Space Weather

RESEARCH ARTICLE
10.1029/2024SW004236

Imbalanced Regression Artificial Neural Network Model for Auroral Electrojet Indices (IRANNA): Can We Predict Strong Events?

Xiangning Chu[1], Lucas Jia[2], Robert L. McPherron[3], Xinlin Li[1,4], and Jacob Bortnik[5]

[1]Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, CO, USA, [2]Department of Electrical & Computer Engineering, University of California Santa Barbara, Santa Barbara, CA, USA, [3]Department of Earth, Planetary, and Space Sciences, University of California Los Angeles, Los Angeles, CA, USA, [4]Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, USA, [5]Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, CA, USA

Key Points:

- Imbalanced Regression Artificial Neural Network model for Auroral Electrojet Index (IRANNA) addresses data imbalance
- The IRANNA model can predict strong-to-extreme geomagnetic events for the first time, especially their peak amplitudes
- IRANNA can provide physical insights into the magnetospheric response to solar wind driving and potential underlying triggering mechanisms

Correspondence to:
X. Chu,
chuxiangning@gmail.com

Citation:
Chu, X., Jia, L., McPherron, R. L., Li, X., & Bortnik, J. (2025). Imbalanced Regression Artificial Neural Network model for auroral electrojet indices (IRANNA): Can we predict strong events? *Space Weather*, *23*, e2024SW004236. https://doi.org/10.1029/2024SW004236

Received 2 NOV 2024
Accepted 1 MAY 2025

Author Contributions:
Conceptualization: Xiangning Chu
Data curation: Xiangning Chu, Lucas Jia, Xinlin Li
Formal analysis: Xiangning Chu
Funding acquisition: Xiangning Chu
Investigation: Xiangning Chu
Methodology: Xiangning Chu, Robert L. McPherron, Jacob Bortnik
Project administration: Xiangning Chu
Resources: Xiangning Chu
Software: Xiangning Chu, Lucas Jia
Supervision: Xiangning Chu
Validation: Xiangning Chu, Robert L. McPherron, Xinlin Li
Visualization: Xiangning Chu
Writing – original draft: Xiangning Chu
Writing – review & editing: Xiangning Chu, Lucas Jia, Robert L. McPherron, Xinlin Li, Jacob Bortnik

Abstract We develop an Imbalanced Regression Artificial Neural Network model for the Auroral electrojet index (IRANNA) to predict the SuperMAG SML index, addressing the heavily imbalanced distribution of the SML data set. The data set contains mostly quiet-time values of lesser importance and very few strong-to-extreme values of interest, such as those associated with super substorms. Traditional prediction models, which minimize mean squared error uniformly across the whole data set, are often skewed by this imbalance, prioritizing the lower, quiet-time values and consequently underestimating strong geomagnetic events. The IRANNA model addresses this issue by using a customized weighting scheme in the loss function, enabling it to predict strong-to-extreme events accurately for the first time. The model takes solar wind parameters as inputs and predicts the logarithm of the absolute SML values. It does not rely on past values of the SML index, differentiating it from other models that use historical data for prediction. The model has demonstrated its ability to predict the peak amplitudes of strong-to-extreme events across various statistical analyses, event studies, and virtual experiments. Despite this success, challenges remain, particularly during localized electrojet events and when upstream solar wind data propagation is unreliable. This study emphasizes the importance of using imbalanced regression techniques, especially in space physics, where data sets are inherently skewed. It also highlights the potential of the IRANNA model to provide valuable insights into the magnetosphere's response to solar wind driving, improving space weather forecasting and offering new tools for investigating magnetospheric dynamics.

Plain Language Summary We developed an Imbalanced Regression Artificial Neural Network model for Auroral electrojet indices (IRANNA) to predict SuperMAG SML index. This index measures electrical currents in the upper atmosphere during magnetospheric activity, particularly substorms, which can affect satellites, power grids, and other technologies. The challenge is that most SML data reflect quiet periods of lesser importance, with few data points representing powerful events like super substorms. Traditional models often cannot predict these rare, strong events. Our IRANNA model solves this problem by deliberately prioritizing extreme events. This allows it to accurately predict when these powerful events will occur and how strong they will be. Unlike other models, ours does not rely on past values of the SML index and, therefore, focuses on solar wind conditions. While the IRANNA model has shown that it can accurately predict the intensity of super events, some challenges remain when predicting very concentrated magnetic disturbances or when solar wind data are inaccurate. This study highlights the importance of using advanced methods to handle uneven data sets, especially in space weather forecasting. It also provides new ways to study how the Earth's magnetosphere responds to the solar wind, improving our understanding of space weather events.

1. Introduction

1.1. Introduction to Auroral Electrojet Indices

The auroral Electrojet indices (AE) are critical for understanding the magnetosphere's response to solar wind driving, particularly through tail magnetic reconnection and substorm-associated dynamics (Davis and Sugiura, 1966). When solar wind energy is transferred into the magnetosphere via dayside reconnection, it is stored as magnetic energy in the magnetotail lobe (Dungey, 1961). The stored energy accumulates progressively

in the lobe, leading to magnetic reconnection in the magnetotail, which provides energy and flux for geomagnetic activity such as substorms (Angelopoulos et al., 2008; Pu et al., 2010). Earthward fast flows generated by the reconnection bring energy and magnetic flux toward the Earth, then slow down, enhancing currents in the near-Earth region such as the substorm current wedge (Angelopoulos et al., 2008; Baker et al., 1996; Baumjohann & Nakamura, 2001; Birn & Hesse, 2014; Chu et al., 2015; Hones et al., 1970; Kepko et al., 2015; McPherron, 1970; McPherron et al., 1973, 2011; Xing et al., 2009, 2011; Yao et al., 2012). The westward auroral electrojet, the ionospheric part connecting the current systems in the near-Earth region, induces magnetic perturbations on the ground. The auroral electrojet indices are calculated as the envelope of the northward magnetic perturbations measured by ground-based magnetometers distributed across the auroral zone (Davis and Sugiura, 1966). Specifically, the auroral electrojet upper index (AU) reflects the eastward electrojet, the lower index (AL) captures the strength of the westward electrojet, and AE is the sum of AU and AL, representing the overall intensity of auroral electrojet activity. The auroral electrojet indices capture enhanced auroral electrojet activity associated with a variety of geomagnetic activity, including substorms, Steady Magnetospheric Convection events (Kissinger et al., 2011; Pytte et al., 1978; Sergeev et al., 1996; Sergeev and Lennartsson, 1988), High-Intensity, Long-Duration, Continuous AE Activity (HILDCAA) events (Gonzalez et al., 1995; Kamide et al., 1998; Tsurutani and Gonzalez, 1987), sawtooth injections (Borovsky et al., 1993; X. Cai et al., 2006; DeJong and Clauer, 2005; Henderson et al., 2006; Huang et al., 2003), and other activity such as enhanced convection. The geomagnetic activity, as captured by the indices, not only controls the dynamics of the tail region but also contributes to a wide range of phenomena across other regions of the magnetosphere, influencing processes such as radiation belt particle dynamics (Fok et al., 2001; Jaynes et al., 2015), ionospheric currents, and plasmaspheric dynamics (Goldstein, 2007; Maruyama, 2020).

1.2. Previous Prediction Models and Their Potential Problems

Numerous prediction models have been developed to forecast geomagnetic indices, including Dst (e.g., Temerin and Li, 2002, 2006), auroral electrojet indices (AE, AU, AL), and Kp (e.g., Chakraborty & Morley, 2020). These models are essential for space weather forecasting and provide insights into the solar wind driving of geomagnetic activity. Among these indices, the Dst index has been successfully predicted in past models (e.g., Temerin and Li, 2002, 2006), likely due to its 1-hr resolution, which results in fewer data points and a smoother, well-defined pattern during geomagnetic storms, which typically last for several days. In contrast, predicting auroral electrojet indices has proven significantly more challenging. Substorm activity, captured by these indices, is relatively short-lived, with the expansion phase lasting only 10–15 min (Chu et al., 2015) and the entire substorm period spanning 1–3 hr. During the expansion phase, the auroral electrojet index can intensify dramatically, from near zero to over 1,000 nT, making it much more volatile. Additionally, the auroral electrojet indices are noisy due to the influence of many contributing factors, resulting in a highly imbalanced data set.

These prediction models for auroral electrojet indices can be categorized based on the technology used for prediction: (a) empirically derived equations (Li, Oh et al., 2007; Li, Baker et al., 2003; Luo et al., 2013), (b) linear prediction filters (Bargatze et al., 1985; Blanchard & McPherron, 1995; Clauer et al., 1981; Iyemori et al., 1979; Klimas et al., 1997; McPherron et al., 1988, 2013, 2015), (c) ARMAX and NARX models (Amariutei & Ganushkina, 2012; L. Cai et al., 2010; Gu et al., 2019), (d) Support Vector Machines (Gavrishchaka & Ganguli, 2001a), (e) Multivariate Relevance Vector Machines (Andriyas & Andriyas, 2017), (f) Artificial Neural Networks (Amata et al., 2009; Bala & Reiff, 2012; Ferreira & Borges, 2021; Gavrishchaka & Ganguli, 2001b; Gleisner & Lundstedt, 1997, 1999, 2001; Gopinath & Prince, 2019; Hernandez et al., 1993; Nakano & Kataoka, 2022; Pallocchia et al., 2008; Takalo & Timonen, 1997; Weigel et al., 1999), (g) Recurrent Neural Networks (RNN) and closely related Long Short-Term Memory (LSTM) models (Wintoft & Wik, 2021; Zou et al., 2024), and (h) nonlinear dynamical models (Chen & Sharma, 2006; Klimas et al., 1992). However, strong-to-extreme events (e.g., super substorms), which are particularly important for understanding magnetospheric dynamics, remain challenging for most models to predict, as they tend to underestimate these extreme events.

Prediction models can be categorized into two broad groups based on whether or not the input parameters include past values of the target output, in our case, the auroral electrojet SML index. Solar wind-driven models rely exclusively on solar wind or other external parameters, with no past SML values as input. Thus, solar wind-driven models are a function of instantaneous or time series solar wind parameters represented as $\mathrm{SML}(t) = f(\mathrm{sw}(t), \mathrm{sw}(t-1)\ldots)$, where $\mathrm{sw}(t)$ represents the upstream solar wind observations at time t, such as speed, pressure, magnetic field, and coupling function. On the other hand, similar to autoregressive and

persistence models, history-dependent models incorporate past values of the target output alongside solar wind parameters to predict current or future values. Thus, history-dependent models can be expressed as $\mathrm{SML}(t) = f(\mathrm{SML}(t-1), \mathrm{SML}(t-2), \ldots, \mathrm{sw}(t), \mathrm{sw}(t-1)\ldots)$. History-dependent models emphasize the persistence and patterns within the time series of geomagnetic indices. In contrast, solar-wind-driven models focus on the geomagnetic activity's direct response to upstream solar wind conditions. In this study, we develop a solar wind-driven model, focusing on the magnetospheric response purely based on solar wind driving, without relying on past geomagnetic indices.

Several challenges have been identified in the prediction of auroral electrojet indices in past studies listed above, stemming from both model limitations and the complex nature of the underlying physical processes. These challenges include:

1. Data Imbalance and Skewed Distributions: Auroral electrojet index values are highly imbalanced, with predominating quiet-time values (of lesser interest) and relatively few extreme events (of immense importance). This imbalance causes traditional models, especially those using mean squared error (MSE), to be dominated by quiet-time data, leading to the underestimation of strong substorm events.
2. Difficulty in Capturing Short-Term Variations: Many models, including empirical and machine learning approaches, perform well when predicting long-term or averaged trends in auroral electrojet indices. However, short-term variations, such as those occurring during substorm expansion phases, remain challenging to predict due to their rapid, localized, and complex dynamics. This challenge is compounded by the highly variable nature of substorms, where the index can surge from near zero to over 1,000 nT in a short period. This issue also relates to data imbalance, as models trained on such imbalanced data sets tend to predict smooth, quiet-time variations and underestimate extreme events. Our model addresses this issue by handling the data imbalance more effectively.
3. Propagation of Solar Wind Errors: The accuracy of solar wind parameters at Earth's nominal bow shock, especially during their propagation from the L1 point to Earth's magnetosphere, is crucial for prediction models. Propagation errors can lead to inaccuracies in predicting auroral electrojet indices. In particular, the interplanetary magnetic field (IMF) Bz component, which significantly influences geomagnetic activity, is challenging to predict accurately, contributing to large uncertainties in the resulting auroral indices.
4. Local versus Global Phenomena: Some models struggle to distinguish between global magnetospheric responses and localized electrojet events. For example, localized electrojets dominating the auroral electrojet index during a specific event may not reflect broader substorm dynamics, leading to prediction discrepancies.
5. Nonlinearity of Magnetospheric Response (for linear models): The magnetosphere's response to solar wind variations is generally not linear. Different levels of geomagnetic activity result in varied response properties, limiting the ability of linear models to fully capture the magnetosphere's complex behavior.

1.3. Imbalanced Regression Technique

In addition to predicting geomagnetic indices, many problems in space physics and other real-world applications inherently involve data imbalance and require supervised regression techniques. A significant challenge arises when regression models encounter imbalanced, non-uniformly distributed data sets. Such data sets typically exhibit highly skewed distributions (e.g., with a long tail), where specific ranges (e.g., weak, quiet-time values) are overrepresented but of lesser interest, while critical ranges (e.g., intense, active-time values) have far fewer data samples. This imbalance poses substantial challenges: traditional regression models, typically trained to minimize MSE, are disproportionately influenced by the abundance of quiet-time data (e.g., Figure 1 in Chu et al., 2024). As a result, these models tend to "regress to the mean," where large active-time values are significantly underestimated and quiet-time data are overestimated. This leads to unrealistic biases, particularly when predicting strong-to-extreme values and events in space physics, which are typically of the greatest importance.

In this study, we adopt the imbalanced regression method proposed by Chu et al. (2023, 2024), previously applied to a highly imbalanced data set of whistler-mode chorus and hiss waves. This approach successfully predicted both quiet-time background values and strong chorus waves. Similarly, the SML database exhibits a pronounced imbalance, with far more quiet-time background values than strong-to-extreme geomagnetic events, which are of greater interest for space physics studies. To address this, we adopt an imbalanced regression artificial neural network model tailored specifically for the auroral electrojet index from SuperMAG (SML), enabling more

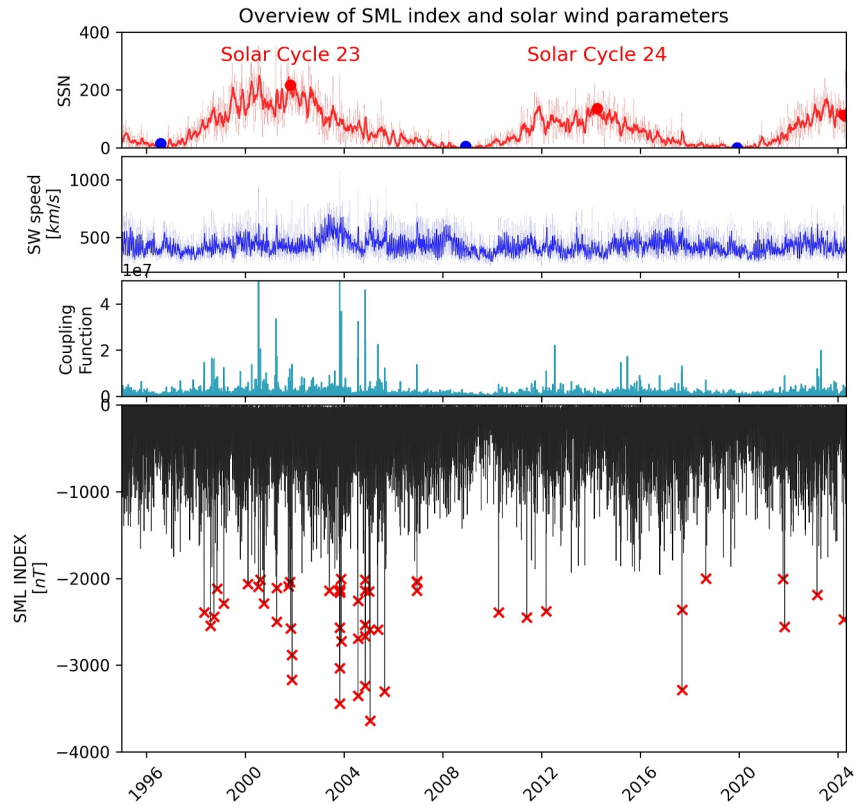

Figure 1. Overview of the SML index and solar wind parameters from 1 January 1995, to 1 May 2024. (a) and (b) show the daily (thin) and 30-day averages (thick) of sunspot number and solar wind speed (SW speed). (c) and (d) show the daily values of the coupling function ($v^2 B_{tot}^{1/2} B_T \sin^6(\theta_c/2)$) and 1-min SML with super substorms (|SML| > 2,000 nT) marked by red crosses.

accurate predictions of strong-to-extreme geomagnetic events (compared to traditional MSE-based models) for the first time.

2. Data Description

In this study, we utilize solar wind parameters obtained from the OMNI data set, available at a 1-min resolution (Papitashvili and King, 2020), alongside auroral electrojet indices from SuperMAG, specifically the SML index, also at a 1-min resolution (Gjerloev, 2012). The OMNI data set provides essential solar wind measurements, such as the IMF, solar wind speed, proton density, and dynamic pressure—key variables for understanding magnetospheric dynamics and their influence on auroral activity. The SML index, representing the intensity of the westward auroral electrojets, is derived from a network of ground-based magnetometers distributed across the auroral zone, making it a critical tool for tracking geomagnetic disturbances. We use the SML index over the World Data Center's AL index due to the latter's reduced data availability in recent years, particularly during periods of intense geomagnetic activity, which are of special interest for this study.

2.1. Distribution Analysis of SML

Figure 1 illustrates the SML index and key solar wind drivers over 30 years, from 1 January 1995, to 1 May 2024, covering approximately three solar cycles. Figures 1a and 1b display the daily (thin) and 30-day averages (thick) of the sunspot number (SSN) and the solar wind speed. Figure 1c shows the daily values of the coupling function ($v^2 B_{tot}^{1/2} B_T \sin^6(\theta_c/2)$) (McPherron et al., 2015; Newell et al., 2007), with solar minima and maxima indicated. In Figure 1d, the 1-min SML index is plotted to indicate geomagnetic activity, with super substorms defined as events where |SML| exceeds 2,000 nT, representing extreme events. The geomagnetic activity closely correlates with solar activity, as reflected in the SSN and extreme values in the coupling function. Solar cycle 23 exhibits stronger solar and geomagnetic activity, characterized by a higher frequency of super substorms. In addition, there

10.1029/2024SW004236

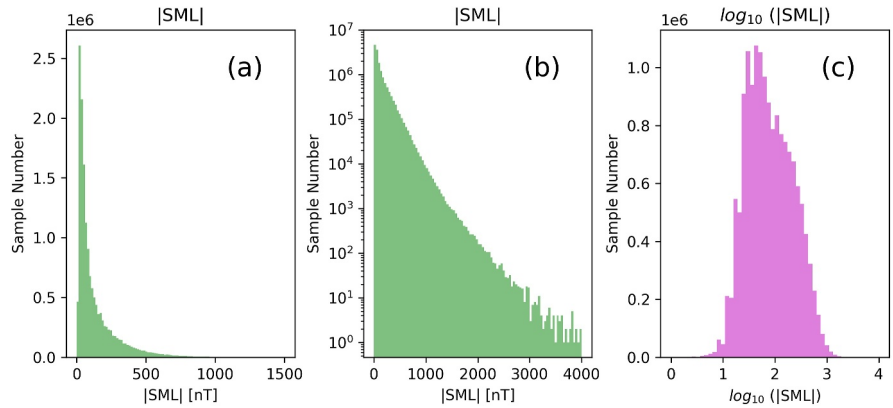

Figure 2. Statistical distribution of the SML index. (a) shows the histogram of |SML|, revealing a long-tailed distribution with a large number of low values. (b) Plots the same data on a logarithmic scale, emphasizing the exponential decrease in sample numbers with increasing |SML|. (c) Presents the histogram of $\log_{10}$(|SML|), which produces a more balanced bell-shaped distribution, justifying the choice to predict $\log_{10}$(|SML|) rather than |SML| directly.

are more super substorms during the solar declining phase than the rising phase, which aligns with findings by Chu et al. (2015), Hajra et al. (2016), and Fu et al. (2021). In contrast, solar cycles 24 and 25 (up to 1 May 2024) demonstrate relatively weaker solar and geomagnetic activity.

Figure 2 provides a statistical overview of the SML data set, emphasizing its distribution and the inherent data imbalance. Figure 2a shows the histogram of |SML|, revealing a long-tailed distribution with a large concentration of data below 50 nT and a comparatively small number of samples (0.33%) above 1,000 nT. Figure 2b further illustrates this imbalance, with the number of samples decreasing exponentially as |SML| increases, emphasizing the rarity of high-intensity geomagnetic activity. Quiet-time observations dominate the data set: 80% correspond to values below 200 nT, with only 0.33% of |SML| values exceeding 1,000 nT (representing strong events), 0.01% of observations exceeding 2,000 nT (representing extreme events), and an even smaller fraction (0.0006%) surpasses 3,000 nT. This extreme skewness suggests that quiet-time samples would overwhelm any MSE-based models, which distribute weight evenly across all samples in the loss function, making it difficult for such models to predict strong geomagnetic events accurately.

Figure 2c presents the histogram of $\log_{10}$(|SML|), which approximates a more balanced Gaussian-like distribution. The logarithmic transformation reduces skewness, resulting in a distribution where extreme events (i.e., high |SML| values) are less rare relative to the bulk of the data. Consequently, in this study, we develop a model to predict $\log_{10}$(|SML|) rather than |SML|, as this transformation better addresses the imbalance and allows for improved prediction of strong geomagnetic events.

3. Methodology

3.1. Model Description

In this study, we develop an imbalanced regression model for the SML index using a fully connected neural network, following the workflow outlined in Chu, Ma et al. (2021), Chu et al. (2023, 2024). The architecture of the neural network is similar to that employed in these past studies, which successfully modeled global dynamic distributions of plasma density, electron and ion fluxes, and whistler-mode chorus and hiss waves (Chu, Bortnik, Li, Ma, Angelopoulos, Thorne, 2017, Chu, Bortnik, Li, Ma, Denton, et al., 2017; Chu, Ma et al. (2021), Chu et al. (2023, 2024). The model consists of a linear input layer, three hidden layers with sigmoid activation functions, each followed by batch normalization (Ioffe & Szegedy, 2015) and dropout (Srivastava et al., 2014) to avoid overfitting, and a linear output layer. Although the SML index represents time-series data, we chose a fully connected neural network architecture for simplicity and computational efficiency, following similar successful applications in previous studies (Chu et al., 2024). However, since this study focuses on the importance of imbalanced regression to predict strong-to-extreme geomagnetic events, incorporating temporal models such as RNNs and LSTM could be explored in future work to capture sequential dependencies in the solar wind parameters. The input parameters include time series of solar wind parameters and coupling functions from the

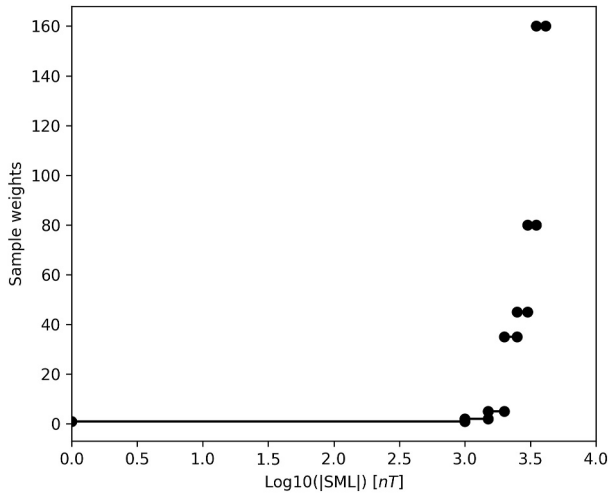

Figure 3. Sample weight distribution for the Imbalanced Regression Artificial Neural Network Model for Auroral Electrojet indices model. The weights are plotted as a function of $\log_{10}$(|SML|), with higher weights for larger |SML| values to account for the data imbalance.

OMNI data set, which are detailed below. The target parameter, or model output, is the base-10 logarithm of the absolute value of the SML index, $\log_{10}$(|SML|). Before training, the model inputs and outputs are normalized using their means and standard deviations, and then rescaled when making predictions.

Imbalanced regression is crucial due to the highly imbalanced nature of the data set, even after transforming to $\log_{10}$(|SML|), with far more quiet-time samples than active-time samples. To address this, we adopt a weighted approach similar to that of Chu et al. (2023, 2024), assigning different weights to data samples based on their amplitudes, as shown in Figure 3. For |SML| < 1,000, samples are assigned a weight of 1.0, while stronger |SML| values are assigned progressively larger weights, reflecting increasing geomagnetic activity. These weights are empirically determined through extensive experiments, ensuring that the model's predictions closely align with the data, as illustrated by the diagonal line in Figure 4.

The model's loss function is the weighted mean squared error (WMSE) (wmse) of $\log_{10}$(|SML|), defined as

$$\mathrm{WMSE} = \frac{\sum_1^n w_i\left(\log_{10}|\mathrm{SML}_{\mathrm{obs}}| - \log_{10}|\mathrm{SML}_{\mathrm{model}}|\right)^2}{n\sum_1^n w_i}$$

where w_i represents the weight of each data sample. To minimize this loss function, the neural network is trained using the Nesterov-accelerated Adaptive Moment Estimation (Nadam) optimizer (Dozat, 2016). To prevent data leakage, the data set is segmented into daily blocks, ensuring that each 1-day segment is longer than the typical substorm duration of 1–2 hr (Chu et al., 2015). The data segments are then randomly split into training (60%), validation (20%), and test (20%) sets. In addition, the geomagnetic events analyzed in Section 5 are specifically held out as an out-of-sample test data set. To avoid overfitting, early stopping is applied after 15 epochs, with dropout layers added after each hidden layer (Srivastava et al., 2014).

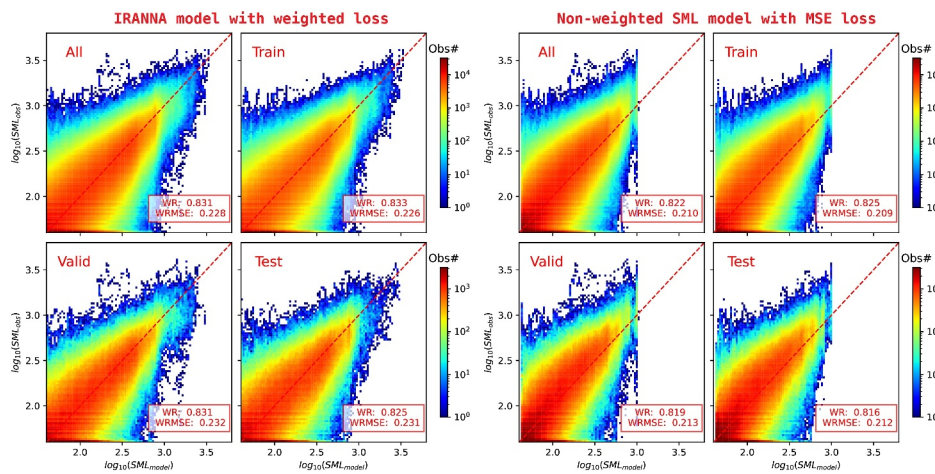

Figure 4. Comparison between observed and model-predicted $\log_{10}$(|SML|). The 2D histograms show observed versus predicted $\log_{10}$(|SML|) values for the entire data set ("All") as well as the training, validation, and test sets for the Imbalanced Regression Artificial Neural Network Model for Auroral Electrojet indices model (left) and the traditional MSE-based model (right). The model demonstrates strong performance, with WR and weighted root mean squared error (WRMSE) values shown in red in the legend for each data set. For the non-weighted traditional model, WR and WRMSE are functionally equivalent to the standard R and RMSE metrics, respectively.

3.2. Feature Selection and Hyperparameter Optimization

We adopted a similar approach for feature selection and hyperparameter optimization (HPO) as described in Chu et al. (2023, 2024). Our method progressively integrates the most informative input parameters into the Imbalanced Regression Artificial Neural Network model for Auroral electrojet indices (IRANNA) model and evaluates their impact on performance (Kuhn and Johnson, 2013). We chose a forward feature selection approach, allowing us to evaluate the contribution of each solar wind parameter independently while minimizing the risk of feature self-correlation.

First, we compiled a list of solar wind drivers from the OMNI data set at 1-min resolution, along with several coupling functions proposed by Newell et al. (2007) and McPherron et al. (2015). Next, we iteratively selected one parameter from this list to use as input and trained the neural network using a 5-min averaged time series over a 6-hr lookback window. After testing all drivers, the coupling function $v^2\, B_{tot}{}^{1/2} B_T \sin^6(\theta_c/2)$ yielded the best model performance among all solar wind parameters. Another series of tests was conducted to decide which parameter would yield the best model performance when paired with the coupling function. This second parameter—solar wind speed—emerged as the next most informative feature and was combined with the coupling function to serve as our model's input. This finding is consistent with previous studies indicating that strong substorms triggered by interplanetary shocks typically occur under preconditions of a southward IMF Bz and fast solar wind (Zong et al., 2021). Although we explored adding additional parameters, incorporating a third feature did not further improve model performance. As a result, we selected the coupling function and solar wind speed as the primary input features for the model.

The optimal length of the lookback time series was determined based on model performance. While increasing the time history generally improved performance, the gains became marginal beyond 6 hr. To balance performance improvements with computational efficiency, we selected a 6-hr time series for model input.

For HPO, the model's architecture—including the number of neurons in each hidden layer and the dropout rates—was fine-tuned using the Tree-structured Parzen Estimator algorithm (Bergstra et al., 2011, 2013) implemented through Optuna (Akiba et al., 2019). The final model architecture consists of three hidden layers with 121, 256, and 39 neurons, respectively, and dropout rates of 0.44, 0.19, and 0.39.

4. Statistical Analysis of Model Performance

We evaluate the performance of the IRANNA model using a series of statistical analyses to demonstrate its ability to effectively handle the challenges of imbalanced data sets, establishing its robustness and reliability in prediction. For comparison, a traditional MSE-based SML model was trained without the weighting scheme, which failed to predict strong events.

Figure 4 shows the correlation between the observed and modeled SML amplitudes, $log_{10}(|SML|)$, for values exceeding quiet background levels, using the entire data set as well as the training, validation, and test data sets for the IRANNA model (left) and the traditional MSE-based model (right). The number of samples in each bin is color-coded. The red dashed diagonal line ($y = x$) represents perfect amplitude prediction by the IRANNA model. Most observed and modeled data pairs roughly cluster near this diagonal line, indicating that the IRANNA model accurately replicates observations most of the time with minimal over- or underestimation of $\log_{10}(|SML|)$ amplitudes. The IRANNA model can predict |SML| with reasonable uncertainty. The weighted root mean squared error (WRMSE), shown in the lower-right corner of each panel, is 0.23 for the test data set $\log_{10}(|SML|)$. This corresponds to 1.70 ($10^{0.23}$) for linear |SML|, representing the uncertainty in the forms of the ratio between the observed and predicted |SML|. Note that the model uncertainty is a ratio instead of an absolute value. Thus, it is different when converted to a linear scale. For |SML| = 1,000 nT, the uncertainty ranges between $1{,}000 \times [1/1.70, 1.70] = [588, 1{,}700]$ nT. While for |SML| = 3,000 nT, the uncertainty expands to between 1764 and 5,100 nT. The current IRANNA model was trained to predict $\log_{10}(|SML|)$ because its distribution is closer to a Gaussian form, making it less skewed and easier to train. An alternative approach would be to develop an IRANNA model that directly predicts |SML| on a linear scale, which is currently in preparation. However, due to the highly imbalanced nature of |SML| in its linear form, such a model struggles to achieve comparably accurate predictions over the bulk of the distribution.

The weighted correlation coefficients are 0.833, 0.831, and 0.825 for the training, validation, and test data sets, respectively, at a 1-min resolution. Notably the weighted correlation coefficient is sensitive to the data time

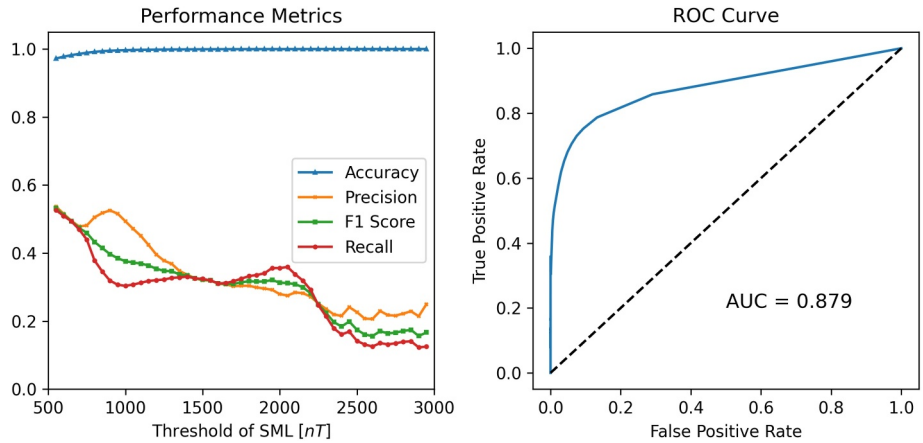

Figure 5. Performance metrics (left) and receiver operating characteristic (ROC) curve (right) of the Imbalanced Regression Artificial Neural Network Model for Auroral Electrojet indices model as a function of the SML index threshold. The left panel shows accuracy, precision, recall, and F1-score for different SML thresholds, illustrating the trade-offs between these metrics. The right panel presents the ROC curve, with an area under the curve of 0.879, indicating the model's overall classification performance.

resolution, partly due to large differences in data variability between data sets of different time resolutions, while the WRMSE remains more consistent across experiments. Therefore, special attention should be paid to data resolution when comparing model performance, as discussed in Section 6.2.

It is important to note that the IRANNA model effectively predicts strong SML values above 1,000 nT, a challenge for traditional statistical models and many neural network approaches (Camporeale, 2019; Chu et al., 2023). Due to the inherent data imbalance, conventional models that minimize MSE are often dominated by quiet-time values, leading to the underestimation of strong |SML| values. In particular, the traditional MSE-based SML model without a weighting scheme exhibits saturation effects, indicating a systematic inability to predict |SML| values beyond the 1,000 nT threshold in both statistical and event-based analyses. This study underscores the necessity of imbalance-aware regression techniques in handling imbalanced data sets and establishes the IRANNA model as the first model capable of accurately predicting high-intensity geomagnetic events.

Figure 5 presents the performance metrics and receiver operating characteristic (ROC) curves of the IRANNA model at different SML thresholds. The IRANNA model is evaluated as a binary classifier, where a prediction is considered positive if the predicted |SML| exceeds a given threshold. For each observation-prediction pair, a confusion matrix is computed at different thresholds to determine the number of true positives, true negatives, false positives, and false negatives. The performance metrics are defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F1 = 2.0 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The left panel of Figure 5 illustrates the variation of these performance metrics with different SML thresholds. Accuracy remains consistently high and close to 100%, primarily due to the class imbalance in the SML data set, where quiet-time values dominate and are easier to predict. Precision measures the proportion of predicted |SML| values that exceed the threshold and are correctly identified, while Recall represents the percentage of observed |SML| values above the threshold that are correctly identified. The F1 score, which balances precision and recall, is approximately 0.4 at |SML| = 1,000 nT and 0.3 at |SML| = 2,000 nT. This suggests that about 40% of the large

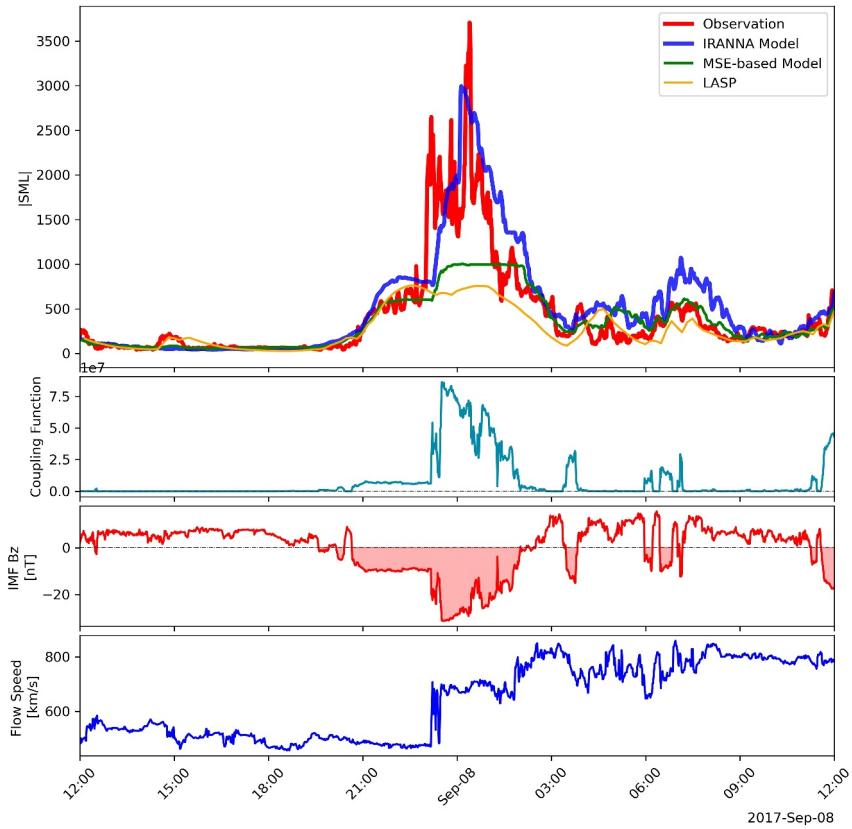

Figure 6. Event analysis of a super substorm on 8 September 2017. The first panel shows the comparison between the observed (red), Imbalanced Regression Artificial Neural Network Model for Auroral Electrojet indices model predicted (blue) SML index, traditional MSE-based model (green), and predictions at LASP (yellow). The second to fourth panels show the coupling function $v^2 B_{tot}^{1/2} B_T \sin^6(\theta_c/2)$, interplanetary magnetic field Bz, and solar wind speed, respectively.

and quiet-time |SML| values are correctly predicted. Therefore, the IRANNA model is more effective at predicting strong and long-duration geomagnetic activity, particularly when multiple |SML| data points exceed the threshold. This is because the likelihood of at least one correct prediction increases when geomagnetic activity (e.g., a substorm event) spans multiple time steps with consistently elevated |SML| values.

The right panel of Figure 5 presents the ROC curve with an area under the curve (AUC) of 0.879. The ROC curve illustrates the model's performance across all thresholds, with the AUC serving as a measure of overall effectiveness. A higher AUC indicates better performance across varying thresholds. A perfect classifier, which completely distinguishes between positive and negative samples, has an AUC of 1.0. In contrast, an AUC of 0.5 (represented by the dashed line) indicates a model that performs no better than random guessing. Generally, an AUC between 0.7 and 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and values above 0.9 are regarded as outstanding (Hosmer and Lemeshow, 2000; Mandrekar, 2010). With an AUC of 0.879, the IRANNA model demonstrates excellent classification performance, effectively distinguishing between substorm and non-substorm conditions.

5. Event Analysis

In this section, the IRANNA model is evaluated on a selected set of strong geomagnetic events, particularly super substorms. Examples demonstrate both strong and weak performance and potential contributing factors are discussed. Notably, these events were held out as an out-of-sample test data set.

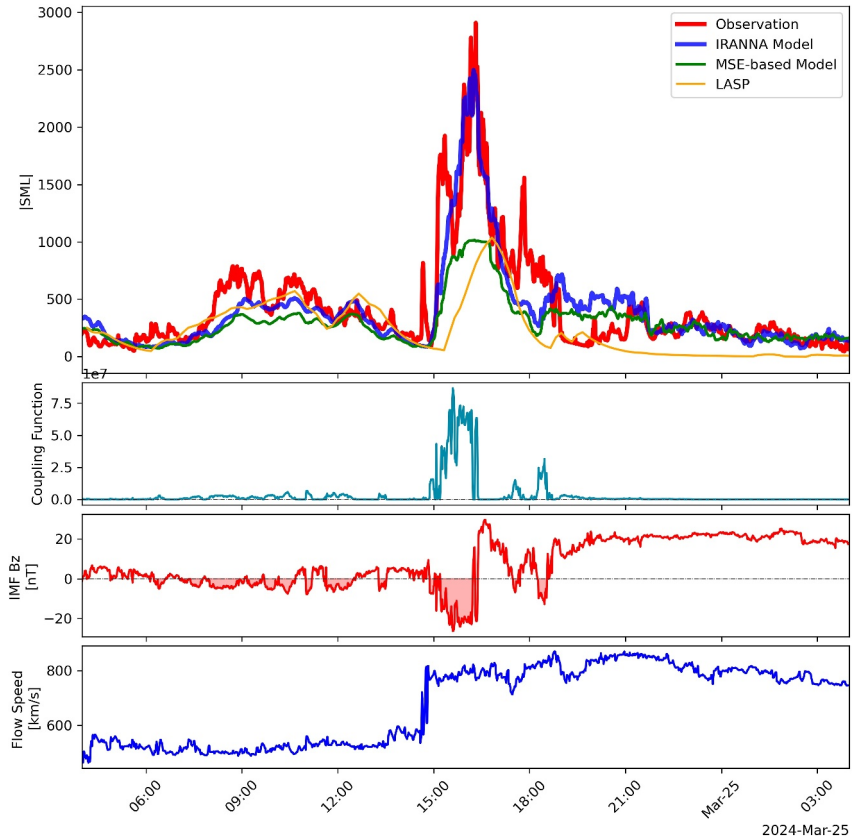

Figure 7. Event analysis of a super substorm on 24 March 2024, in the same format as Figure 6.

5.1. Super Substorm on 7 September 2017

Figure 6 presents the super substorm observed on 7 September 2017, which occurred in isolation, with quiet conditions before and after the event. Multiple coronal mass ejections (CMEs) erupted from the same active region that drove this event between September 4 to 6, 2017 (Werner et al., 2019). The magnetic clouds from these CMEs reached Earth, causing a geomagnetic storm and triggering a super substorm. On September 7 around 2000 UT, the IMF Bz turned southward to −10 nT when the magnetic cloud arrived, remaining negative until 0200 UT on September 8. At 2,300 UT, the solar wind speed sharply increased from 460 km/s to 700 km/s with the arrival of an interplanetary shock, accompanied by a strongly negative Bz of −30 nT. The coupling function rose to 0.75×10^7 at the arrival of the magnetic cloud and increased significantly to 8.5×10^7 after the shock.

The CME induced a super substorm with |SML| > 3,500 nT. The SML index first began to intensify at around 2,000 UT, reaching 1,000 nT at the arrival of the magnetic cloud. A second, faster increase in the |SML| index was observed around 2,300 UT, coinciding with the shock arrival, during which the strongly negative Bz resulted in enhanced solar wind driving, as indicated by the coupling function. The IRANNA model effectively predicted the quiet-time values preceding the super substorm onset. Additionally, the model accurately captured the timing of the first substorm onset and the second enhancement. Most notably, the IRANNA model reproduced the peak values of the super substorm (3,000 nT vs. the observed 3,500 nT) with impressive accuracy. In contrast, the traditional MSE-based models only predict peak values around 1,000 nT in such cases (see discussion in Section 4).

5.2. Super Substorm on 24 March 2024

Figure 7 shows the super substorm observed on 24 March 2024, which occurred in isolation, with quiet conditions before and after the event. A magnetic cloud from a CME was observed around 1,430 UT on March 24. The solar wind speed increased sharply from 580 to over 800 km/s. The IMF Bz turned southward around 1,450 UT,

10.1029/2024SW004236

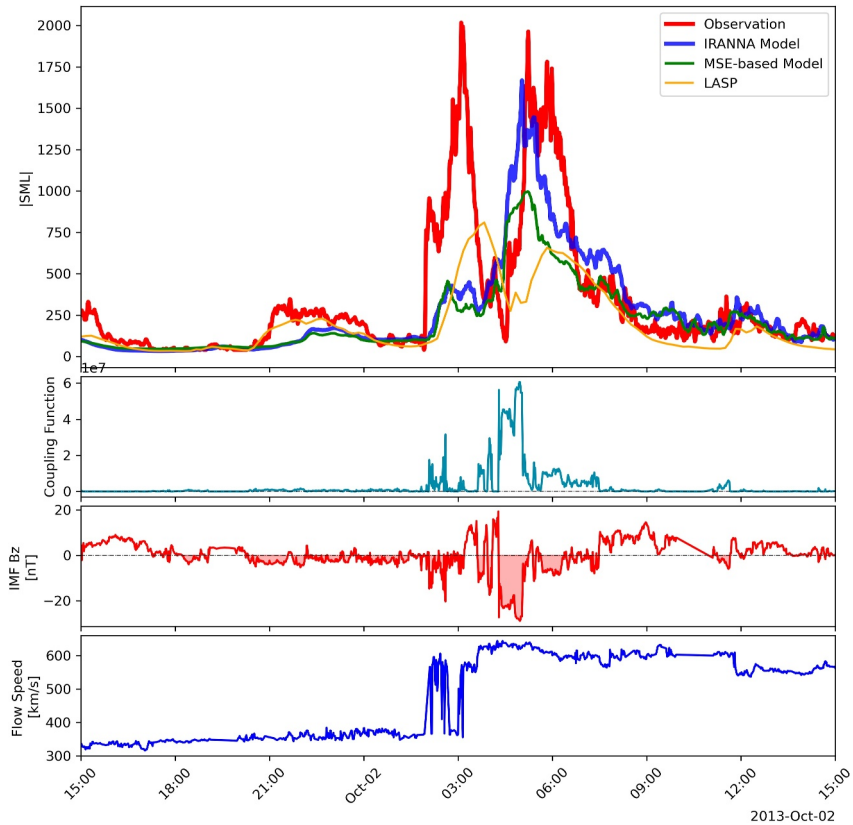

Figure 8. Event analysis of a super substorm on 2 October 2013, in the same format as Figure 6, based on solar wind parameters from OMNI.

reaching a minimum of −25 nT, before turning northward again at 1,625 UT, where it remained thereafter. The coupling function increased with the southward turning of the IMF Bz to a peak of 8.5×10^7 and dropped to near zero after the northward turning.

The magnetic cloud triggered a super substorm with an |SML| of 2,900 nT. The IRANNA model successfully predicted the relatively weak substorm activity and its subsequent decay leading up to the super substorm. Additionally, the model accurately captured the timing of the super substorm onset within 5 min. Most importantly, the IRANNA model again very accurately predicted the peak amplitude of the super substorm. In comparison, the traditional MSE-based model only predicts peak values up to 1,000 nT in such cases, demonstrating the advantage of using a weighted model for strong events.

5.3. Super Substorm on 2 October 2013: The Importance of Solar Wind Propagation

Figure 8 illustrates the super substorm recorded on 2 October 2013, which was induced by the magnetic cloud of a CME arriving around 0200 UT. The solar wind speed increased from 350 to 640 km/s, while the IMF Bz was slightly negative (−10 nT) upon the magnetic cloud's arrival between 0200 UT and 0315 UT. The IMF briefly turned northward before turning strongly southward to −30 nT at 0418 UT, where it remained for 60 min. The coupling function was slightly elevated during the first southward Bz period and increased significantly during the second period of southward IMF Bz.

This super substorm exhibited two peaks, each reaching ∼2,000 nT, with quiet conditions before and after. Both peaks corresponded to periods of strong solar wind driving. While the IRANNA model accurately captured the timing (±10 min) and amplitude of the second super substorm, it only predicted the onset timing of the first substorm, with an underestimated peak of ∼500 nT, deviating significantly from the observed 2,000 nT. This underestimation is consistent with the behavior of the coupling function, which was much weaker during the first

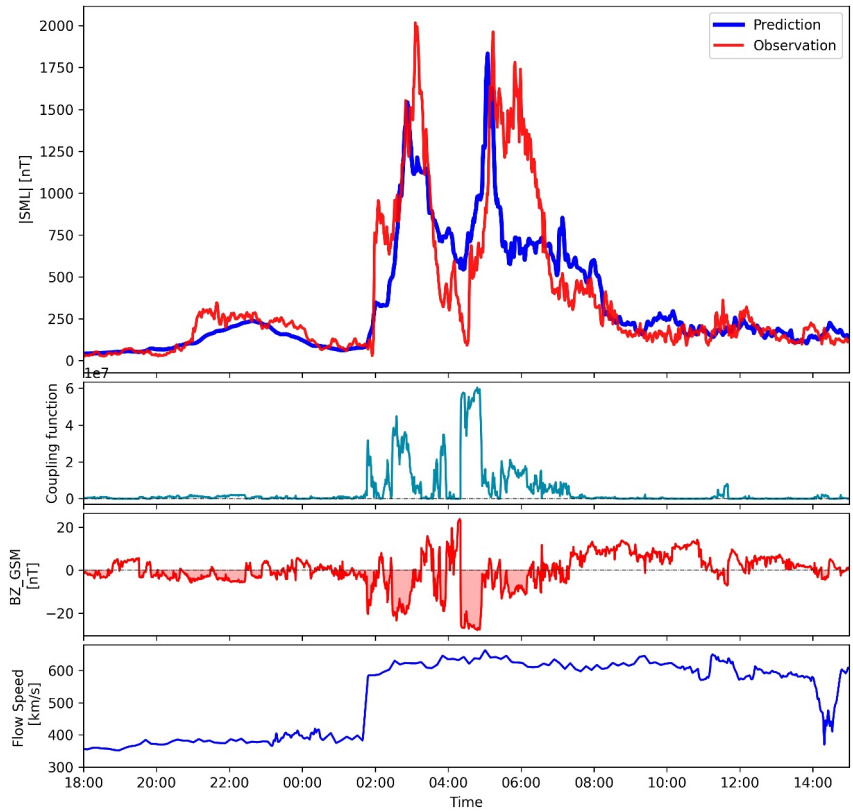

Figure 9. Event analysis of a super substorm on 2 October 2013, in the same format as Figure 6, based on solar wind parameters from THEMIS-C located upstream inside the solar wind.

southward Bz, indicating that the model's prediction of a weaker substorm was reasonable given the weaker driver.

Figure 9 shows the same event, but now using solar wind observations from the THEMIS/ARTEMIS mission. THEMIS-C was orbiting the Moon around $(x, y, z) = (50, -34, -4)$ Re in GSE coordinates and was located upstream of the Earth inside the solar wind. At solar wind speeds of 350 and 600 km/s, it takes approximately 12 and 17 min for the solar wind to propagate to Earth's bow shock. Note that solar wind propagation time was not applied in Figure 9. The solar wind speeds and the IMF Bz from THEMIS-C differed from OMNI values, which resulted in different coupling functions. Consequently, the IRANNA model was able to capture both the timing and peak amplitude of the first super substorm. However, during the second period, the southward Bz recorded by THEMIS-C was shorter in duration and slightly weaker. As a result, although the model captured the timing of the second substorm, the substorm duration was underestimated.

Upon further investigation, the discrepancy in the model's prediction arises from a combination of (a) uncertainties in solar wind propagation, which affects OMNI-based predictions during the first super substorm, and (b) differences between solar wind conditions observed upstream (e.g., L1 or at THEMIS-C) and those reaching Earth (e.g., Walsh et al., 2019), affecting predictions during the second super substorm using THEMIS-C, which was upstream near Earth. For example, ACE and Wind satellites can be positioned ∼100 Re away from the Sun-Earth line, while THEMIS-C, even though it was upstream near Earth, was ∼33 Re azimuthally displaced from the Sun-Earth line. This spatial offset means that the measured solar wind structure may not be identical to that encountered directly upstream of Earth. (c) Furthermore, the solar wind is filled with a network of entangled magnetic flux tubes. Their texture impacts the flow properties and turbulence in the solar wind, leading to further difficulties in solar wind propagation (Borovsky, 2008).

Figure 10 compares solar wind measurements from multiple satellite monitors and OMNI during the October 2013 event. The panels illustrate the solar wind flow speed and IMF Bz from OMNI, THEMIS-C, ACE, and

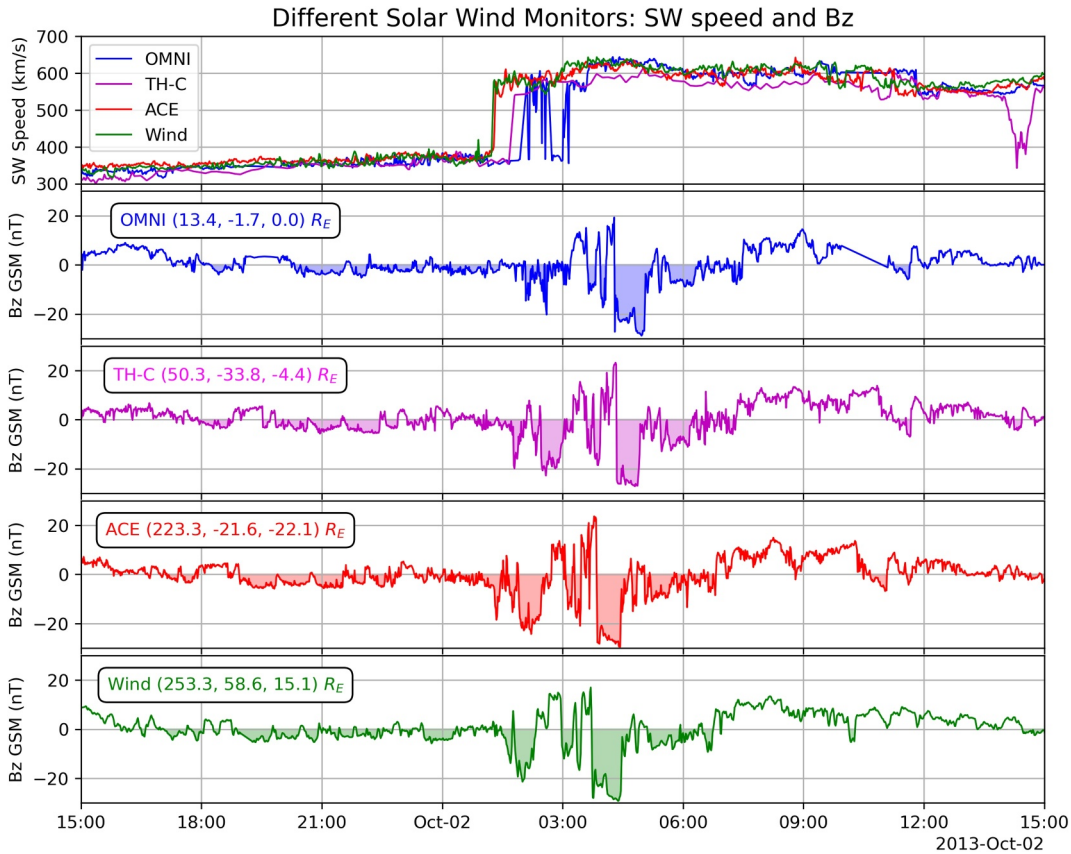

Figure 10. Comparison of solar wind speed and interplanetary magnetic field (IMF) Bz (in GSM coordinates) from OMNI, THEMIS-C, ACE, and Wind during the 2 October 2013 event. The top panel shows the solar wind speed measurements, while the lower panels display the IMF Bz variations observed at their respective locations in GSE coordinates.

Wind, along with their respective satellite positions in GSE coordinates. A key observation is that while Wind, ACE, and THEMIS-C all recorded a consistent increase in solar wind speed from 400 to 600 km/s, the OMNI data set showed an anomalous pattern—an increase, followed by a decrease, and then another increase. This discrepancy suggests that uncertainties in solar wind propagation could be responsible for deviations in OMNI data compared to direct observations. Additionally, the IMF Bz data from THEMIS-C, ACE, and Wind reveal two distinct southward Bz periods following the flow speed increase. However, in OMNI data, the first period of southward Bz appears significantly weaker, while the second period is more pronounced. Since the coupling function depends on both solar wind flow speed and IMF Bz, the underestimation of the first southward Bz period in OMNI data leads to a lower calculated coupling function, resulting in an underestimated SML index during the first super substorm. Conversely, during the second southward Bz period, the OMNI, ACE, and Wind data indicate a longer duration of southward Bz, whereas THEMIS-C records a shorter duration. This results in an underestimated coupling function when using THEMIS-C data, which subsequently leads to an underestimation of the SML index for the second super substorm.

This event analysis highlights the critical importance of accurately measuring solar wind parameters and properly accounting for solar wind propagation in space weather forecasting models, as differences in upstream solar wind conditions can significantly impact predictions.

5.4. Super Substorm on 14 October 2016: The Difference Between Global and Localized Substorms

Figure 11 presents the super substorm observed at 01:00 UT on 14 October 2016, which occurred during a geomagnetic storm with a Dst minimum of −100 nT. The SML index reached a maximum of 2,000 nT but lasted only about 10 min, which is unusually short for a super substorm. The IMF Bz was southward at −15 nT, and the coupling function increased to 1.0×10^7 during this event. However, this solar wind driving was too weak to

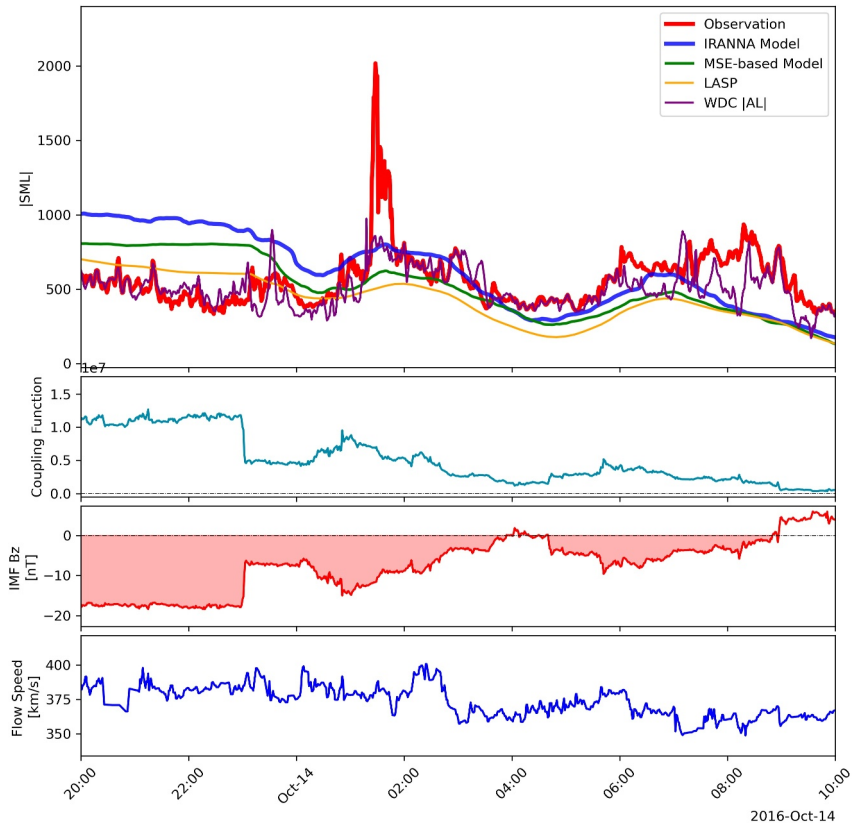

Figure 11. Event analysis of a super substorm on 14 October 2016, in the same format as Figure 6 above.

produce a super substorm, consistent with the IRANNA model prediction of an SML index peaking at only reached 800 nT. Similarly, the AL index from WDC (green) followed a comparable trend, peaking at 900 nT. Both the solar wind driving and the WDC AL index indicate that a super substorm was unlikely to have occurred during this period.

Upon closer examination, one magnetometer station, RAN (Rankin Inlet), at magnetic coordinates (−23.12°, 72.44°) near Hudson Bay, Canada, recorded a magnetic impulse of up to 2,000 nT, while nearby stations observed magnetic perturbations of ∼1,000 nT or lower. The observation indicates that a highly localized electrojet was located overhead at RAN, with an amplitude of 2,000 nT, while the broader substorm-related westward electrojet reached only about 1,000 nT. Since the SML index reflects the envelope of the most extreme northward perturbation, the SML index was dominated by the strong localized electrojet. The IRANNA model, which is designed to capture the global response of the magnetosphere-ionosphere system to external solar wind driving, did not capture this localized electrojet event. This event highlights the importance of station coverage for global geomagnetic indices like the auroral electrojet indices, which can be impacted by both a sparse station distribution that misses regional electrojets or a dense distribution that is skewed by localized currents.

5.5. Virtual Substorm Events

Figure 12 demonstrates the IRANNA model's ability to study the magnetospheric response to solar wind driving using controlled virtual solar wind events. This study focuses on varying the IMF Bz while keeping other parameters constant. The solar wind speed was set to 500 km/s, and the Bx and By components were set to zero.

In event A, the IMF Bz was set to −10 nT (southward) for 2 hr, resulting in a coupling function of approximately 8.0×10^6. A substorm was triggered with a peak amplitude of ∼1,000 nT, lasting about 3 hr. In event B, the IMF Bz was increased to −20 nT, producing a stronger substorm with a peak amplitude of 2,200 nT, and the event duration remained the same at 3 hr. In event C, the IMF Bz was kept at −20 nT but persisted for 3 hr, leading to a substorm with the same peak amplitude (1,400 nT) as event B but with a longer duration of 4 hr. In event D, the

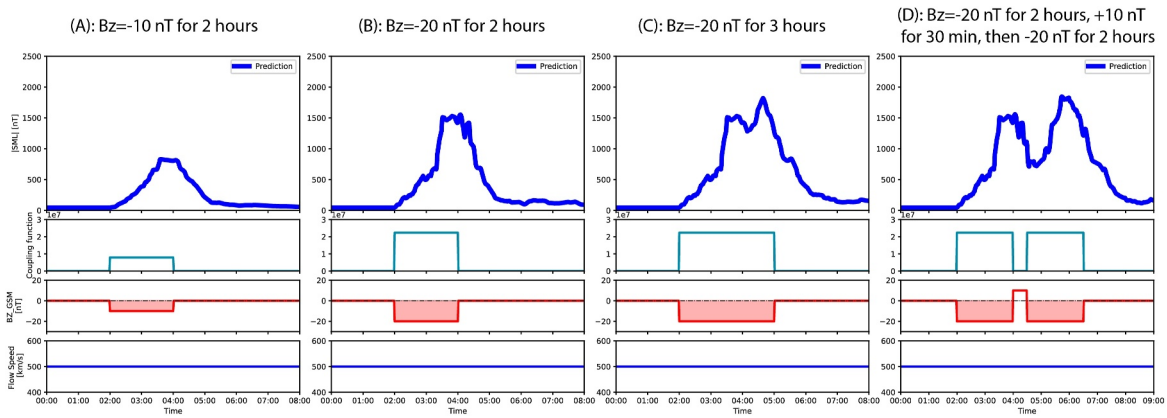

Figure 12. Virtual substorm events: (a) Southward Bz = −10 nT for 2 hr. (b) Southward Bz = −20 nT for 2 hr. (c) Southward Bz = −20 nT for 3 hr. (d) Southward Bz = −20 nT for 2 hr, followed by northward turning with Bz = 10 nT for half an hour, and another period of Bz = −20 nT for 2 hr.

IMF Bz was initially set to −20 nT for 2 hr, followed by a northward turning to +10 nT for half an hour, before turning southward again to −20 nT for another 2 hr. This produced a multiple-onset substorm, as expected. The first onset was identical to event B, with a peak of 2,200 nT before the northward turning. The second onset had a lower peak of 1,500 nT and lasted about 3 hr.

These virtual substorm experiments offer valuable physical insights into the magnetospheric response to variations in solar wind driving. Specifically, the strength and duration of the coupling function directly influenced the onset timing, duration, and intensity of substorm activity, including events with multiple onsets. While a detailed analysis of these results is beyond the scope of this study, these events highlight the potential of using virtual experiments to explore substorm dynamics in greater detail, which will be investigated in future work.

6. Discussion

6.1. The Importance of Imbalanced Regression

We developed an Imbalanced Regression Artificial Neural Network model for the Auroral electrojet index (IRANNA) from SuperMAG |SML|, with a focus on addressing the pronounced imbalance in the SML data set. About 99.7% of |SML| values are below 1,000 nT, with only 0.33% of |SML| values exceeding 1,000 nT (representing strong events), and 0.01% of observations exceeding 2,000 nT (representing extreme events). Traditional models based on minimizing the MSE between model predictions and observations tend to be biased by the dominant quiet-time values, effectively "regressing to the mean" and underestimating strong |SML| values, which are critical for space physics studies and for producing meaningful space weather forecasts. To overcome this, we employed a WMSE loss function, assigning higher weights to larger |SML| values, thus allowing the model to focus more on the underrepresented strong geomagnetic activity.

For the first time, we can accurately predict strong-to-extreme geomagnetic events. The IRANNA model successfully predicts strong |SML| values across various statistical metrics, event studies, and virtual experiments. In contrast, the predictions of a traditional MSE-based model without weighting are capped at a certain value, preventing it from accurately forecasting strong-to-extreme geomagnetic activity. The IRANNA model uses a 6-hr long lookback window with 5-min averages of time series of the coupling function and solar wind speed to predict $\log_{10}$(|SML|) at a 1-min resolution. Importantly, |SML| history is not used as an input, making the model more robust for real-time applications. This high temporal resolution is crucial for accurately capturing substorm onsets. The model's uncertainty for $\log_{10}$(|SML|) is ∼0.23, translating to an uncertainty factor of approximately 1.70 ($10^{0.23}$) for |SML|. Given the prevalence of imbalanced data in space physics and space weather studies in general (i.e., extreme events are inherently infrequent), the emphasis on imbalanced regression techniques is increasingly critical for advancing research in these fields.

The IRANNA model predicts strong geomagnetic activity within the above uncertainty range. However, prediction errors increase under certain conditions, primarily due to uncertainties in the input solar wind parameters and the quality of the target index itself. First, the uncertainty in the coupling function is sensitive to the accuracy

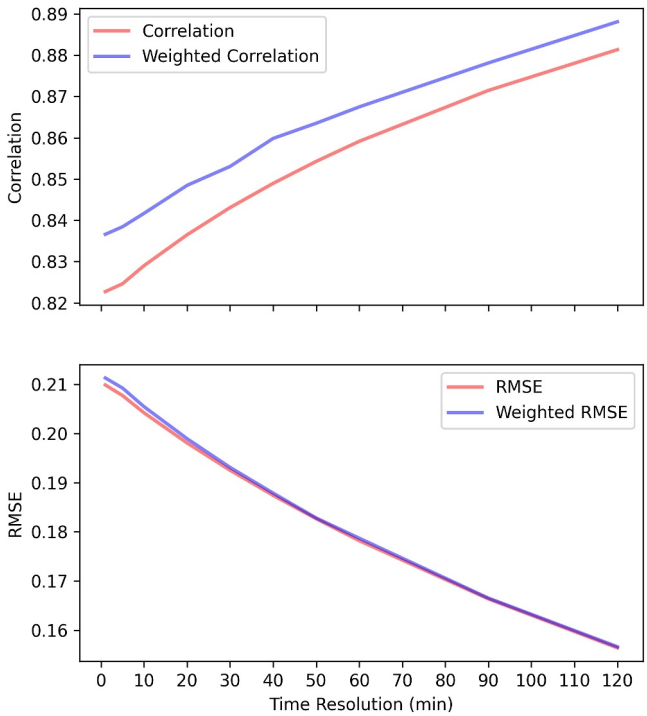

Figure 13. Model performance as a function of the time resolution of |SML| in minutes shows that model performance improves when time resolution decreases.

of solar wind propagation from the L1 point to Earth's bow shock, particularly the IMF Bz, which largely controls the coupling function. The challenge of accurately propagating the solar wind, especially during periods of unreliable solar wind data, contributes to larger prediction uncertainties. Additionally, the L1 observatory may not be strictly at L1 but could have very large (∼100 Re) deviations from the Sun-Earth line. Furthermore, the texture of the solar wind, with flux tubes, may lead to additional difficulties in the propagation of solar wind. Second, prediction errors are significant for localized electrojet events, where the SML index is dominated by large, impulsive magnetic perturbations that do not reflect the strength of global westward electrojets. These localized impulsive magnetic perturbations dominate the SML index, leading to larger discrepancies in the model's predictions.

6.2. The Importance of Time Resolution

The performance of the IRANNA model depends on several factors, with time resolution being a critical aspect to consider when comparing it to other models. Since these models are driven solely by external solar wind parameters and do not incorporate past SML values, their performance is influenced by the time resolution of the SML index. This dependence is expected, as variance and bias change with different time resolutions, thereby affecting model performance. The trade-off between the variance and biases implies that the lower the variance, the lower the biases, and thus the better model performance. For example, as time resolution increases to a larger cadence (e.g., from 1 to 60 min), the data becomes smoother and less noisy, making it easier for the model to predict and reducing the overall error. In Figure 13, the correlation coefficient (Pearson R) and the weighted correlation coefficient (WR) were calculated between the observed and IRANNA predicted SML at a 1-min resolution. Subsequently, the observed and predicted SML values were smoothed and interpolated at various time resolutions, up to 2 hr, and their correlation coefficients were recalculated at each time resolution.

Notably, Pearson R increases from 0.82 to 0.88 as the time resolution increases. The weighted correlation coefficient follows a similar trend, consistently exceeding Pearson R across all resolutions. On the other hand, both the RMSE and weighted RMSE decrease from 0.21 to 0.16 as time resolution increases since spurious impulses tend to get smoothed out. This indicates that model performance improves with larger time resolutions, which makes this analysis useful for comparing models operating at different time scales. For example, McPherron et al. (2015) used the optimal function (opn) to predict the WDC AL index at a 60-min resolution, achieving a prediction efficiency (R^2) of 68%, corresponding to an R of 0.82. On the other hand, the IRANNA model at 60-min resolution has an R of 0.859 and a WR of 0.867, which indicates a higher performance for the IRANNA model.

6.3. The Difference Between Driving Models and Persistence Models

The IRANNA model is a solar wind-driven model, meaning it does not use past |SML| values as input. We develop a solar wind-driven model instead of a history-dependent model because adding SML (t−1 min) into the input can lead to a model dominated by historical values rather than external solar wind driving. The auto-regressive approach tends to predict changes between two time steps in ΔSML rather than the actual values of SML, reducing the model's interpretability, particularly regarding the magnetospheric response to external solar wind driving.

In addition, special consideration is needed when comparing IRANNA to auto-regressive models that incorporate past |SML| values. By using past |SML| as input, these auto-regressive models tend to resemble persistence models, which rely solely on past values, without any external parameters, to make predictions. Figure 14 shows the model performance (correlation coefficient and RMSE) for $log_{10}(|SML|)$ and |SML| in persistence models over different time delays.

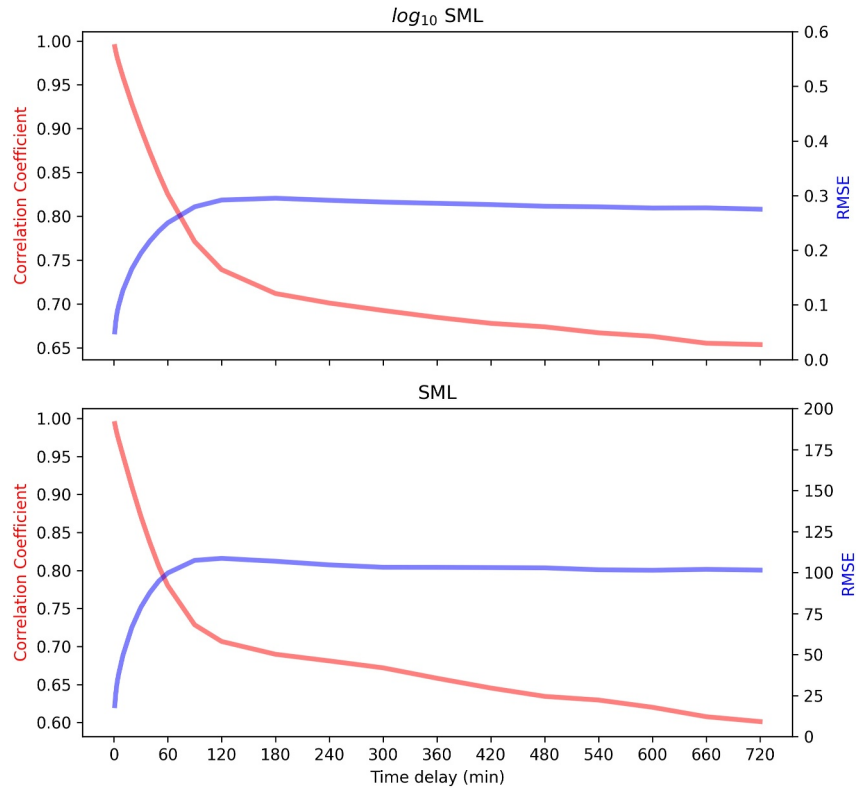

Figure 14. The model performance of persistence models at different time delays.

For instance, a persistence model that predicts $SML_{t+1\ min} = SML_t$ achieves an extremely high correlation coefficient of 0.994. However, as the time delay increases, the correlation drops significantly. For a persistence model predicting $SML_{t+60\ min} = SML_t$, the correlation coefficient falls to 0.825. Models that use past |SML| values in combination with solar wind parameters should outperform purely persistence models. Therefore, it is essential to note that the performance of such models should not be directly compared to that of solar wind-driven models like the IRANNA, as the latter does not rely on historical |SML| data.

For example, an LSTM model was developed to predict auroral electrojet indices, using historical values of geomagnetic indices and solar wind parameters at a 1-hr resolution as inputs to forecast the index 1 hr ahead (Zou et al., 2024). The LSTM model achieved a Pearson correlation coefficient (R) of 0.854 in predicting the AL index, outperforming the persistence model, which has a Pearson R of 0.825 at the same 1-hr resolution.

6.4. Implications for Space Physics and Weather Forecasting

The IRANNA model offers significant potential for improving space weather forecasting by accurately predicting auroral electrojet indices, which are key indicators of geomagnetic activity and a variety of space weather phenomena driven by these events. For instance, the dynamics of Earth's radiation belts, including the acceleration and loss of energetic electrons, are influenced by auroral electrojet indices more than by the ring current index (Hua and Bortnik, 2024; Ma et al., 2023, 2024; Reeves et al., 2003). Similarly, substorm activity impacts and directly relates to phenomena such as magnetic reconnection, auroral activations, ionospheric and field-aligned currents, ionospheric outflow, plasmaspheric dynamics, polar cap dynamics, and electric field patterns. Thus, the IRANNA model has the potential to improve predictions and provide valuable forecasts for space weather studies.

Furthermore, the IRANNA model offers a valuable tool for gaining physical insights into space weather phenomena. One area of interest is the onset timing of magnetic reconnection, which has long been a challenging issue to predict accurately. The IRANNA model has demonstrated the ability to capture substorm onsets reasonably well in these event analyses, suggesting that it could potentially be used to predict the timing of

magnetic reconnection with a reasonable degree of uncertainty. This result was unexpected as most previous models produce smooth, slowly varying onsets without the sharp changes in the slope of the SML index seen in observations. A possible explanation may be pulses in the solar wind coupling function that have recently been shown to be associated with substorm onset (McPherron, 2023). It was shown that a slow increase in coupling followed by either a rapid increase or decrease is frequently associated with a substorm onset 20 min later. Positive changes in the coupling function produce stronger substorms than negative changes. It is likely that our high-resolution model has captured this behavior, which needs further investigation to confirm and demonstrate.

This capability of the model to predict both the amplitude and timing of substorm onset could be leveraged in several ways. For example, numerical simulations of magnetospheric dynamics could be compared against the predictions of the IRANNA model to assess whether these simulations correctly capture magnetic reconnection, thereby improving their accuracy and providing validation.

6.5. Summary and Conclusions

Here, we describe the development of an Imbalanced Regression Artificial Neural Network model for the Auroral electrojet index (IRANNA) that predicts the magnitude of SuperMAG |SML| using a weighted loss function scheme and thus captures strong-to-extreme geomagnetic activity. We summarize the major conclusions as follows:

1. For the first time, the IRANNA model successfully predicted strong-to-extreme geomagnetic activity (e.g., super substorms), particularly the peak amplitudes of strong |SML| values, by implementing a customized weighting scheme for these large |SML| values in the loss function.
2. Model prediction errors can be significant under certain conditions, particularly when upstream solar wind parameters are inaccurately measured or improperly propagated, or when the SML index reflects spatially localized or temporally transient electrojets rather than reflecting global westward electrojets.
3. The prediction of virtual substorm events provides valuable physical insights into the magnetospheric response to solar wind driving and the underlying triggering mechanisms.
4. This study highlights the importance of developing and applying imbalanced regression techniques, given the prevalence of imbalanced data sets in space physics, space weather forecasting, and other real-world scenarios.

Data Availability Statement

Solar wind parameters and geomagnetic indices are sourced from the OMNI data set (Papitashvili and King, 2020), from SuperMAG (Gjerloev, 2012), and LASP (Luo et al., 2013; Temerin and Li, 2002, 2006). The neural network models were developed using the open-source TensorFlow package (TensorFlow Developers, 2023). The code and model are publicly available (Chu, 2025).

Acknowledgments
XC would like to thank grant NASA ECIP 80NSSC19K0911, 80NSSC20K0196, 80NSSC22K1023, 80NSSC20K1325, 80NSSC23K0096, 80NSSC18K1227, 80NSSC24K1112, NSF grant AGS-2247255, and AFOSR YIP FA9550-23-1-0359. In addition, JB acknowledges support from subgrants to the University of California, Los Angeles, from the University of Colorado Boulder under NASA Prime Grant agreement 80NSSC20K1580, and 80NSSC22K1023, and NSF award AGS-2025706. This work uses the Casper system provided by the NSF National Center for Atmospheric Research, sponsored by the National Science Foundation and Extreme Science and Engineering Discovery Environment (XSEDE) Bridges GPU at the PSC through allocation TG-PHY190033.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Paper presented in proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining (KDD' 19)* (pp. 2623–2631). Association for Computing Machinery. https://doi.org/10.1145/3292500.3330701

Amariutei, O. A., & Ganushkina, N. Y. (2012). On the prediction of the auroral westward electrojet index. *Annales Geophysicae*, *30*(5), 841–847. https://doi.org/10.5194/angeo-30-841-2012

Amata, E., Consolini, G., Pallocchia, G., & Marcucci, M. F. (2009). ANN forecast of hourly averaged AE index based on L1 IMF and plasma measurements. *Acta Geophysica*, *57*(1), 185–196. https://doi.org/10.2478/s11600-008-0083-1

Andriyas, T., & Andriyas, S. (2017). Use of multivariate relevance vector machines in forecasting multiple geomagnetic indices. *Journal of Atmospheric and Solar-Terrestrial Physics*, *154*, 21–32. https://doi.org/10.1016/j.jastp.2016.11.002

Angelopoulos, V., McFadden, J. P., Larson, D., Carlson, C. W., Mende, S. B., Frey, H., et al. (2008). Tail reconnection triggering substorm onset. *Science*, *321*(5891), 931–935. https://doi.org/10.1126/science.1160495

Baker, D. N., Pulkkinen, T. I., Angelopoulos, V., Baumjohann, W., & McPherron, R. L. (1996). Neutral line model of substorms: Past results and present view. *Journal of Geophysical Research*, *101*(A6), 12975–13010. https://doi.org/10.1029/95JA03753

Bala, R., & Reiff, P. (2012). Improvements in short-term forecasting of geomagnetic activity. *Space Weather*, *10*(6), S06001. https://doi.org/10.1029/2012SW000779

Bargatze, L. F., Baker, D. N., McPherron, R. L., & Hones, E. W. E. W., Jr. (1985). Magnetospheric impulse response for many levels of geomagnetic activity. *Journal of Geophysical Research*, *90*(A7), 6387–6394. https://doi.org/10.1029/JA090iA07p06387

Baumjohann, W., & Nakamura, R. (2001). Updating the near-earth neutral line model. *arXiv*. https://doi.org/10.48550/arXiv.physics/0111145

Bergstra, J., Yamins, D., & Cox, D. (2013). Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. *Proceedings of the 30th International Conference on Machine Learning*, In *Proceedings of Machine Learning Research*. 28(1):115–123. Retrieved from https://proceedings.mlr.press/v28/bergstra13.html

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems* (Vol. 24). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

Birn, J., & Hesse, M. (2014). The substorm current wedge: Further insights from MHD simulations. *Journal of Geophysical Research: Space Physics*, *119*(5), 3503–3513. https://doi.org/10.1002/2014JA019863

Blanchard, G. T., & McPherron, R. L. (1995). Analysis of the linear response function relating AL to VB for individual substorms. *Journal of Geophysical Research*, *100*(A10), 19155–19165. https://doi.org/10.1029/95JA01341

Borovsky, J. E. (2008). Flux tube texture of the solar wind: Strands of the magnetic carpet at 1 AU? *Journal of Geophysical Research*, *113*(A8), A08110. https://doi.org/10.1029/2007JA012684

Borovsky, J. E., Nemzek, R. J., & Belian, R. D. (1993). The occurrence rate of magnetospheric-substorm onsets: Random and periodic substorms. *Journal of Geophysical Research*, *98*(A3), 3807–3813. https://doi.org/10.1029/92JA02556

Cai, L., Ma, S. Y., & Zhou, Y. L. (2010). Prediction of SYM-H index during large storms by NARX neural network from IMF and solar wind data. *Annales Geophysicae*, *28*(2), 381–393. https://doi.org/10.5194/angeo-28-381-2010

Cai, X., Henderson, M. G., & Clauer, C. R. (2006). A statistical study of magnetic dipolarization for sawtooth events and isolated substorms at geosynchronous orbit with GOES data. *Annals of Geophysics*, *24*(12), 3481–3490. https://doi.org/10.5194/angeo-24-3481-2006

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, *17*(8), 1166–1207. https://doi.org/10.1029/2018SW002061

Chakraborty, S., & Morley, S. K. (2020). Probabilistic prediction of geomagnetic storms and the Kp index. *Journal of Space Weather and Space Climate*, *10*, 36. https://doi.org/10.1051/swsc/2020037

Chen, J., & Sharma, A. S. (2006). Modeling and prediction of the magnetospheric dynamics during intense geospace storms. *Journal of Geophysical Research*, *111*(A4), A04209. https://doi.org/10.1029/2005JA011359

Chu, X. (2015). *Configuration and generation of substorm current wedge (PhD Thesis)*. University of California. Retrieved from https://escholarship.org/uc/item/4sw949rf

Chu, X. (2025). Imbalanced regression model of auroral electrojet indices: Can we predict super substorms? *Zenodo*. https://doi.org/10.5281/zenodo.14871764

Chu, X., Bortnik, J., Li, W., Ma, Q., Denton, R., Yue, C., et al. (2017). A neural network model of three-dimensional dynamic electron density in the inner magnetosphere. *Journal of Geophysical Research: Space Physics*, *122*(9), 9183–9197. https://doi.org/10.1002/2017JA024464

Chu, X., Bortnik, J., Li, W., Shen, X.-C., Ma, Q., Ma, D., et al. (2023). Distribution and evolution of chorus waves modeled by a neural network: The importance of imbalanced regression. *Space Weather*, *21*(10), e2023SW003524. https://doi.org/10.1029/2023SW003524

Chu, X., Bortnik, J., Shen, X.-C., Ma, Q., Li, W., Ma, D., et al. (2024). Imbalanced regression neural network model for whistler-mode hiss waves: Spatial and temporal evolution. *Journal of Geophysical Research: Space Physics*, *129*(8), e2024JA032761. https://doi.org/10.1029/2024JA032761

Chu, X., Hsu, T.-S., McPherron, R. L., Angelopoulos, V., Pu, Z., Weygand, J. J., et al. (2014). Development and validation of inversion technique for substorm current wedge using ground magnetic field data. *Journal of Geophysical Research: Space Physics*, *119*(3), 1909–1924. https://doi.org/10.1002/2013JA019185

Chu, X., Ma, D., Bortnik, J., Tobiska, W. K., Cruz, A., Bouwer, S. D., et al. (2021). Relativistic electron model in the outer radiation belt using a neural network approach. *Space Weather*, *19*(12), e2021SW002808. https://doi.org/10.1029/2021SW002808

Chu, X., McPherron, R., Hsu, T.-S., Angelopoulos, V., Weygand, J. M., Liu, J., & Bortnik, J. (2021). Magnetotail flux accumulation leads to substorm current wedge formation: A case study. *Journal of Geophysical Research: Space Physics*, *126*(1), 2020JA028342. https://doi.org/10.1029/2020JA028342

Chu, X., McPherron, R. L., Hsu, T.-S., & Angelopoulos, V. (2015). Solar cycle dependence of substorm occurrence and duration: Implications for onset. *Journal of Geophysical Research: Space Physics*, *120*(4), 2808–2818. https://doi.org/10.1002/2015JA021104

Chu, X. N., Bortnik, J., Li, W., Ma, Q., Angelopoulos, V., & Thorne, R. M. (2017). Erosion and refilling of the plasmasphere during a geomagnetic storm modeled by a neural network. *Journal of Geophysical Research: Space Physics*, *122*(7), 7118–7129. https://doi.org/10.1002/2017JA023948

Clauer, C. R., McPherron, R. L., Searls, C., & Kivelson, M. G. (1981). Solar wind control of auroral zone geomagnetic activity. *Geophysical Research Letters*, *8*(8), 915–918. https://doi.org/10.1029/GL008i008p00915

Davis, T. N., & Sugiura, M. (1966). Auroral electrojet activity index *AE* and its universal time variations. *Journal of Geophysical Research*, *71*(3), 785–801. https://doi.org/10.1029/JZ071i003p00785

DeJong, A. D., & Clauer, C. R. (2005). Polar UVI images to study steady magnetospheric convection events: Initial results. *Geophysical Research Letters*, *32*(24), L24101. https://doi.org/10.1029/2005GL024498

Dozat, T. (2016). Incorporating Nesterov momentum into Adam. *Proc. ICLR*, 1–4.

Dungey, J. W. (1961). Interplanetary magnetic field and auroral zones. *Physical Review Letters*, *6*(2), 47–48. https://doi.org/10.1103/PhysRevLett.6.47

Ferreira, A. A., & Borges, R. A. (2021). Performance analysis of distinct feed-forward neural networks structures on the AE index prediction. In *2021 IEEE Aerospace conference (50100)* (pp. 1–7). Big Sky. https://doi.org/10.1109/AERO50100.2021.9438504

Fok, M.-C., Moore, T. E., & Spjeldvik, W. N. (2001). Rapid enhancement of radiation belt electron fluxes due to substorm dipolarization of the geomagnetic field. *Journal of Geophysical Research*, *106*(A3), 3873–3881. https://doi.org/10.1029/2000JA000150

Fu, H., Yue, C., Zong, Q.-G., Zhou, X.-Z., & Fu, S. (2021). Statistical characteristics of substorms with different intensity. *Journal of Geophysical Research: Space Physics*, *126*(8), e2021JA029318. https://doi.org/10.1029/2021JA029318

Gavrishchaka, V. V., & Ganguli, S. B. (2001a). Optimization of the neural-network geomagnetic model for forecasting large-amplitude substorm events. *Journal of Geophysical Research*, *106*(A4), 6247–6257. https://doi.org/10.1029/2000JA900137

Gavrishchaka, V. V., & Ganguli, S. B. (2001b). Support vector machine as an efficient tool for high-dimensional data processing: Application to substorm forecasting. *Journal of Geophysical Research*, *106*(A12), 29911–29914. https://doi.org/10.1029/2001JA900118

Gjerloev, J. W. (2012). The SuperMAG data processing technique. *Journal of Geophysical Research*, *117*(A9), A09213. https://doi.org/10.1029/2012JA017683

Gleisner, H., & Lundstedt, H. (1997). Response of the auroral electrojets to the solar wind modeled with neural networks. *Journal of Geophysical Research*, *102*(A7), 14269–14278. https://doi.org/10.1029/96JA03068

Gleisner, H., & Lundstedt, H. (1999). Ring current influence on auroral electrojet predictions. *Annales Geophysicae*, *17*(10), 1268–1275. https://doi.org/10.1007/s00585-999-1268-x

Gleisner, H., & Lundstedt, H. (2001). Auroral electrojet predictions with dynamic neural networks. *Journal of Geophysical Research*, *106*(A11), 24541–24549. https://doi.org/10.1029/2001JA900046

Goldstein, J. (2007). Plasmasphere response: Tutorial and review of recent imaging results. In D. N. Baker, B. Klecker, S. J. Schwartz, R. Schwenn, & R. Von Steiger (Eds.), *Solar Dynamics and Its Effects on the Heliosphere and Earth* (pp. 203–216). Springer. https://doi.org/10.1007/978-0-387-69532-7_14

Gonzalez, W. D., Gonzalez, A. L. C. de, & Tsurutani, B. T. (1995). Geomagnetic response to large-amplitude interplanetary Alfvén wave trains. *Physica Scripta*, *1995*(T60), 140. https://doi.org/10.1088/0031-8949/1995/T60/018

Gopinath, S., & Prince, P. R. (2019). A comparison of machine-learning techniques for the prediction of the auroral electrojet index. *Journal of Earth System Science*, *128*(7), 172. https://doi.org/10.1007/s12040-019-1194-6

Gu, Y., Wei, H.-L., Boynton, R. J., Walker, S. N., & Balikhin, M. A. (2019). System identification and data-driven forecasting of AE index and prediction uncertainty analysis using a new cloud-NARX model. *Journal of Geophysical Research: Space Physics*, *124*(1), 248–263. https://doi.org/10.1029/2018JA025957

Hajra, R., Tsurutani, B. T., Echer, E., GjerloevGonzalez, W. J. W. D., & Gjerloev, J. W. (2016). Supersubstorms (SML 2500nT): Magnetic storm and solar cycle dependences. *Journal of Geophysical Research: Space Physics*, *121*(8), 7805–7816. https://doi.org/10.1002/2015JA021835

Henderson, M. G., Reeves, G. D., Skoug, R., Thomsen, M. T., Denton, M. H., Mende, S. B., et al. (2006). Magnetospheric and auroral activity during the 18 April 2002 sawtooth event. *Journal of Geophysical Research*, *111*(A1), A01S90. https://doi.org/10.1029/2005JA011111

Hernandez, J. V., Tajima, T., & Horton, W. (1993). Neural net forecasting for geomagnetic activity. *Geophysical Research Letters*, *20*(23), 2707–2710. https://doi.org/10.1029/93GL02848

Hones, E. W. E. W. Jr., Akasofu, S.-I., Perreault, P., Bame, S. J., & Singer, S. (1970). Poleward expansion of the auroral oval and associated phenomena in the magnetotail during auroral substorms: 1. *Journal of Geophysical Research*, *75*(34), 7060–7074. https://doi.org/10.1029/JA075i034p07060

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed., pp. 160–164). John Wiley and Sons. Chapter 5.

Hua, M., & Bortnik, J. (2024). Upper limit of outer belt electron acceleration and their controlling geomagnetic conditions: A comparison of storm and non-storm events. *Geophysical Research Letters*, *51*(13), e2024GL109612. https://doi.org/10.1029/2024GL109612

Huang, C.-S., Reeves, G. D., Borovsky, J. E., Skoug, R. M., Pu, Z. Y., & Le, G. (2003). Periodic magnetospheric substorms and their relationship with solar wind variations. *Journal of Geophysical Research*, *108*, 1255. https://doi.org/10.1029/2002JA009704

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*. https://doi.org/10.48550/arXiv.1502.03167

Iyemori, T., Maeda, H., & Kamei, T. (1979). Impulse response of geomagnetic indices to interplanetary magnetic field. *Journal of Geomagnetism and Geoelectricity*, *31*(1), 1–9. https://doi.org/10.5636/jgg.31.1

Jaynes, A. N., Baker, D. N., Singer, H. J., Rodriguez, J. V., Loto'aniu, T. M., Ali, A. F., et al. (2015). Source and seed populations for relativistic electrons: Their roles in radiation belt changes. *Journal of Geophysical Research: Space Physics*, *120*(9), 7240–7254. https://doi.org/10.1002/2015JA021234

Kamide, Y., Baumjohann, W., Daglis, I. A., Gonzalez, W. D., Grande, M., Joselyn, J. A., et al. (1998). Current understanding of magnetic storms: Storm-substorm relationships. *Journal of Geophysical Research*, *103*(A8), 17705–17728. https://doi.org/10.1029/98JA01426

Kepko, L., McPherron, R. L., Amm, O., Apatenkov, S., Baumjohann, W., Birn, J., et al. (2015). Substorm current wedge revisited. *Space Science Reviews*, *190*(1–4), 1–46. https://doi.org/10.1007/s11214-014-0124-9

Kissinger, J., McPherron, R. L., Hsu, T.-S., & Angelopoulos, V. (2011). Steady magnetospheric convection and stream interfaces: Relationship over a solar cycle. *Journal of Geophysical Research*, *116*(A5), A00I19. https://doi.org/10.1029/2010JA015763

Klimas, A. J., Baker, D. N., Roberts, D. A., Fairfield, D. H., & Büchner, J. (1992). A nonlinear dynamical analogue model of geomagnetic activity. *Journal of Geophysical Research*, *97*(A8), 12253–12266. https://doi.org/10.1029/92JA00794

Klimas, A. J., Vassiliadis, D., & Baker, D. N. (1997). Data-derived analogues of the magnetospheric dynamics. *Journal of Geophysical Research*, *102*(A12), 26993–27009. https://doi.org/10.1029/97JA02414

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Li, X., Baker, D. N., Larson, D., Temerin, M., Reeves, G., & Kanekal, S. G. (2003). The predictability of the magnetosphere and space weather. *Eos, Transactions American Geophysical Union*, *84*(37), 361–370. https://doi.org/10.1029/2003EO370002

Li, X., Oh, K. S., & Temerin, M. (2007). Prediction of the AL index using solar wind parameters. *Journal of Geophysical Research*, *112*(A6), A06224. https://doi.org/10.1029/2006JA011918

Luo, B., Li, X., Temerin, M., & Liu, S. (2013). Prediction of the AU, AL, and AE indices using solar wind parameters. *Journal of Geophysical Research: Space Physics*, *118*(12), 7683–7694. https://doi.org/10.1002/2013JA019188

Ma, D., Bortnik, J., Chu, X., Claudepierre, S. G., Ma, Q., & Kellerman, A. (2023). Opening the black box of the radiation belt machine learning model. *Space Weather*, *21*(4), e2022SW003339. https://doi.org/10.1029/2022SW003339

Ma, D., Bortnik, J., Ma, Q., Hua, M., & Chu, X. (2024). Machine learning interpretability of outer radiation belt enhancement and depletion events. *Geophysical Research Letters*, *51*(1), e2023GL106049. https://doi.org/10.1029/2023GL106049

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Maruyama, N. (2020). Chapter 9 - storms and substorms—The new whole system approach and future challenges. In M. Materassi, B. Forte, A. J. Coster, & S. Skone (Eds.), *The Dynamical Ionosphere* (pp. 87–119). Elsevier. https://doi.org/10.1016/B978-0-12-814782-5.00009-1

McPherron, R. L. (1970). Growth phase of magnetospheric substorms. *Journal of Geophysical Research*, *75*(28), 5592–5599. https://doi.org/10.1029/JA075i028p05592

McPherron, R. L. (2023). Substorm triggering by the solar wind. *Journal of Geophysical Research: Space Physics*, *128*(6), e2022JA031147. https://doi.org/10.1029/2022JA031147

McPherron, R. L., Baker, D. N., Baker, D. N., Bargatze, L. F., Bargatze, L. F., & Bargatze, L. F. (1985). Linear filters as a method of real-time prediction of geomagnetic activity, *126*, 85. https://doi.org/10.1007/978-90-277-2303-1_5

McPherron, R. L., Baker, D. N., Bargatze, L. F., Clauer, C. R., & Holzer, R. E. (1988). IMF control of geomagnetic activity. *Advances in Space Research*, *8*(9), 71–86. https://doi.org/10.1016/0273-1177(88)90114-7

McPherron, R. L., Baker, D. N., Pulkkinen, T. I., Hsu, T.-S., Kissinger, J., & Chu, X. (2013). Changes in solar wind–magnetosphere coupling with solar cycle, season, and time relative to stream interfaces. *Journal of Atmospheric and Solar-Terrestrial Physics*, *99*, 1–13. https://doi.org/10.1016/j.jastp.2012.09.003

McPherron, R. L., Hsu, T.-S., & Chu, X. (2015). An optimum solar wind coupling function for the AL index. *Journal of Geophysical Research: Space Physics*, *120*(4), 2494–2515. https://doi.org/10.1002/2014JA020619

McPherron, R. L., Hsu, T. S., Kissinger, J., Chu, X., & Angelopoulos, V. (2011). Characteristics of plasma flows at the inner edge of the plasma sheet. *Journal of Geophysical Research*, *116*(A5), A00I33. https://doi.org/10.1029/2010JA015923

McPherron, R. L., Russell, C. T., & Aubry, M. P. (1973). Satellite studies of magnetospheric substorms on August 15, 1968: 9. Phenomenological model for substorms. *Journal of Geophysical Research*, *78*(16), 3131–3149. https://doi.org/10.1029/JA078i016p03131

Nakano, S., & Kataoka, R. (2022). Echo state network model for analyzing solar-wind effects on the AU and AL indices. *Annales Geophysicae*, *40*(1), 11–22. https://doi.org/10.5194/angeo-40-11-2022

Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. J. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research*, *112*(A1), A01206. https://doi.org/10.1029/2006JA012015

Pallocchia, G., Amata, E., Consolini, G., Marcucci, M. F., & Bertello, I. (2008). AE index forecast at different time scales through an ANN algorithm based on L1 IMF and plasma measurements. *Journal of Atmospheric and Solar-Terrestrial Physics*, *70*(2), 663–668. https://doi.org/10.1016/j.jastp.2007.08.038

Papitashvili, N. E., & King, J. H. (2020). OMNI 1-min data [Dataset]. *OMNI 1-min data*. Retrieved from https://omniweb.gsfc.nasa.gov/

Pu, Z. Y., Chu, X. N., Cao, X., Mishin, V., Angelopoulos, V., Wang, J., et al. (2010). THEMIS observations of substorms on 26 February 2008 initiated by magnetotail reconnection. *Journal of Geophysical Research*, *115*(A2), 2009JA014217. https://doi.org/10.1029/2009JA014217

Pytte, T., McPherron, R. L., Hones, E. W., Jr., & West, H. I., Jr. (1978). Multiple-satellite studies of magnetospheric substorms: Distinction between polar magnetic substorms and convection-driven negative bays. *Journal of Geophysical Research*, *83*(A2), 663–679. https://doi.org/10.1029/JA083iA02p00663

Reeves, G. D., McAdams, K. L., Friedel, R. H. W., & O'Brien, T. P. (2003). Acceleration and loss of relativistic electrons during geomagnetic storms. *Geophysical Research Letters*, *30*(10), 1529. https://doi.org/10.1029/2002GL016513

Sergeev, V. A., & Lennartsson, W. (1988). Plasma sheet at X ≈ −20 RE during steady magnetospheric convection. *Planetary and Space Science*, *36*(4), 353–370. https://doi.org/10.1016/0032-0633(88)90124-9

Sergeev, V. A., Pellinen, R. J., & Pulkkinen, T. I. (1996). Steady magnetospheric convection: A review of recent results. *Space Science Reviews*, *75*(3), 551–604. https://doi.org/10.1007/BF00833344

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

Takalo, J., & Timonen, J. (1997). Neural network prediction of AE data. *Geophysical Research Letters*, *24*(19), 2403–2406. https://doi.org/10.1029/97GL02457

Temerin, M., & Li, X. (2002). A new model for the prediction of Dst on the basis of the solar wind. *Journal of Geophysical Research*, *107*(A12), 1472. https://doi.org/10.1029/2001JA007532

Temerin, M., & Li, X. (2006). Dst model for 1995–2002. *Journal of Geophysical Research*, *111*(A4), A04221. https://doi.org/10.1029/2005JA011257

TensorFlow Developers. (2023). TensorFlow (v2.13.0-rc2) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.8071599

Tsurutani, B. T., & Gonzalez, W. D. (1987). The cause of high-intensity long-duration continuous AE activity (HILDCAAs): Interplanetary Alfvén wave trains. *Planetary and Space Science*, *35*(4), 405–412. https://doi.org/10.1016/0032-0633(87)90097-3

Walsh, B. M., Bhakyapaibul, T., & Zou, Y. (2019). Quantifying the uncertainty of using solar wind measurements for geospace inputs. *Journal of Geophysical Research: Space Physics*, *124*(5), 3291–3302. https://doi.org/10.1029/2019JA026507

Weigel, R. S., Horton, W., Tajima, T., & Detman, T. (1999). Forecasting auroral electrojet activity from solar wind input with neural networks. *Geophysical Research Letters*, *26*(10), 1353–1356. https://doi.org/10.1029/1999GL900280

Werner, A. L. E., Yordanova, E., Dimmock, A. P., & Temmer, M. (2019). Modeling the multiple CME interaction event on 6–9 September 2017 with WSA-ENLIL+Cone. *Space Weather*, *17*(2), 357–369. https://doi.org/10.1029/2018SW001993

Wintoft, P., & Wik, M. (2021). Exploring three recurrent neural network architectures for geomagnetic predictions. *Frontiers in Astronomy and Space Sciences*, *8*. https://doi.org/10.3389/fspas.2021.664483

Xing, X., Lyons, L. R., Angelopoulos, V., Larson, D., McFadden, J., Carlson, C., et al. (2009). Azimuthal plasma pressure gradient in quiet time plasma sheet. *Geophysical Research Letters*, *36*(14), L14105. https://doi.org/10.1029/2009GL038881

Xing, X., Lyons, L. R., Nishimura, Y., Angelopoulos, V., Donovan, E., Spanswick, E., et al. (2011). Near-Earth plasma sheet azimuthal pressure gradient and associated auroral development soon before substorm onset. *Journal of Geophysical Research*, *116*(A47), A07204. https://doi.org/10.1029/2011JA016539

Yao, Z. H., Pu, Z. Y., Fu, S. Y., Angelopoulos, V., Kubyshkina, M., Xing, X., et al. (2012). Mechanism of substorm current wedge formation: THEMIS observations. *Geophysical Research Letters*, *39*(13), L13102. https://doi.org/10.1029/2012GL052055

Zong, Q.-G., Yue, C., & Fu, S.-Y. (2021). Shock induced strong substorms and super substorms: Preconditions and associated Oxygen ion dynamics. *Space Science Reviews*, *217*(2), 33. https://doi.org/10.1007/s11214-021-00806-x

Zou, Z., Huang, H., Zuo, P., Ni, B., San, W., Yuan, Q., et al. (2024). A forecast model of geomagnetic indices from the solar wind fluids observations based on long short-term memory neural network. *Physics of Fluids*, *36*(2), 026616. https://doi.org/10.1063/5.0196284