Iranian Journal of Geophysics, Vol 20 No 6, 2027, Page 37 – 48 (Research Article)
DOI: 10.30499/ijg.2025.479953.1637

The investigation into the impacts of solar wind-magnetosphere interactions on ionospheric electron flux using machine learning

Mehdi Heidari[1] and Seyed Majid Mir Rokni[2*]

[1] *Ph.D. Student, Department of Physics, Yazd University, Yazd, Iran*
[2] *Assistant Professor, Department of Physics, Yazd University, Yazd, Iran*

(Received: 24 September 2024, Accepted: 23 September 2025)

Summary

The solar wind plays a crucial role in Earth's magnetosphere, driving nonlinear interactions that result in geomagnetic storms and energy transfer to the ionosphere. This study investigates the relationship between solar wind parameters (e.g., speed, density, and interplanetary magnetic field components) and geomagnetic indices (e.g., AL, AE) with the electron energy flux entering the ionosphere. A ten-year satellite dataset (2004–2014), covering a complete solar cycle, was analyzed. Advanced machine learning techniques, including Random Forest, Support Vector Regression, Ridge Regression, and Principal Component Analysis, were employed to identify the most influential features affecting the electron energy flux. Additionally, preprocessing steps, such as outlier detection, normalization, and feature selection, were implemented to ensure optimal model performance. The analysis revealed that the importance of parameters dynamically varies during periods of solar minimum and maximum activity, with parameters such as Bz and the AL index having distinct effects on electron energy flux. This study not only presents an innovative approach to analyzing the complex nonlinear Sun-Earth interactions but also emphasizes the need for adaptive modeling techniques that account for temporal variations. The findings of this research provide a basis for more accurate space weather and climatology predictions, as well as for improving hybrid data-driven and physical models for future studies.

Keywords: Machine learning, solar wind, ionosphere, feature selection, space weather forecasting

*Corresponding author: mirrokni@yazd.ac.ir

1 Introduction

During solar activity, part of the Sun's mass is ejected as plasma from the corona, the outermost layer of the Sun's atmosphere. This high-energy and hot plasma is known as the "solar wind." The solar wind affects the Earth's atmosphere on a global scale (Islam et al., 2010).

The impact of the solar wind on the Earth's magnetic field results in heat transfer to the ionosphere. The solar wind causes ionization in the ionosphere. Minor changes in solar radiation lead to variations in the Earth's temperature and, ultimately, changes in the Earth's climate (Beer et al., 2010). Therefore, precise observation and analysis of temporal and spatial variations in solar wind parameters are essential for predicting extreme weather and climate events (Ohunakin et al., 2015).

In recent decades, numerous studies have investigated the impacts of the solar wind on the Earth's magnetic field and its consequences for the Earth's atmosphere (Hansteen & Velli, 2012; Parker, 1958). Despite decades of observation and research, the detailed mechanisms of the solar wind remain unclear. After Parker (1958) discovered supersonic solar winds, new scenarios were developed for modeling and explaining the solar wind's effects on the Earth's atmosphere. These models have difficulty describing certain observations (Hansteen & Velli, 2012). Exospheric models based on electric fields and thermal hydrodynamic models can explain slow solar winds (400 km/s); however, none of these models adequately describe the effects of fast solar winds (750 km/s) (Halekas et al., 2022).

Electron precipitation is a key component linking the ionosphere and the magnetosphere. Electrons carry current and transfer energy within the magnetosphere-ionosphere system. In other words, electrons follow magnetic field lines from the magnetosphere to the ionosphere, where they collide with the neutral atmosphere, causing changes in the conductivity tensor. This tensor represents the three-dimensional electric current circuit that flows over vast distances between the magnetosphere and the ionosphere. In fact, the precipitation of incoming particles is crucial for all global circulation models (Ridley et al., 2006).

In recent years, new techniques and algorithms have been developed to predict the consequences of electron precipitation, including traditional models such as empirical models and theoretical parameter-based approaches, as well as newer models that utilize machine learning and artificial intelligence. These models simulate and predict the effects of electron precipitation on the Earth's atmosphere using data from meteorological stations and satellites (Liu et al., 2019; Vardavas & Taylor, 2011).

Given the growing volume and diversity of ionospheric data, more accurate results are expected from theory-based models using modern techniques. Machine learning techniques offer better and more useful results for modeling and predicting the ionosphere using solar activity data (Benoit & Petry, 2021). Due to recent advancements, machine learning has created a wide range of models and techniques related to the Sun and its impacts on the Earth-atmosphere system. Selecting the appropriate configuration for these models directly affects their performance and often leads to a better understanding of model results and improved strategies for solving problems (Guo et al., 2022).

To study and make data-driven decisions using machine learning methods for calculating the total energy flux entering the ionosphere, one of the critical questions is which variables are most important for determining and predicting particle precipitation (Ohunakin et al., 2015). The present study aims to select features using various machine learning techniques to investigate the effects of the solar wind on the Earth-atmosphere system.

A distinctive feature of this study is the use of machine learning approaches to examine the effects of the solar wind over a solar cycle. The structure of the paper is as follows: the data and research methods used in the present study will be introduced in the section on data and methodology. The evaluation and analysis of the output from the applied models will be presented in the third section. Finally, the results of the present study will be detailed in the fourth section.

2 Data and Methodology

The data used in the present study is the electron energy flux based on satellite measurements. The dataset comes from the Defense Meteorological Satellite Program (DMSP) series, specifically satellites F13 to F16, which collect extensive data. Utilizing this dataset provides substantial information on solar wind parameters and geomagnetic parameters (Table 1).

Table 1: Operational Dates, Ascending/Descending Local Times (LT) Relative to the Equator, and Various Satellite Correction Factors (X) for the Northern Hemisphere (NH) and Southern Hemisphere (SH) for the DMSP Weather Satellites F13 to F16 [12].

Name	Start-Stop	Start-End A/D LT	SH/X1	NH/X1
DMSP-F13	95089-present	1711/0511-1830/0630 LT	0.99	1.01
DMSP-F14	97118-05094	2035/0835-1930/0730 LT	1.43	1.43
DMSP-F15	99351-present	2110/0910-2020/0820 LT	1.37	1.35
DMSP-F16	03300-present	2000/0800-2010/0810 LT	0.81	0.75

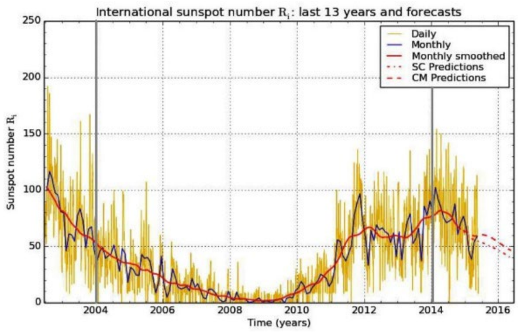

Figure 1. shows solar activity based on the international sunspot number over the past 13 years, presented as daily, monthly, and smoothed monthly averages, along with predictions for the period under study, which includes a complete solar cycle [16].

The dataset spans a 10-year period (2004-2014), covering a complete solar cycle, thus providing a well-sampled representation of the solar cycle. The collected data is prepared at five-minute intervals, and the input dataset includes 19 parameters of solar wind activity and magnetic indices.

This dataset includes a combination of solar wind parameters and geomagnetic index state descriptors (NASA Space Physics Center).. The combination of solar wind parameters and geomagnetic indices provides precise information about the entry and dispersion of electron energy flux in the ionosphere, as supported by prior studies (Borovsky & Denton, 2006; Guo et al., 2022; Kamide et al., 1998; McGranaghan et al., 2021; Ridley et al., 2006). In this study, we used a ten-year period (2004-2014) to cover the solar cycle. The reason for choosing this ten-year period is to obtain comprehensive data on solar magnetic activities and their effects on the magnetosphere and ionosphere, as well as to examine the response of various solar wind features and their effects on the Earth's atmosphere concerning solar activities. The data collection interval is five minutes.

These data were collected using standard methods, and to ensure quality, The data underwent preprocessing, including standardization, outlier removal, and normalization, to ensure their quality and suitability for statistical analysis and machine learning modeling. The cleaning processes included removing incomplete data, correcting obvious errors in the data, and smoothing the time series to ensure high data accuracy for subsequent analyses.

2.1 Data Preprocessing

Data preprocessing is a critical stage in any data analysis that affects the quality and accuracy of the research results. In this study, the dataset underwent several preprocessing steps to ensure that the data were suitable for statistical analyses and machine learning modeling.

1. **Data Cleaning:** The raw data were initially examined to identify and rectify potential deficiencies, including missing data, invalid values, and obvious errors. For any row where the amount of missing values exceeded 40% of the total columns, that row was removed. If less than this amount, the missing values were filled using the nearest neighbor method depending on the column type (García et al., 2015). After handling missing values, we addressed outliers. Outliers were identified and removed using the Interquartile Range (IQR) method to obtain a valid dataset. In this method, the dataset is divided into four quartiles (Q1, Q2, Q3, and Q4), each containing 25% of the sorted data. The IQR is the difference between Q3 and Q1:

$$IQR = Q3 - Q1 \tag{1}$$

Data points that fall below $Q1-1.5\times IQR$ or above $Q3+1.5\times IQR$ are considered outliers and were removed to ensure the dataset's integrity and reliability for analysis. These preprocessing steps ensured that the data were of high quality and ready for accurate and robust statistical and machine learning analyses.

2. **Normalization:** The data were normalized to ensure uniformity and comparability on an appropriate scale. This process involved converting the data to a common scale to reduce the impact of large variances across different scales (Han et al., 2022). For higher accuracy, especially since some algorithms are distance-based, normalization is recommended. Given that some independent variables have negative values, normalization was performed within the range of -1 to 1. After removing outliers and correcting missing values, the data, consisting of 1,424,826 rows and 9 columns, were normalized using the MinMax method.
In the next step, the dataset was split into subsets: a) training, b) validation and c) testing. Subsequently, various machine learning algorithms were applied. Finally, different statistical indices were used to quantify the accuracy of the predictive algorithms. The training set was used to build and train the supervised model. The validation set was used to fine-tune the model's hyperparameters. The test set was used to estimate the model's performance on data not used during training. The training set comprises 70% of the total data, the validation set 10%, and the test set 20%.

3. **Trend Analysis**: To analyze long-term trends and the behavior of key indices related to solar and geomagnetic activity, two parameters and the AL index were examined for the 2004–2014 period. Figures 2(a) and 2(b) illustrate the annual variations of these two parameters.

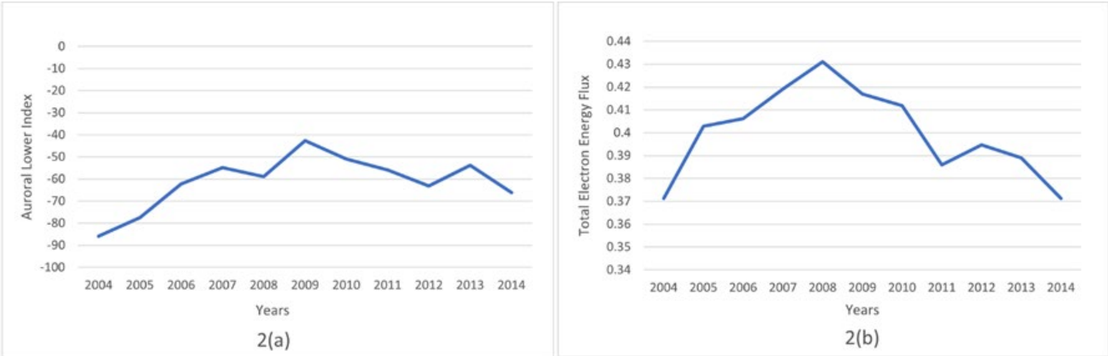

Figure 2. The figures depict the annual trends of the AL index (a) and Total Electron Energy Flux (b) from 2004 to 2014. Both parameters show an increasing trend from 2004 to 2008, followed by a decreasing trend after 2008. These trends highlight a significant.

These two graphs reveal a significant correlation between the AL index and the Total Electron Energy Flux. As observed, both parameters exhibit an increasing trend from 2004 to 2008, followed by a decreasing trend after 2008. This correlation reflects the dynamic relationship between the intensity of electric currents in the magnetosphere (represented by the AL index) and the energy of electrons entering the ionosphere.

2.2 Machine Learning Models

To analyze the data and select the effective features related to the solar wind and Earth's magnetic field, a set of machine learning models was utilized. These models were chosen for their outstanding capabilities in processing large volumes of data and uncovering complex patterns within the data. Feature selection in machine learning algorithms is applied to choose the best subset of variables that represent the original data, thereby reducing data size and model complexity and improving predictive performance (Mera-Gaona et al., 2021). The following is an explanation of these models:

Random Forest (RF): Random Forest, a leading algorithm in machine learning, consists of a collection of decision trees, each trained on a random sample of the data. This technique is popular for its high ability to reduce variance without increasing bias and is used for classification and regression, effectively handling data complexities (Breiman, 2001).

Support Vector Regression (SVR): In SVR, the goal is to find a function f(x) that keeps the prediction error within a specified margin of error ϵ, while minimizing the model's complexity.

Recursive Feature Elimination (RFE): RFE uses models that assign weights to features and eliminates less important features, improving model accuracy. This method is robust against redundant features and helps reduce model complexity (Guyon et al., 2002).

Ridge Regression: Also known as Tikhonov regularization, Ridge Regression manages high variance issues in linear regression models by adding a regularization term to the cost function (Hoerl & Kennard, 1970).

Least Absolute Shrinkage and Selection Operator (Lasso): Lasso is a regularization method that simultaneously reduces model complexity by penalizing the sum of the absolute values of the coefficients and performs feature selection (Tibshirani, 1996).

Principal Component Analysis (PCA): PCA is a statistical technique for reducing data dimensionality while preserving the main information, creating new components that have higher linear correlation with the data (Jolliffe, 2002).

Voting Average Technique (VOT): VOT is an ensemble method that combines predictions from multiple models. It uses the average of different predictions to achieve a more accurate and final prediction (Kittler et al., 1998).

Analysis and Discussion:

In the present study, we aim to understand how feature selection varies with changes in the solar wind and determine whether feature selection remains consistent across different years for different machine learning algorithms or if it changes with solar activities over different years. Feature selection can be aggregated using multiple feature selection algorithms to achieve better results (Solano et al., 2022). We examine the performance of various machine learning algorithms for feature selection across different years. The algorithms used are: a) Random Forest, b) Support Vector Regression, c) Recursive Feature Elimination, d) Ridge Regression, e) Lasso, and f) Principal Component Analysis.

Using the Voting Average Technique, we identify the most effective features in predicting the total electron energy flux entering the ionosphere across different years. The Voting Average Technique is an ensemble algorithm that combines predictions from several machine learning algorithms (An & Meng, 2010). In regression problems, VOT computes the average of all models to obtain a final prediction. By combining different models, the risk of poor performance by one model is mitigated by the strong performance of others, leading to better overall results. Since voting utilizes multiple machine learning algorithms, it is computationally more intensive.

In supervised learning, training samples are denoted by (x_i, y_i), and the function y = f(x) is used to predict unknown values, where x represents real-valued independent variables and y represents the dependent variable results. If the classifications that each learning algorithm performs for the training set S are denoted by h, the conditional probability distribution is defined as follows:

$$h(x) = P(f(x) = y|x, h) \quad (2)$$

Given a new data point x and a training sample S, the problem of predicting the value of f(x) can be considered as the problem of calculating:

$$P(f(x) = y|x, S) \quad (3)$$

This can be rewritten as a weighted sum over all hypotheses H, where H is the set of all formed hypotheses:

$$P(f(x) = y|x, S) = \sum_{h \in H} h(x)P(h|s) \quad (4)$$

This is an ensemble method in which the ensemble is composed of all hypotheses H, each weighted by the probability P(h|S) (Ridley et al., 2006). The necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its members is that the classifiers be both accurate and diverse. A classifier is accurate if its error rate is better than random guessing on new values of x. Two classifiers are diverse if they make different errors on new data points (Liu et al., 2019). To create this diversity, we used algorithms with different computational bases, such as decision-making and distance-based algorithms.

Predicting particle precipitation in the ionosphere poses many challenges for researchers striving to optimize these predictions. One of the most significant challenges is selecting features that can make predictions more accurate and faster. Additionally, the classification of solar wind parameters and geomagnetic indices must be precise.

After preprocessing the dataset, we used machine learning approaches to examine the validated dataset to select the best features for the dependent variable. A high number of input features may lead to poor performance of machine learning algorithms. In fact, the correlation between different solar wind features and various geomagnetic activity indices means that multiple features may contain information about others, and a high number of input features can only slow down the algorithm and increase computational error. The performance of the algorithm can be improved by removing features with less information (Parker, 1958).

In the present study, the total electron energy flux entering the ionosphere is the dependent variable, and solar wind data and geomagnetic indices, which describe the state of the ionosphere and magnetosphere, are the independent variables. Feature examination reveals linear and nonlinear relationships between the independent and dependent variables. Typically, the first step in scientific research, including space physics, is detection. Detecting effects and outcomes is the first step in identifying mechanisms and processes, which is the ultimate goal of scientific research.

For example, in Figure3, 15 years of solar wind data are classified at time intervals of 5, 10, 15, 30, 45 minutes, and 1, 3, 5, 6 hours using two algorithms, PCA and Random Forest. The results show that these two algorithms have different opinions in classifying input parameters.

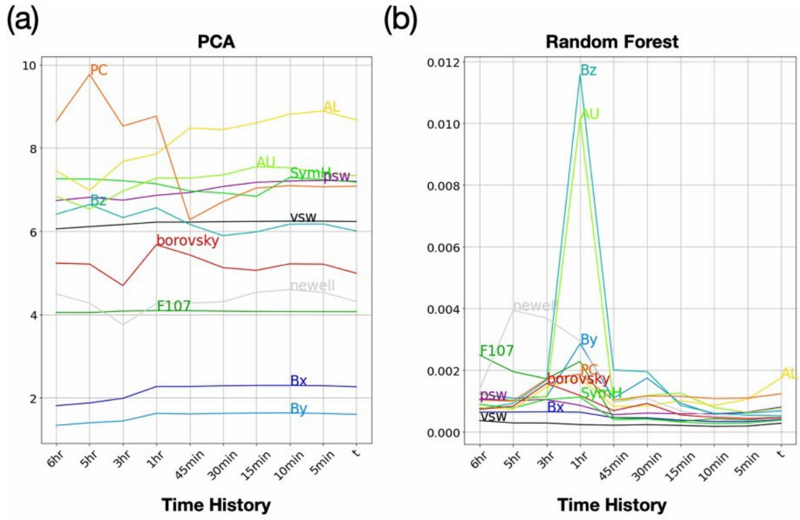

Figure 3. Classification of solar wind parameters and geomagnetic indices for different 15-year time frames using PCA and Random Forest methods.

The analysis of Figure 3 indicates that the importance of independent features for evaluation and use in PCA and Random Forest algorithms varies across different time intervals. For example, the PC feature has higher importance over other solar wind parameters and geomagnetic indices in the 1-hour intervals. In contrast, for 5-minute intervals, the AL parameter is more significant for the PCA algorithm. However, for the Random Forest algorithm, the Bz parameter is particularly important in the 1-hour interval, while in the 5-minute interval, the importance lies with the AL and PC parameters (McGranaghan et al., 2021).

Additionally, over a period of 15 years, various parameters hold the most value in different time intervals for the algorithms. Therefore, we proposed a group voting technique using the Voting Average Technique (VOT) among different algorithms. Instead of using ten consecutive years of data for optimal feature selection, we evaluated each year separately and plotted the timeline for each feature.

First, we performed a Pearson correlation analysis to reveal the correlation between variables (Han et al., 2022). Figure 4 shows the correlation matrix between the variables. A correlation coefficient (r = 0) indicates no linear relationship between variables, whereas a coefficient closer to 1 or -1 signifies a stronger relationship between the variables.

In Figure 4, each cell represents the Pearson correlation coefficient between two different parameters. The values within each cell indicate the magnitude and direction of the linear correlation between the parameters. For example, a correlation coefficient of 0.94 between AL and AE signifies a strong positive correlation. This means that an increase in the AE index is generally accompanied by an increase in the AL index. Both indices indicate geomagnetic activity, so it is expected that their variations align.

In the presented Pearson correlation chart, the correlation coefficient between the parameters By and Bz is -0.0018. This coefficient is very close to zero, indicating no significant linear correlation between these two parameters. In other words, changes in one of these parameters do not have a notable effect on the other. Practically, changes in By and Bz occur independently of each other. Although correlation coefficients close to zero indicate a lack of strong linear correlation, it does not imply the absence of any relationship between the parameters. There may be nonlinear relationships or more complex interactions that the Pearson correlation coefficient cannot detect.

The Pearson correlation coefficient only measures the linear correlation between parameters. This analysis can serve as a preliminary basis for further studies and more complex modeling in the field of solar wind and its impacts on the space environment and Earth.

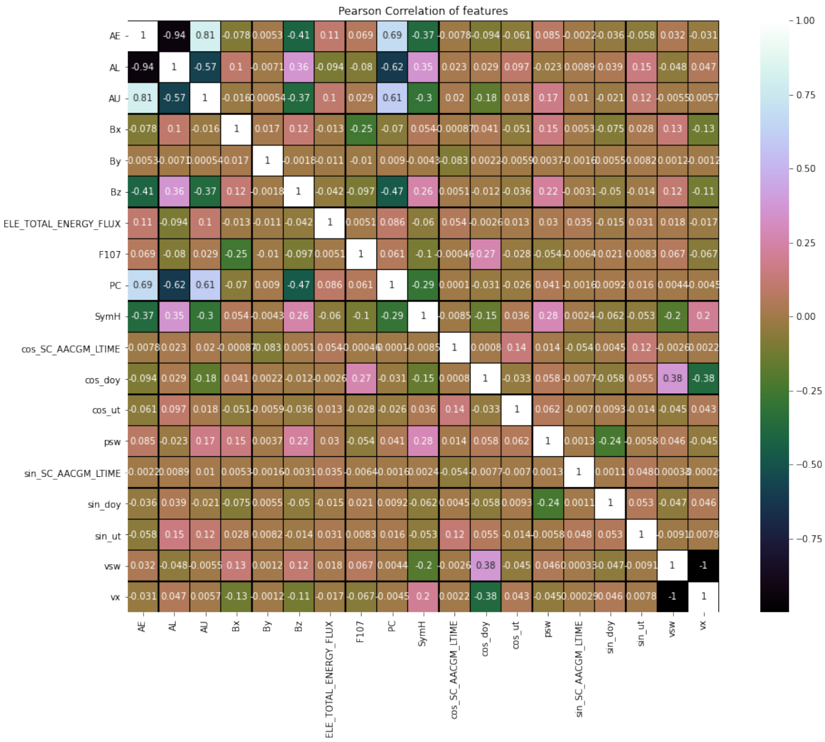

Figure 4. Pearson correlation for the independent and dependent variables of solar wind parameters and geomagnetic indices.

For further study using machine learning algorithms, the data for each year were evaluated separately using six different algorithms. Then, using the VOT technique, the results for each parameter were calculated separately, and their 10-year timelines were plotted to identify the independent parameters that are most important for calculating the total electron energy flux entering the ionosphere.

In this study, the data for each year were separately evaluated using six different machine learning algorithms, including Random Forest, Support Vector Regression, Recursive Feature Elimination, Ridge Regression, Least Absolute Shrinkage and Selection Operator, and Principal Component Analysis. This algorithmic diversity allowed us to leverage the strengths and capabilities of each method in data analysis to obtain more accurate results.

To combine and evaluate the results obtained from these six algorithms, we used the VOT (Voting Technique). This technique allowed us to consider the results of each algorithm separately and reach a reliable overall conclusion for each parameter. Using the VOT technique, the results for each parameter were calculated separately, and their ten-year trends were plotted and analyzed.

Identifying the independent parameters that are crucial for calculating the total electron energy flux entering the ionosphere is essential. The analyses showed that combining different algorithms and the voting technique led to more accurate and reliable identification of these parameters.

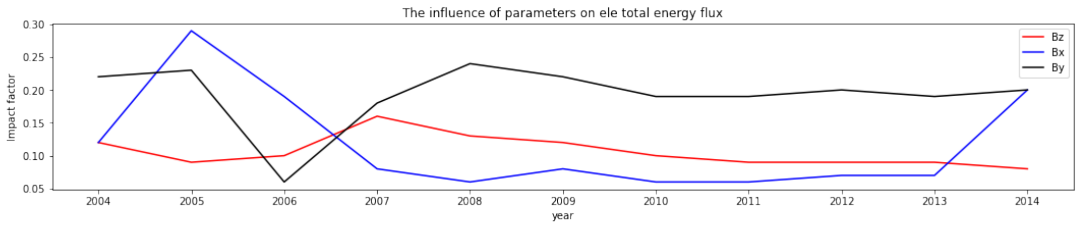

Figure 5. The impact of solar wind parameters BX, By, and Bz (components of the interplanetary magnetic field) on the total electron energy flux entering the ionosphere over a 10-year period from 2004 to 2014.

In Figure5, the influence coefficient of each magnetic field component on the electron energy flux is displayed on the vertical axis (y), and the years are displayed on the horizontal axis (x). The chart indicates significant variations in the influence of magnetic field components over the period under review. These variations are expected due to the dynamic nature of the solar wind and its interactions with the Earth's magnetosphere. The data covers a complete solar cycle characterized by alternating periods of solar maximum (around 2009-2014) and solar minimum (around 2004-2009).

The influence coefficient of the magnetic components shows more regular patterns during the solar maximum period and irregular behaviors during the solar minimum period. The importance of the Bx parameter shows significant variations, with a notable peak around 2005, followed by a sharp decline, indicating that Bx interacts more dynamically with the electron energy flux during years of decreasing solar activity than during years of increasing solar activity. The selection of the By feature shows variable importance during the years of declining solar activity from 2004 to 2009. During increasing solar activity, the importance of this parameter stabilizes with slight fluctuations, suggesting a more stable impact throughout the solar activity increase cycle. This indicates By's role in the stable conditions of solar wind interactions with the Earth's magnetosphere.

The Bz component shows a relatively steady influence over the period, with slight increases and decreases. It peaks around 2007 and then slightly declines but remains relatively constant. The stability in the Bz component's influence is notable, as Bz often plays a significant role in geomagnetic storms and energy transfer to the magnetosphere.

This chart demonstrates that the impact of solar wind magnetic field components on the electron energy flux significantly varies with solar activity. During the solar maximum period, the impact is more substantial and structured, whereas during the solar minimum period, it becomes more irregular.

In Figure 6, the impact coefficient of the geomagnetic indices AL, AU, and AE on the

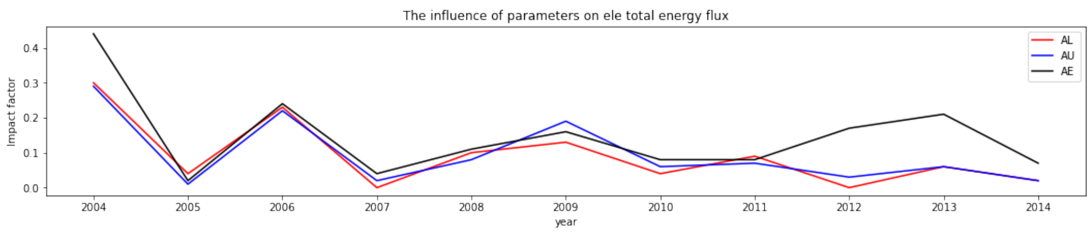

Figure 6. The impact of the geomagnetic indices AL, AU, and AE on the total electron energy flux entering the ionosphere over a 10-year period from 2004 to 2014.

electron energy flux is displayed on the vertical axis (y), with the years shown on the horizontal axis (x). The AL index indicates the intensity of geomagnetic currents moving towards the poles. The AU index represents the intensity of geomagnetic currents moving towards the equator, and the AE index is the sum of the AL and AU indices, defined as:

$$AL - AU = AE \tag{5}$$

This mathematical relationship succinctly shows the difference between the maximum and minimum magnetic deviations during geomagnetic storms, observed as geomagnetic currents in the auroral region. The AE index is used as an overall measure to assess the intensity of geomagnetic activities.

As indicated by the Pearson correlation coefficient and the above chart, these geomagnetic indices have a significant correlation with each other, and their importance in feature selection follows almost identical behaviors. Therefore, in feature selection tasks, one of these indices can be used to avoid overfitting and issues arising from an increased number of parameters in machine learning.

This correlation allows for more streamlined modeling, ensuring that the data is not overfitted, and simplifies the machine learning process without losing significant information on geomagnetic activity impacts.

In Figure7, for the parameters of solar wind speed and pressure, as illustrated in the chart, the importance of these features for measuring the total electron energy flux entering the ionosphere is more unstable and irregular during periods of decreasing solar activity. In contrast, during periods of increasing solar activity, the impact of these parameters is more stable and structured, suggesting a more significant and predictable influence during solar activity increases.

The multiple peaks observed in Figures 5, 6, and 7 during the 2004-2008 period, despite the overall decreasing trend of solar activity shown in Figure 1, can be attributed to several factors. First, short-term solar events such as Coronal Mass Ejections (CMEs) and Corotating Interaction Regions (CIRs) can temporarily enhance solar wind parameters (e.g., Vsw, Bz) and geomagnetic indices (e.g., AL, AE). These events occur sporadically multiple times within a year, leading to dynamic peaks in the datasets. Second, the non-linear interactions within the magnetosphere-ionosphere system can amplify these effects, resulting in complex variations over shorter timescales. Lastly, the year-to-year resolution of Figures 5, 6, and 7 captures less apparent fluctuations in the broader trends of solar activity shown in Figure 1. Previous studies have discussed such observations(Tsurutani et al., 2006), emphasizing the role of transient solar events and non-linear dynamics in driving geomagnetic variability.

Similar results were obtained for other solar wind parameters and geomagnetic activities, indicating that feature selection for creating machine learning models is dependent on solar activity. For models that consider a relatively long period, feature selection is particularly crucial, as the importance of features varies across the years studied.

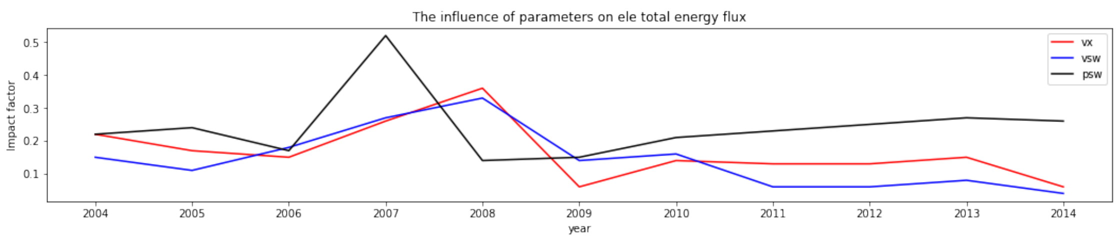

Figure 7. The impact of solar wind speed and pressure parameters on the total electron energy flux entering the ionosphere over a 10-year period from 2004 to 2014.

4 Conclusion

This study analyzed the influence of solar wind parameters and geomagnetic indices on electron energy flux entering the ionosphere using a comprehensive ten-year dataset (2004–2014). The results demonstrate that the importance of different features changes dynamically over time, reflecting the variability in solar activity phases. During periods of solar minimum, the behavior of features is more irregular, whereas in solar maximum, these features exhibit a more structured and predictable pattern.

By employing a diverse set of machine learning algorithms, including Random Forest, Support Vector Regression, Ridge Regression, Lasso, Principal Component Analysis (PCA), and the Voting Ensemble Technique, the study identified key parameters influencing electron energy flux. This methodological diversity enabled the integration of algorithmic strengths, improving predictive accuracy and providing a robust framework for feature selection.

A key finding of this research is the temporal variability in feature importance, which highlights the dynamic nature of Sun-Earth interactions. The study emphasizes the need for adaptive models that consider such variations to enhance the accuracy and reliability of predictions. The outcomes of this analysis provide practical insights for space weather modeling, particularly in forecasting geomagnetic storms and understanding their impacts on communication systems, navigation, and satellite technologies.

In conclusion, this research offers a novel perspective on the complexities of solar wind-magnetosphere interactions and their influence on ionospheric dynamics. The methodologies and findings presented herein lay the groundwork for future studies to integrate advanced machine learning techniques with physical modeling to achieve more accurate and comprehensive predictions in space science.

References

An, K., & Meng, J. (2010). Voting-averaged combination method for regressor ensemble. International Conference on Intelligent Computing,

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., & Bonan, G. B. (2010). Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science*, *329*(5993), 834-838.

Benoit, A. G. M. d. S., & Petry, A. (2021). Evaluation of F10. 7, Sunspot Number and Photon Flux Data for Ionosphere TEC Modeling and Prediction Using Machine Learning Techniques. *Atmosphere*, *12*(9), 1202.

Borovsky, J. E., & Denton, M. H. (2006). Differences between CME-driven storms and CIR-driven storms. *Journal of Geophysical Research: Space Physics*, *111*(A7).

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer.

Guo, Y., Ni, B., Fu, S., Wang, D., Shprits, Y., Zhelavskaya, I., Feng, M., & Guo, D. (2022). Identification of controlling geomagnetic and solar wind factors for magnetospheric chorus intensity using feature selection techniques. *Journal of Geophysical Research: Space Physics*, *127*(1), e2021JA029926.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, *46*, 389-422.

Halekas, J., Whittlesey, P., Larson, D., Maksimovic, M., Livi, R., Berthomier, M., Kasper, J., Case, A., Stevens, M., & Bale, S. (2022). The radial evolution of the solar wind as organized by electron distribution parameters. *The Astrophysical Journal*, *936*(1), 53.

Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.

Hansteen, V. H., & Velli, M. (2012). Solar wind models from the chromosphere to 1 AU. *Space Science Reviews*, *172*, 89-121.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Islam, M., Alili, A., Kubo, I., & Ohadi, M. (2010). Measurement of solar-energy (direct beam radiation) in Abu Dhabi, UAE. *Renewable Energy*, *35*(2), 515-519.

Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.

Kamide, Y., Baumjohann, W., Daglis, I., Gonzalez, W., Grande, M., Joselyn, J., McPherron, R., Phillips, J., Reeves, E., & Rostoker, G. (1998). Current understanding of magnetic storms: Storm-substorm relationships. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17705-17728.

Kittler, J., Hatef, M., Duin, R. P., & Matas, J. (1998). On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence*, *20*(3), 226-239.

Liu, Y., Tan, Q., & Pan, T. (2019). Determining the parameters of the Ångström-Prescott model for estimating solar radiation in different regions of China: Calibration and modeling. *Earth and Space Science*, *6*(10), 1976-1986.

McGranaghan, R. M., Ziegler, J., Bloch, T., Hatch, S., Camporeale, E., Lynch, K., Owens, M., Gjerloev, J., Zhang, B., & Skone, S. (2021). Toward a Next Generation Particle Precipitation Model: Mesoscale Prediction Through Machine Learning (a Case Study and Framework for Progress). *Space Weather*, *19*(6), e2020SW002684. https://doi.org/https://doi.org/10.1029/2020SW002684

Mera-Gaona, M., López, D. M., Vargas-Canas, R., & Neumann, U. (2021). Framework for the ensemble of feature selection methods. *Applied Sciences*, *11*(17), 8122.

Ohunakin, O. S., Adaramola, M. S., Oyewola, O. M., Matthew, O. J., & Fagbenle, R. O. (2015). The effect of climate change on solar radiation in Nigeria. *Solar Energy*, *116*, 272-286.

Parker, E. N. (1958). Dynamics of the interplanetary gas and magnetic fields. *Astrophysical Journal, vol. 128, p. 664*, *128*, 664.

Ridley, A., Deng, Y., & Toth, G. (2006). The global ionosphere–thermosphere model. *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(8), 839-864.

Solano, E. S., Dehghanian, P., & Affonso, C. M. (2022). Solar radiation forecasting using machine learning and ensemble feature selection. *Energies*, *15*(19), 7049.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.

Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L. C., Guarnieri, F. L., Gopalswamy, N., Grande, M., Kamide, Y., Kasahara, Y., Lu, G., Mann, I., McPherron, R., Soraas, F., & Vasyliunas, V. (2006). Corotating solar wind streams and recurrent geomagnetic activity: A review. *Journal of Geophysical Research: Space Physics*, *111*(A7). https://doi.org/https://doi.org/10.1029/2005JA011273

Vardavas, I., & Taylor, F. (2011). *Radiation and Climate: Atmospheric energy budget from satellite remote sensing* (Vol. 138). International Monographs on Ph.