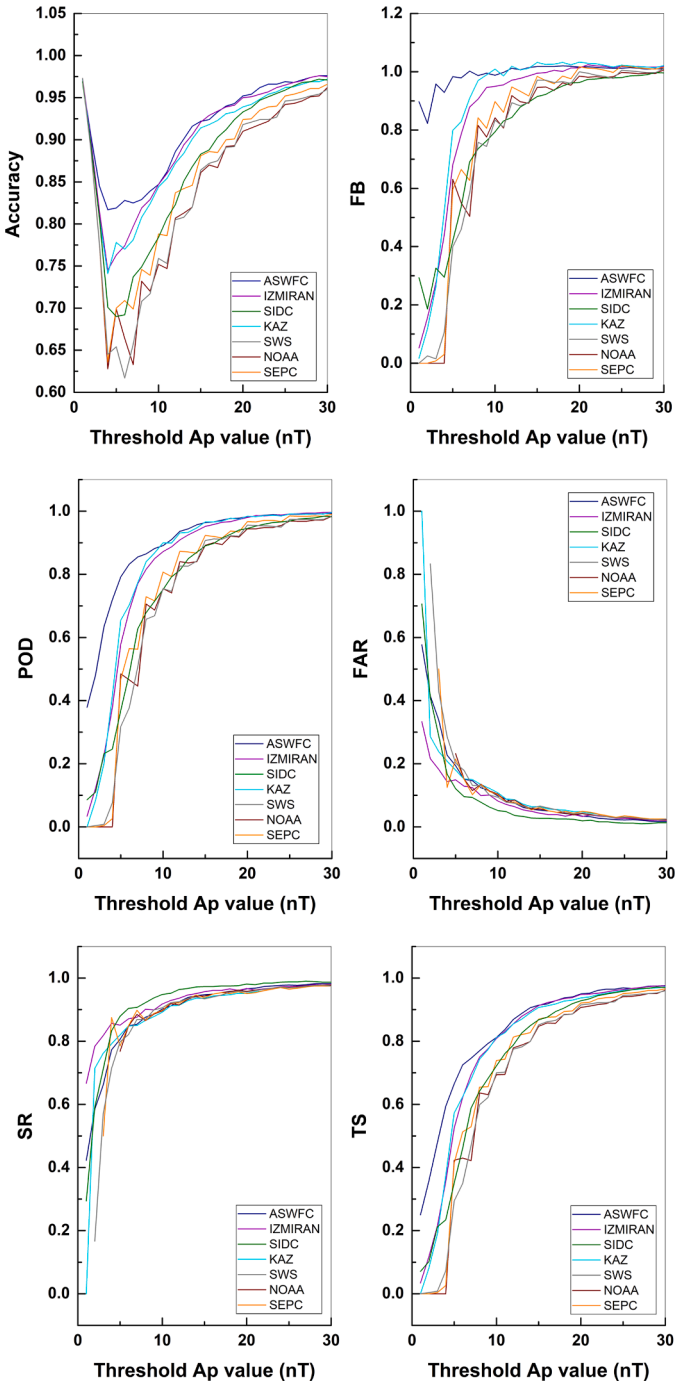

Fig. 4. Accuracy, frequency bias, probability of detection, false alarm ratio, success ratio and threat score as a function of the threshold value for all SWPCs concerning the first day of forecasts. For better visual presentation we focus on the threshold values between 0 and 30 as above that limit the changes in the plots are insignificant.

The ASWFC announces a new collaboration with the Hellenic National Meteorological Service, where ASWFC will be the responsible Center for Space Weather predictions and especially forecasting geomagnetic storms through Ap index.

CRediT authorship contribution statement

Evangelos Paouris: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing, Supervision, Visualization. **Maria Abunina:** Data curation, Investigation,

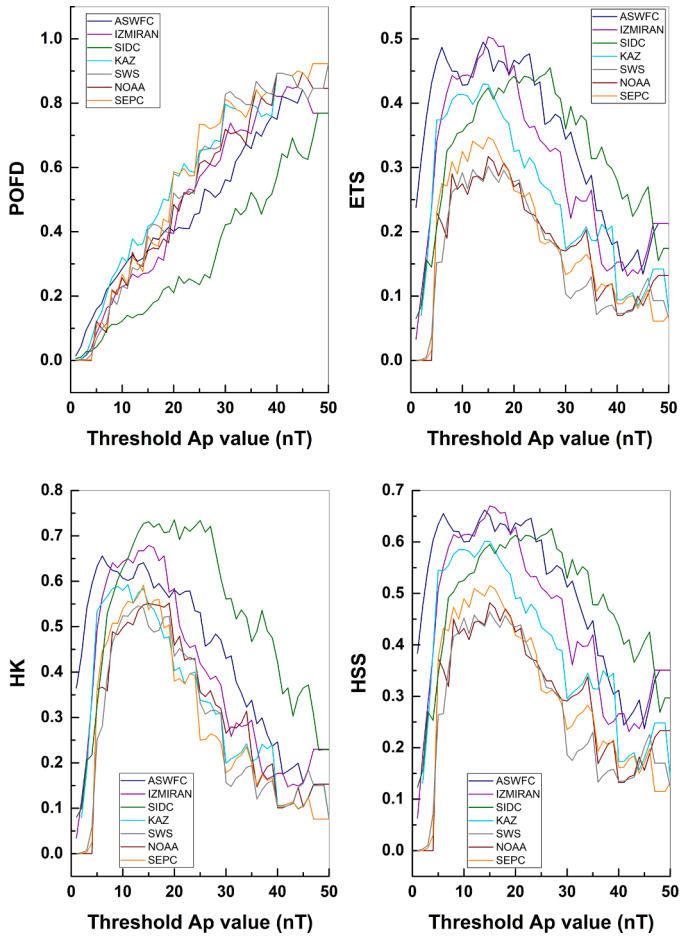

Fig. 5. Probability of false detection, equitable threat score, Hansen-Kuipers skill score and Haidke skill score as a function of the threshold value for all SWPCs concerning the first day of forecasts.

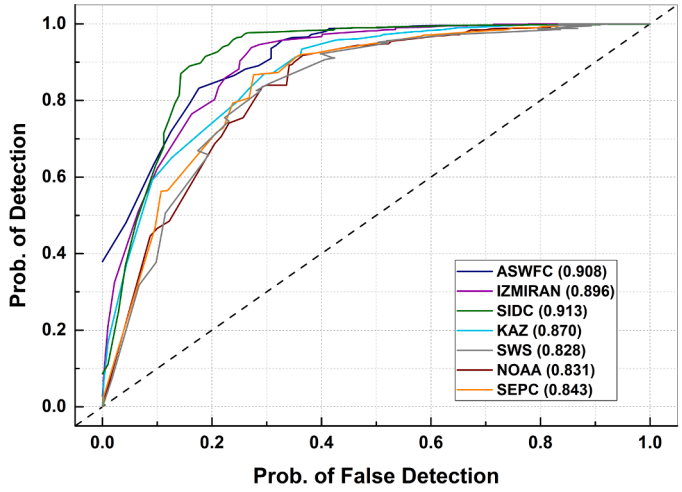

Fig. 6. The Receiver operating characteristic (ROC) curve for each SWPC is presented. The area under the curve for a perfect score should be equal to unity, while the unity slope (dashed) line represents no skill. The calculated area under the curve for each SWPC is presented inside a parenthesis on the legend of the plot.

Writing - review & editing. **Anatoly Belov:** Investigation, Writing - review & editing. **Helen Mavromichalaki:** Investigation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are grateful to the PIs and the groups of each one of the presented here Space Weather Prediction Centers for making their products available. In particular, we acknowledge the PIs and the duty forecasters of ASWFC, NOAA, SIDC, KAZ, SWS, and SEPC for their daily forecasts. Data used on the current research are available at the http://spaceweather.phys.uoa.gr/data.html repository of the ASWFC/NKUA that is supported by the Special Research Account of the National and Kapodistrian University of Athens (70/4/5803). This research is also supported by ESA SSA SWE Space Radiation Expert Service Centre activities (ESA contract number 4000113187/15/D/MRP) and the European Neutron Monitor Services research funded by the ESA SSA SN IV-3 Tender: RFQ/3-13556/12/D/MRP.

E.P. supported by the project "PROTEAS II" (MIS 5002515), which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund). M.A. and A.B. are supported by the Russian Science Foundation under grant 20-72-10023.

Appendix I. – Category I

The mathematical equations of the Category I which are presented in subsection 3.1 are illustrated in this appendix. These metrics are the linear equation between the forecasted and the observed values, the mean error, the mean average error, the root mean square error, skill score and the bias.

The forecast model (F) is predicting an observed (O) value, and these two must be associated through a linear relationship (Liemohn et al., 2018) of the form:

$$F = B{\cdot}O + A. \tag{A1}$$

As much as the slope (B) is closer to 1 and the intercept (A) is closer to zero, the better the model performs. The intercept reveals the bias at the lowest observational values. At the same time, the slope quantifies whether the trend of the model results with increasing observational values keeps pace with the observed increase, undershoots, or overshoots it (Liemohn et al., 2018).

The mean error (ME) between the forecasted values (F_i) and the observed ones (O_i) answers the question of what is the average forecast error. It ranges from minus infinity up to plus infinity with a perfect score at zero, and the equation calculates it:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i) \tag{A2}$$

while the mean average error is:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|F_i - O_i| \tag{A3}$$

and it ranges from 0 up to infinity with a perfect score at 0.

The root mean squared error (RMSE) answers to the question of what is the average magnitude of the forecast errors and defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2} \tag{A4}$$

RMSE ranges from zero up to infinity with a perfect score at zero. The square term inside the summation puts more considerable influence on significant errors than the smaller ones, something that is noticeable, especially at active periods.

The skill score (SS) is defined in terms of the mean squared error (MSE) as follows:

$$SS = 1 - \frac{MSE}{MSE_{ref}} \tag{A5}$$

ranging between negative infinity and +1. The equation defines MSE is:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2. \tag{A6}$$

The value of MSE_{ref} is the MSE of a model, which is used as a reference model, and in the current research is the persistence model. As a closer, the SS is to one, the better the model performs, with one corresponds to a perfect fit. A forecast with a skill score of 0 has the same MSE as the reference model, while a negative score reflects a performance worse than the reference model (Devos et al., 2014). Another metric is the bias, which is the degree of correspondence between the mean forecast and the mean observation. It indicates whether observations are overestimated with bias > 1 or underestimated with bias < 1.

$$bias = \frac{\frac{1}{N}\sum_{i=1}^{N}(F_i)}{\frac{1}{N}\sum_{i=1}^{N}(O_i)} \tag{A7}$$

and it is different from the frequency bias (Liemohn et al., 2018), which is used for categorical forecasts (Devos et al., 2014).

Appendix II. – Category II

The following metrics calculated using a threshold value to create a contingency table of "yes/no" events as it was described in subsection 3.2. The contingency table consists of four basic elements in each case: hit, false alarm, miss and correct negative. In particular:

- The number of the correct forecasted events or hits. When the predicted and observed values are less than or equal to the threshold value, this pair is considered as a "hit" (Term A).
- The number of false alarms. A false alarm is a forecast of an event, while no event was observed. When a forecasted value is less or equal with the threshold value, while the observed value was higher than the threshold, this pair is considered as a "false alarm" (Term B).
- The number of misses. A miss is an event that was not forecasted. In that case, the predicted value is higher than the threshold value, and the observed value is less or equal to the threshold value. This pair is considered as a "miss" (Term C).
- The number of true negatives or correct rejections. A correct rejection is a forecast of a non-event, while indeed, no event was observed. In that case, the forecasted and the observed values are greater than the threshold, and this pair is considered as "correct negatives" (Term D).

The sum of A+B+C+D is equal to the total number of pairs or the sample size (n).

These metrics are:

The accuracy (AC) or the fraction of correct events which is the ratio of the number of hits (A) and correct negatives (D) to the total number of forecasted events or sample size (n):

$$AC = \frac{A + D}{n} \tag{A8}$$

Accuracy ranges from 0 to 1, with 1 represents a perfect score.

The bias score or the frequency bias (FB) defined as the ratio of the number of forecasts (A+B) to the number of actual occurrences (A+C):

$$FB = \frac{A + B}{A + C} \tag{A9}$$

FB ranges from 0 to infinity, while the value of 1 represents the perfect score. If FB < 1, then the forecast system tends to under forecast or over forecast when FB > 1.

The probability of detection (POD) or the hit rate, which is the ratio of the number of hits (A) to the number of events (A+C):

$$POD = \frac{A}{A + C} \tag{A10}$$

POD ranges from 0 to 1, while 1 represents the perfect score.

The false alarm ratio (FAR) which is the ratio of the number of false alarms (B) to the number of event forecasts (A+B):

$$FAR = \frac{B}{A + B} \tag{A11}$$

FAR ranges from 0 to 1, while 0 represents the perfect score.

The probability of false detection (POFD) which is the ratio of the number of false alarms (B) to the number of observed non-events (B+D):

$$POFD = \frac{B}{B + D} \tag{A12}$$

POFD ranges from 0 to 1, while 0 is corresponding to a perfect score.

The success ratio (SR), which is the number of hits (A) to the number of event forecasts (A+B):

$$SR = \frac{A}{A + B} \tag{A13}$$

SR ranges from 0 to 1, while 1 represents the perfect score. SR is equal to SR = 1 – FAR.

The threat score (TS) or the critical success index, which is the ratio of the number of hits (A) to the number of hits, misses, and false alarms (A+C+B):

$$TS = \frac{A}{A + C + B} \tag{A14}$$

TS ranges from 0 to 1, while 0 indicates no skill, and 1 indicates perfect score and measures the fraction of observed and/or forecast events that were correctly predicted.

The equitable threat score (ETS) or Gilbert skill score (also denoted as GSS), which is the ratio of the difference between hits (A) and random hits (Ar) to the number of hits, misses and false alarms (A+C+B) minus the Ar:

$$ETS = \frac{A - Ar}{A + C + B - Ar} \tag{A15}$$

The relation given Ar is:

$$Ar = \frac{(A + C)(A + B)}{n} \tag{A16}$$

ETS measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance. ETS answers the question of how well did the forecast "yes" events correspond to the observed "yes" events. Actually, shows the hits due to chance. It ranges from –(1/3) to 1, while 0 indicates no skill, and 1 indicates a perfect score.

The Hanssen and Kuipers (HK) discriminant or the true skill statistic (TSS) or Peirce's skill score (PSS), which is the difference between POD and POFD:

$$HK = \frac{A}{A+C} - \frac{B}{B+D} \tag{A17}$$

HK ranges from -1 to 1, while 0 indicates no skill, and 1 indicates a perfect score. HK shows how well did the forecast separate the "yes" events from the "no" events.

The Heidke skill score (HSS) or Cohen's κ, which answers the question of what was the accuracy of the forecast relative to that of random chance, with a relation:

$$HSS = \frac{(A+D) - ECr}{n - ECr} \tag{A18}$$

where ECr is equal to:

$$ECr = \frac{1}{n}[(A+C)(A+B) + (D+C)(D+B)] \tag{A19}$$

HSS measures the fraction of correct forecasts after eliminating those forecasts, which would be correct due purely to random chance. HSS ranges from -1 to 1 while 0 indicates no skill, and 1 indicates a perfect score.

References

Bartels, J., Heck, N.H., Johnston, H.F., 1939. The three-hour range index measuring geomagnetic activity. J. Geophys. Res. 44, 411–454. https://doi.org/10.1029/TE044I004P00411.

Bartels, J., 1951. An attempt to standardize the daily international magnetic character figure, IATME Bull. 12e. International Union of Geodesy. and Geophysics 109. Publ. Off., Paris.

Bothmer, Volker, Daglis, Ioannis, 2007. Springer, https://doi.org/10.1007/978-3-540-34578-7.

Cade, W.B., III, C., Chan-Park, C., 2015. The Origin of "Space Weather. Space Weather 13, 99–103. https://doi.org/10.1002/2014SW001141.

Cui, Yanmei, Liu, Siqing, Aa, Ercha, Zhong, Qiuzhen, Luo, Bingxian, Ao, XianZhi., 2016. Verification of SPE probability forecasts at the Space Environment Prediction Center (SEPC). Science China Earth Sciences 59. https://doi.org/10.1007/s11430-016-5284-x.

Devos, A., Verbeeck, C., Robbrecht, E., 2014. Verification of space weather forecasting at the regional warning center in Belgium. Space Weather Space Clim 4 (27). https://doi.org/10.1051/swsc/2014025. A29.

Eastwood, J.P., Biffis, E., Hapgood, M.A., Green, L., Bisi, M.M., et al., 2017. Risk Analysis 37, 206. https://doi.org/10.1111/risa.12765.

Eroshenko, E.A., Belov, A.V., Boteler, D., Gaidash, S.P., Lobkov, S.L., et al., 2010. Effects of strong geomagnetic storms on Northern railways in Russia. Adv. in Space Res. 46, 9. https://doi.org/10.1016/j.asr.2010.05.017, 1102.

Gopalswamy, N., 2009. In: Tsuda, T., Fujii, R., Shibata, K., Geller, M.A. (Eds.). Terrapub, Tokyo, pp. 77–120.

Jolliffe, I.T., Stephenson, D.B., 2012. Forecast verification. A practitioners Guide in Atmospheric Science, 2nd ed. Wiley-Blackwell, p. 2012.

Kane, R.P., 2006. The idea of space weather. Adv. Space Res. 37, 1261. https://doi.org/10.1016/j.asr.2006.01.014.

Koskinen, H.E.J., 2011. Physics of Space Storms: From the Surface of the Sun to the Earth. Springer Praxis Books, New York, NY.

Kubo, Yuki, Den, Mitsue, Ishii, Mamoru, 2017. Verification of operational solar flare forecast: Case of Regional Warning Center Japan. Journal of Space Weather and Space Climate 7. https://doi.org/10.1051/swsc/2017018.

Liemohn, M.W., McCollough, J.P., Jordanova, V.K., Ngwira, C.M., Morley, S.K., Cid, C., et al., 2018. Model evaluation guidelines for geomagnetic index predictions. Space Weather 16, 2079–2102. https://doi.org/10.1029/2018SW002067.

Menvielle, M., Iyemori, T., Marchaudon, A., Nose, M., 2011. Geomagnetic indices. In: Mandea, M., Korte, M. (Eds.), Geomagnetic observations and models, IAGA Special Sopron Book Series, 5. Springer, Dordrecht, pp. 183–228, 10.1007/978-90-481-9858-0.

Murray, S. A., S. Bingham, M. Sharpe, and D. R. Jackson, 2017, Flare forecasting at the Met Office Space Weather Operations Centre, Space Weather, 15, 577-588, DOI: 10.1002/2016SW001579.

Paouris, E., 2013. Ineffectiveness of Narrow CMEs for Cosmic Ray Modulation. Solar Phys 284, 589–597. https://doi.org/10.1007/s11207-012-0166-7.

Paouris, E., Mavromichalaki, H., 2015. The CME-index for short-term estimation of Ap geomagnetic index based on the new ICME list. In: Proceedings of the 12th Hellenic Astronomical Conference, 28 June – 2 July, 2015, Thessaloniki, Greece. Available at: https://helas.gr/conf/2015/talks/S_1/paourisAbstract.pdf.

Paouris, E., Mavromichalaki, H., 2017a. Interplanetary Coronal Mass Ejections Resulting from Earth-Directed CMEs Using SOHO and ACE Combined Data During Solar Cycle 23. Solar Phys 292, 30. https://doi.org/10.1007/s11207-017-1050-2.

Paouris, E., Mavromichalaki, H., 2017b. Effective Acceleration Model for the Arrival Time of Interplanetary Shocks driven by Coronal Mass Ejections. Solar Phys 292, 180. https://doi.org/10.1007/s11207-017-1212-2.

Paouris, E., Čalogović, J., Dumbović, M., Mays, L., Vourlidas, A., Papaioannou, A., Anastasiadis, A., Balasis, G., 2021a. Propagating Conditions and the Time of ICME Arrival: A Comparison of the Effective Acceleration Model with ENLIL and DBEM Models. Solar Phys 296, 12. https://doi.org/10.1007/s11207-020-01747-4.

Paouris, E., Vourlidas, A., Papaioannou, A., Anastasiadis, A., 2021b. Assessing the projection correction of Coronal Mass Ejection speeds on Time-of-Arrival prediction performance using the Effective Acceleration Model. Space Weather 19 e2020SW002617. https:// doi.org/10.1029/2020SW002617.

Riley, Pete, Mays M., Leila, Andries, Jesse, Amerstorfer, Tanja, et al., 2018. Forecasting the Arrival Time of Coronal Mass Ejections: Analysis of the CCMC CME Scoreboard. Space Weather 16, 1245–1260. https://doi.org/10.1029/2018SW001962.

Rostoker, G., 1972. Geomagnetic Indices. Reviews of Geophysics and Space Physics 10 (4), 935–950. https://doi.org/10.1029/RG010i004p00935.

Schwenn, R., 2006. Space Weather: The Solar Perspective. Living Rev. Solar Phys. 3 (2) https://doi.org/10.12942/lrsp-2006-2.

Thomsen, M.F., 2004. Why Kp is such a good measure of magnetospheric convection. Space Weather 2, S11004. https://doi.org/10.1029/2004SW000089.

Tsurutani, B.T., et al., 2006. Corotating solar wind streams and recurrent geomagnetic activity: A review. J. Geophys. Res. 111, A07S01. https://doi.org/10.1029/2005JA011273.

Vybostokova, T., Svanda, M., 2019. Statistical analysis of the correlation between anomalies in the Czech electric power grid and geomagnetic activity. Space Weather 17. https://doi.org/10.1029/2019SW002181.