Medium Energy Electron Flux in Earth's Outer Radiation Belt (MERLIN): A Machine Learning Model

Artem Smirnov[1], Max Berrendorf[2], Yuri Shprits[3], Elena A. Kronberg[2], Hayley J Allison[4], Nikita A Aseev[5], Irina S. Zhelavskaya[6], Steven K. Morley[7], Geoffrey D. Reeves[8], Matthew R. Carver[7], and Frederic Effenberger[9]

[1]Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences
[2]Ludwig Maximilian University of Munich
[3]GFZ Helmholtz Center Potsdam
[4]GFZ
[5]GFZ German Research Centre for Geosciences
[6]Helmholtz Centre Potsdam, GFZ German Research Centre For Geosciences, Potsdam, Germany
[7]Los Alamos National Laboratory (DOE)
[8]Space Science and Applications Group, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.
[9]Helmholtz Centre Potsdam, GFZ, German Research Centre for Geosciences, Potsdam, Germany

November 26, 2022

Abstract

The radiation belts of the Earth, filled with energetic electrons, comprise complex and dynamic systems that pose a significant threat to a variety of satellite systems. While various models of the relativistic electron flux have been developed for geostationary orbit (GEO), the behaviour of the medium energy (120-600 keV) electrons below GEO remains poorly quantified. In this paper we present a Medium Energy electRon flux In Earth's outer radiatioN belt (MERLIN) model based on the Light Gradient Boosting (LightGBM) machine learning algorithm. The MERLIN model takes as input the satellite position, a combination of geomagnetic indices and solar wind parameters including the time history of velocity, and does not use persistence. MERLIN is trained and validated on \$>\$15 years of the GPS electron flux data, and tested on more than \$1.5\$ years of measurements. 10-fold cross validation (CV) yields that the model predicts the MEO radiation environment well, both in terms of dynamics and amplitudes of flux. Evaluation on the test set yields high correlation between the predicted and observed electron flux (0.8) and low values of absolute error. The MERLIN model can have wide Space Weather applications, providing information for the scientific community in the form of radiation belts reconstructions, as well as industry for satellite mission design, nowcast of the MEO environment and surface charging analysis.

Medium Energy Electron Flux in Earth's Outer Radiation Belt (MERLIN): A Machine Learning Model

A.G. Smirnov[1,2], M. Berrendorf[3], Y.Y. Shprits[1,2,5], E.A. Kronberg[4,6], H.J. Allison[1], N.A. Aseev[1,2], I.S. Zhelavskaya[1,2], S.K. Morley[7], G.D. Reeves[7], M.R. Carver[7], F. Effenberger[1]

[1]Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany.
[2]Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany.
[3]Department of Database Systems and Data Mining, Ludwig-Maximilians University of Munich, Munich, Germany.
[4]Department of Earth and Environmental Sciences, Ludwig Maximilians University of Munich, Munich, Germany.
[5]Department of Earth, Planetary and Space Sciences, University of California, Los Angeles, California, USA.
[6]Max Planck Institute for Solar System Research, Göttingen, Germany.
[7]Space Science and Applications, Los Alamos National Laboratory, New Mexico, USA.

Key Points:

- A machine learning model is created to predict electron flux at MEO for energies 120-600 keV.
- The model requires solar wind parameters and geomagnetic indices as input and does not use persistence.
- MERLIN model yields high accuracy and high correlation with observations (0.8).

Corresponding author: Artem Smirnov, `artem.smirnov@gfz-potsdam.de`

Abstract

The radiation belts of the Earth, filled with energetic electrons, comprise complex and dynamic systems that pose a significant threat to a variety of satellite systems. While various models of the relativistic electron flux have been developed for geostationary orbit (GEO), the behaviour of the medium energy (120-600 keV) electrons below GEO remains poorly quantified. In this paper we present a Medium Energy electRon flux In Earth's outer radiatioN belt (MERLIN) model based on the Light Gradient Boosting (LightGBM) machine learning algorithm. The MERLIN model takes as input the satellite position, a combination of geomagnetic indices and solar wind parameters including the time history of velocity, and does not use persistence. MERLIN is trained and validated on $>$15 years of the GPS electron flux data, and tested on more than 1.5 years of measurements. 10-fold cross validation (CV) yields that the model predicts the MEO radiation environment well, both in terms of dynamics and amplitudes of flux. Evaluation on the test set yields high correlation between the predicted and observed electron flux (0.8) and low values of absolute error. The MERLIN model can have wide Space Weather applications, providing information for the scientific community in the form of radiation belts reconstructions, as well as industry for satellite mission design, nowcast of the MEO environment and surface charging analysis.

Plain Language Summary

The radiation belts of the Earth, which are the zones of charged energetic particles trapped by the geomagnetic field, comprise complex and dynamic systems posing a significant threat to a variety of commercial and military satellite systems. While the inner belt is relatively stable, the outer belt is highly variable and depends substantially on solar activity; therefore, accurate and improved models of electron flux in the outer radiation belt are essential to understand the underlying physical processes. Although many models have been developed for the geostationary orbit and relativistic energies, prediction of electron flux in the 120-600 keV energy range still remains challenging. We present a data-driven model of the medium energies (120-600 keV) differential electron flux in the outer radiation belt based on machine learning. We use 17 years of electron observations by Global Positioning System (GPS) satellites. We set up a 3D model for flux prediction in terms of L-values, MLT and satellite latitude. The model gives reliable predictions of the radiation environment in the outer radiation belt and has wide space weather applications.

1 Introduction

The Van Allen radiation belts, discovered by the array of Explorer satellites [Van Allen and Frank, 1959], are zones of charged energetic particles, mainly electrons and protons, trapped by the magnetic field of the Earth. Protons form a single radiation belt with maximum flux intensities between L values from $\sim$ 3 to 4 [Ganushkina et al., 2011]. Energetic electrons ($>$ 100 keV) are mainly confined to two regions - the inner belt, within L from 1.2 to 2.5, and the outer belt located between L from $\sim$ 3 to 7 [Lyons et al., 1972; Summers et al., 2004]. The inner and outer electron radiation belts are separated with a so-called slot region, usually devoid of energetic electrons [Lyons and Thorne, 1973; Kavanagh et al., 2018]. The inner radiation belt is known to exhibit long-term stability, while the outer belt is highly dynamic and depends substantially on solar activity [Meredith et al., 2006].

The dynamics of the outer radiation belt is governed by a complex interplay between acceleration and loss processes [Reeves et al., 2003]. Electrons with energies of tens of keV, called the *source population*, are injected into the inner magnetosphere during substorms and produce waves, for instance, the whistler mode chorus [e.g., Boyd et al., 2014, 2016; Jaynes et al., 2015]. Electrons with energies of hundreds of keV, called the *seed population* electrons, are also injected in the magnetosphere during substorm activity. These electron populations can accumulate at the surface of the spacecraft and lead to satellite loss due to the so-called surface charging effects [e.g., Garrett, 1981; Lanzerotti et al., 1998]. Furthermore, the seed population electrons can be accelerated to relativistic energies by waves. These relativistic (>1 MeV) particles can penetrate through satellite shielding and damage the equipment onboard, also leading to satellite loss [e.g., Fennell et al., 2000].

To date, there are more than 2200 operational satellites in the Earth's orbit and many of them systematically pass through the radiation belts region. Approximately 1400 spacecraft are in Low Earth Orbit (LEO) at altitudes up to 1000 km. The LEO satellites cross the inner radiation belt in the region of the South Atlantic magnetic anomaly (SAA) and the outer belt at higher latitudes. The second most populated is the geostationary orbit (GEO) with more than 560 satellites flying at altitudes of $\sim$36000 km synchronously with the rotation of the Earth. GEO satellites generally fly close to the outer edge of the outer radiation belt ($L \sim 6.6$). Satellites flying below GEO and above LEO follow the so-called Medium Earth Orbit (MEO). Many Global Navigation Satellite System (GNSS) satellites fly at MEO, for instance, the Global Positioning System (GPS), GLObal NAvigation Satellite System (GLONASS) and Galileo. Furthermore, in order to reach GEO, an increasing number of spacecraft are using the electric orbit raising method and can spend hundreds of days in the MEO region [Horne and Pitchford, 2015; Glauert et al., 2018]. Satellites following MEO systematically pass through the heart of the outer radiation belt and are exposed to the largest values of electron flux. The number of satellites in Earth's orbit will increase significantly in the following years, and in order to ensure the long-term satellite operation stability, it is necessary to have reliable models of electron intensities at different energies (from tens of keV up to several MeV) and locations.

The existing radiation belt models can be divided into three main categories: physics-based, data-driven, and data assimilation models. Several physics-based models of electron flux have been created for the radiation belts and ring current region. Among them, there are the Versatile Electron Radiation Belt (VERB) [e.g., Subbotin and Shprits, 2009], British Antarctic Survey Radiation Belts Model (BAS-RBM) [Glauert et al., 2014] and Dynamic Radiation Environment Assimilation Model (DREAM) [Reeves et al., 2012; Tu et al., 2013, 2014] codes based on solving the three-dimensional Fokker-Planck equation to reproduce the dynamics and variability of the MeV radiation belts electrons. The physics-based models typically include the radial diffusion, losses due to pitch-angle scattering and magnetopause shadowing [Glauert et al., 2018]. Recently, the VERB-4D code has been developed to extend the VERB code to lower-energy ring current electrons by including advection terms [Shprits et al., 2015; Aseev et al., 2016]. The physics-based Inner Magnetosphere Particle Transport Model (IMPTAM) was developed and shown to give reasonable flux predictions at energies from several eV up to <150 keV [Ganushkina et al., 2019]. The low energy electrons are also modeled by the coupled Fok Ring Current (FRC) [Fok and Moore, 1997] and Comprehensive Inner-Magnetosphere Ionosphere (CIMI) [Fok et al., 2001] models operating online. It should be noted that there is generally a gap in modeling the electron flux at energies 100-600 keV. Physical modeling at these energies is considered difficult due to the fact that electric field effects have to be considered [Ganushkina et al., 2011] and also because the physics governing the dynamics of electrons at medium energies is not entirely understood [Horne et al., 2013].

The data-driven models can be subdivided into static and dynamic ones. AE8 [Vampola, 1997] and AE9 [Ginet et al., 2013] are examples of the static models providing the values of integral flux at energies 40 keV - 7 MeV, although we note that different versions exist for the times of solar minimum and maximum. AE8 and AE9 models overcome the limitations of the individual data sets by combining large scale statistics and are currently used as a reference for engineering purposes [Glauert et al., 2018]. Dynamic data-driven models typically depend on a combination of solar wind parameters and geomagnetic indices. Several data-driven models have been developed for the GEO orbit. Denton et al. [2015] developed an empirical model of electron flux for low energies (10 eV - 40 keV) based on 82 LANL satellites data, driven by the Kp index. Later, the upstream solar wind conditions were incorporated into the model [Denton et al., 2016], and it was also expanded to 6-20 R_E [Denton et al., 2019] using Cluster data. For relativistic energies, Balikhin et al. [2011] employed a Nonlinear AutoRegressive Moving Average with eXogeneous inputs (NARMAX) technique to predict daily flux at GEO orbit at energies 800 keV and 2 MeV, using solar wind parameters and the previous time-series of daily flux from GOES satellites. The NARMAX model was further extended to a broader energy range, including electron flux at energies of hundreds of keV [Boynton et al., 2016]. However, the model needs satellite time-series as inputs and therefore for now is confined to the geostationary orbit. Several models of the relativistic electron flux are based on the Artificial Neural Networks (ANN), e.g., Ling et al. [2010]; Kitamura et al. [2011]; Chen et al. [2019]; de Lima et al. [2020]. Other empirical models of relativistic flux developed for the GEO region include the Relativistic Electron Flux Model (REFM) driven by solar wind velocity [Baker et al., 1990], an empirical function D_0 dependent on several solar wind parameters and Kp [Li, 2004], and linear regression models [e.g., Sakaguchi et al., 2015; Simms et al., 2014, 2016; Osthus et al., 2014] which take as inputs the solar wind parameters and previous values of flux at GEO. Tsutai et al. [1999] used linear filter to predict the values of $>$ 2MeV flux at GEO 1 day ahead using GOES magnetic field data over the preceding 6 days.

Although a variety of models have been developed for the geostationary orbit, few data driven models exist that give reliable electron flux predictions at MEO. This is due to the fact that many GEO satellites provide continuous high quality observations of electron flux, while at MEO the temporal and spatial coverage of observations remains rather sparse [Sakaguchi et al., 2015]. Indeed, only $\sim$ 100 satellites reside in MEO, and only few of them provide measurements of the radiation belt populations. Among other data sets of electron flux measurements in the MEO region, the recently released GPS energetic particle data have notable advantages such as, for instance, the large number of satellites (23) and uniform MLT coverage, as well as availability of 18 years of observations covering almost 2 solar cycles. Furthermore, most of the GPS satellites carry onboard identical Combined X-ray Dosimeter (CXD) detectors measuring electron flux at energies 0.12-10 MeV. The GPS/CXD data have been inter-calibrated with Van Allen Probes electron flux measurements and the two missions were in good agreement at energies below 4 MeV [Morley et al., 2016].

In the current paper we present the data-driven Medium EneRgy ELectron flux In the outer RadiatioN Belt (MERLIN) model, based on machine learning. For model training we employ the Light Gradient Boosting Machine (LightGBM) algorithm, which is known for its high efficiency and accuracy [Ke et al., 2017]. The model takes as input satellite position in an Lshell - MLT - latitude frame, solar wind parameters with history of velocity, and geomagnetic indices. The model returns values of spin-averaged electron flux at energies 120-600 keV as outputs. The structure of the paper takes the form of five parts, including this introductory section. Section 2 describes the data set used for model construction. Section 3 is concerned with the methodology used for this study. Section 4 presents the results. The conclusions are drawn in the final section.

2 Data set

2.1 GPS electron flux data

The GPS spacecraft are distributed across six orbital planes, nominally inclined at 55°. The satellites follow near-circular medium Earth orbit, with 12h revolution period, at a constant altitude. As of 2020, the constellation consists of 74 spacecraft, of which 31 are operational, 9 reserve, 2 being tested and 32 no longer in use [www.gps.gov]. Due to its fixed altitude of 20,200 km (R$\sim$ 4.2), the GPS constellation travels through a range of L-shells providing the particle measurements in the outer radiation belt. We note, however, that the inclination of the GPS orbit restricts the range of equatorial pitch angles as a function of L-shell. The satellite at off-equatorial magnetic latitudes (MLAT) cannot observe the particles mirroring at lower MLATs and therefore samples only a part of the equatorial pitch angle distribution. Furthermore, GPS does not resolve local pitch angles.

Since the early 2000s, newer GPS satellites are equipped with either of the two instrument series: the improved BDD-IIR or the Combined X-ray Dosimeter (CXD). Most of the satellites currently carry aboard the identical CXD detectors. Their response is well-known and their electron flux data have been used in previous radiation belts studies [e.g., Olifer et al., 2018; Pinto et al., 2020]. The CXD instruments measure the electron flux using two sensors: the low energy particle (LEP) and the High Energy Particle (HEP) sensors, with the typical sampling rate of 240 seconds [Morley et al., 2016]. In the present study we use data of the first 5 evaluated CXD energies, namely 120, 210, 300, 425 and 600 keV. As of 2020, 21 GPS satellites are equipped with the CXD detectors, providing more than 200 years of satellite data. The CXD measurements were previously inter-calibrated with the Van Allen Probes electron flux and showed good agreement. Having identified 140 physical GPS - RBSP conjunctions, Morley et al. [2016] obtained the ratios between the count rates of two instruments, which were close to 1 at energies below 4 MeV, while at higher energies larger variance was observed due to the unaccounted instrumental backgrounds.

2.2 Solar wind and geomagnetic indices

The relationship between the electron flux intensities in the outer radiation belts and solar wind parameters has long been recognized [e.g., Paulikas and Blake, 1979; Reeves et al., 2011]. Numerous studies have analyzed contributions of the solar wind parameters to flux enhancements. The independent contributions of solar wind velocity and number density were investigated, for instance, in Balikhin et al. [2011]; Kellerman and Shprits [2012]; Simms et al. [2014]. A combination of velocity and density, pressure and geomagnetic indices, combined with the previous daily flux value, was examined by Sakaguchi et al. [2015]. Long-term relationship between velocities and MeV electron fluxes was discussed in Reeves et al. [2013].

It has been well established that the radiation belts flux enhancements are connected with changes in solar wind speed. Reeves et al. [2011] analyzed the relativistic electron flux at GEO with respect to the solar wind speed and noted that the resulting distribution resembled a triangle. Such a shape was explained as follows. The V_{sw} values rarely fall below 300 km/s, and this leads to a left-hand side of the triangle [see also Wing et al., 2016]. The top side of the triangle forms due to the fact that the flux values seem to have a sharp maximum at higher V_{sw}, for which multiple explanations have been given. One of the most puzzling features of the triangular distribution is that the variability of electron flux at lower V_{sw} is much larger than at higher V_{sw}. Reeves et al. [2011] noted that the electron flux can exhibit large values under any V_{sw} values. The triangular form demonstrates that using the values of the solar wind velocity and density is not enough to fully explain the variability of flux and therefore other parameters have to be taken into account.

We consider the following solar wind parameters and geomagnetic indices obtained at the OMNIWeb database [omniweb.gsfc.nasa.gov]. First, amongst the solar wind drivers, we analyze the solar wind velocity, and its components V_x, V_y, V_z. We analyze the IMF magnitude, B_x, B_y and B_z components, and also solar wind density n_{sw}. We employ the derived solar wind quantities: magnetosonic and alfvenic Mach numbers (Mach_a and Mach_m,respectively), solar wind temperature T_{sw}, electric field (v*B_z), dynamic pressure (Pdyn) and plasma Beta. From geomagnetic indices, we select SYM-H, SYM-D, ASYM-H and ASYM-D indices, planetary Kp index and auroral AL, AU and AE indices. It has been previously established that many of these features are, in fact, correlated with one another. In Figure 1, we show the Pearson linear correlations between different solar wind and geomagnetic parameters in order to examine which features can be used for the model set up.

We find that V_x is perfectly anticorrelated with V, with the -1.0 coefficient, which is as expected since V_x consitutes most of the V amplitude. V_y correlates with V with only 0.19 correlation, and V_z shows zero correlation with velocity magnitude. SYM-D does not correlate with any of the features, except for very weak (0.14) relationship with SYM-H. SYM-H, on the other hand, correlates with several parameters. For instance, it exhibits a moderate positive correlation with the solar wind velocity (0.43), and negative correlation with ASYM-H (-0.6) and ASYM-D (-0.47). Furthermore, it correlates with auroral geomagnetic indices with approximately 0.5 correlation coefficient, and is also weakly anti-correlated with the electric field (v*B_z). ASYM-D index is correlated with ASYM-H with the R-value of 0.65, and also with auroral indices with the absolute value of the linear correlation ∼0.6. In turn, ASYM-H index shows weak linear correlation with solar wind velocity (0.32), IMF (0.49), dynamic pressure (0.36), and exhibits higher correlation with the auroral indices with the corresponding R values of up to 0.7. Kp index exhibits moderate positive correlation with the IMF magnitude (R=0.56), solar wind velocity (0.59), temperature and dynamic pressure (R∼0.5), stronger positive correlation with AE (R=0.7) and the corresponding AU and AL indices, along with 0.5 correlation with SYM-H. It should be noted that the Kp index has a 3-hour cadence, and therefore shows lower correlation with AE than one would expect. By averaging the AE index to the same 3h cadence, one obtains a correlation of 0.82. IMF magnitude shows weak correlation (0.3) with solar wind density and temperature, moderate (0.5) correlation with dynamic pressure and Mach numbers.

We note that although the gradient boosting regression trees are prone to the multi-collinearity of features [e.g., Maloney et al., 2012; Ding et al., 2016], using highly correlated inputs can pose a disadvantage for machine learning studies. For example, when 2 parameters are correlated we can achieve the same reduction in variance as by using only one of them. Here we remove several correlated and derived quantities leaving the more in-depth analysis of the influence that different parameters have on the electron flux for further studies. First, we exclude directional components of magnetic field and velocity, since their information is already contained in the magnitude values (e.g., V and V_x correlate with $R = -1$). Furthermore, we exclude all of the derived quantities, because they encompass information of their original constituent variables (for instance, dynamic pressure strongly correlates with density). Magnetosonic and Alfvenic Mach numbers essentially represent the normalized velocity and it is enough to consider the velocity itself. We apply the same reasoning to the geomagnetic indices selection: AE is a product of AL and AU indices, and also correlates with them, which is why we only use AE for model set up. AE and Kp indices are generally strongly correlated. While Kp is a measure of the planetary geomagnetic activity, substorms are better resolved by the AE index. Furthermore, Smirnov et al. [2019] reported the long-term positive correlation of electron flux at energies up to 400 keV with AE index along the solar cycles 23 and 24. The same conclusion was drawn in [Smirnov et al., 2020] by analyzing the long-term phase space density (PSD) variations of low-μ electrons, indicating the importance of the substorm activity for radiation belts transport processes. For these reasons we include both Kp and AE indices as inputs.

The inner edge of the outer belt is highly dynamic and can move inwards during slot-filling events and outwards during the quiet periods. Li et al. [2006] reported a correlation between the 30-day averages of the innermost edge of the outer belt and the plasmapause location (Lpp) using 12 years of SAMPEX data. The flux values in the slot region, located below the Lpp, are lower than those beyond the plasmapause due to loss processes attributed to storm-enhanced EMIC and plasmaspheric hiss waves [Li et al., 2006] . O'Brien and Moldwin [2003] presented a model of the plasmapause location parametrized as a function of the maximum AE value over preceding 36 hours. Furthermore, the Lpp model based on the AE index was found to perform better than that using the Dst values. In order to account for the dynamics of the inner edge of the outer belt, we include the maximum value of the AE over 36 hours as an input parameter. We do not apply the linear regression coefficients to convert the max(AE) to Lpp, as the Regression Trees are invariant to linear scaling operations [e.g., Druzhkov et al., 2011].

After the enhancement events, the flux of medium energy electrons decay to their pre-storm values gradually over a period of up to 20 days [e.g., Meredith et al., 2006]. Such a slow decay can be explained by the longer hiss lifetimes, which by different estimates vary from several up to tens of days [Orlova et al., 2016]. Hence, it is crucial to include some indication of the previous state of the radiation belts into the model. This is usually done by adding the preceding values of flux as model inputs [e.g., Simms et al., 2016, 2014; Boynton et al., 2016]. Instead, in the MERLIN model we include the history of solar wind velocity as a proxy of the previous activity. We use the averaged v_{sw} values over the preceding 1, 2, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42 hours and 2, 3, 7 and 14 days. It should be noted that the averages over longer periods of time also carry part of the information from shorter scales averages. In this study we only consider the history of solar wind velocity, while adding the history of number density leads to overfitting, as discussed in Section 4.1. In sum, as input parameters we select the satellite position in the L-MLT-latitude frame, geomagnetic indices SYM-H, Kp and AE, solar wind parameters - number density, electric field (v*B_z) and velocity with 2 weeks of history (see also figures 2 and 4). More features can be incorporated into the model in future, however preserving the methodology, and the more refined feature selection will be performed in a separate study.

3 Methodology

Large and growing volumes of data have been provided by the satellite missions in the Earth's radiation belts region. One of the efficient ways to utilize these long-term data sets for modeling is to apply machine learning (ML) techniques. Over the years, machine learning has found numerous applications in the field of space physics research. ML methods have been employed for the geomagnetic indices forecast [e.g., Bala et al., 2009; Shprits et al., 2019], global reconstructions of the plasmaspheric dynamics [Zhelavskaya et al., 2017; Bortnik et al., 2018], solar activity prediction [e.g., Colak and Qahwaji, 2009]. Prediction of the electron flux in the outer radiation belt remains one of the most challenging tasks in the space weather research [Camporeale, 2019]. In the present study, we use the Light Gradient Boosting approach, described in detail below, to predict the flux of medium energy electrons using the GPS particle data.

3.1 Light Gradient Boosting Machine (LightGBM)

One of the predictive approaches, widely used in machine learning, is the so-called Gradient Boosting Decision Tree (GBDT) method. The GBDT algorithms gained popularity for being efficient, highly accurate and interpretable. GBDT is an ensemble model of usually shallow decision trees, also called weak learners, trained in sequence [Friedman, 2001]. Growing each individual tree starts from the source set contained in a *root node* of the tree (shown in Figure 2). When a split is made, the root node is divided into two subsets, and 2 branches are generated [Ren et al., 2019]. The procedure is repeated recursively until either the subset at each node contains all identical values of the target variable, or when the splitting is constrained by the algorithm's hyperparameters (e.g., `max_depth` or `num_leaves` is reached). The GBDTs are grown iteratively and each new tree fits the residuals of the previous iteration to account for the mis-modeled instances [Freund et al., 1999]. GBDTs have been applied successfully to many machine learning problems, performing well for regression and classification tasks alike. Numerous GBDT implementations have been developed, starting from Adaptive boosting (AdaBoost - [Freund et al., 1999]). One of the most popular gradient boosting methods up to date is the Extreme Gradient Boosting Machine (XGBoost) [Chen and Guestrin, 2016], famous for winning machine learning competitions and out-performing even the deep learning neural network (NN) models on tabular data. Even though the gradient boosting methods are capable of giving high quality predictions, their main limitation is the unsatisfactorily long training time and poor scalability [e.g., Zhang et al., 2019].

The main cost in GBDT lies in learning the decision trees, and the most time-consuming part in learning each tree is finding the optimal segmentation points. In the conventional algorithms this is usually done using the so-called pre-sorted algorithm. This method enumerates all possible split points on the pre-sorted feature values. While being simple and effective, this method is also known to be inefficient in training speed and memory consumption [Ke et al., 2017]. Another more recent method is the histogram-based approach. It divides the continuous features into k intervals and selects the split points from those k values [Ju et al., 2019]. Such an approach also has regularization effect and helps prevent overfitting.

One of the GBDT implementations, called the Light Gradient Boosting Machine (LightGBM), has been recently developed by Microsoft [Ke et al., 2017]. LightGBM addresses the traditional GBDT performance issues by, first, using the histogram approach to find segmentation points, and second by uzilizing a different approach to the tree growth. The conventional gradient boosting, and also other tree-based methods such as random forest, grow the trees level-wise. This means that when it is necessary to make a new split, a new level of leaves will be grown. In contrast, the LightGBM method grows the trees leaf-wise which adds only one more leaf and not level when a split is made. Such an approach leads to much faster and less computationally expensive implementation of the gradient boosting [Ke et al., 2017]. It has been demonstrated that LightGBM can be as much as 20 times faster than XGBoost while reducing more loss.

The objective of this study is to predict values of electron flux at a range of L-shells throughout the outer radiation belt. Since the quantity being modeled can be represented by real numbers, we construct a regression gradient boosting model. The model set up is described in detail in section 3.3.

3.2 Test - train splitting of the data

Any supervised machine learning model learns on the so-called training dataset. The training set is seen by the model and usually contains most of the employed data points. Machine learning algorithms are trained iteratively trying to reduce the cost function value at each training iteration [Camporeale, 2019]. At some point, the model learns not only the useful dependencies from the data, but also noise due to reducing the cost function to extremely low values, which results in over-fitting. The performance of the model cannot be adequately evaluated on the data that were used to train it. Therefore, another set is needed to give an unbiased estimate of model performance and also for tuning the hyperparameters. This second set is called a *validation set*. Since these data had not been seen during training, the loss function value would decrease only in case when the model captured the underlying general dependencies from the data. However, it has to be noted that since the loss function is being routinely evaluated on the validation set, the model occasionally sees it as well, although never learns from it. Therefore, it is essential that after training the model and checking its performance on the validation set, the model be evaluated one last time on the fraction of data that had never been used before, neither for training nor validation. The data set used for this purpose is called a *test set*. In many machine learning competitions (e.g., Kaggle), only the training and validation sets are given to the competing teams, while the test set is released after the model submissions and decides the winner.

Multiple ways have been proposed to separate the data set into the train, validation and test parts. In fact, very different strategies can be applied based on the field and objective of the study. For instance, in social sciences the accepted methodology would be to use the random test-train split, that is, when points for train, validation and test sets are selected randomly from the original full dataset. This, however, should not be applied for modeling the time-series and physical processes, due to the fact the validation points (usually 10-20 percent of values) can then be obtained by linearly interpolating the typically larger training set. Such a selection technique might introduce what is referred to as the *data leakage* [e.g., Camporeale, 2019]. Hence, it is important to validate, and then test the model on the *events* unseen during training. For that purpose, we should select the consecutive time intervals for validation and testing. One of the ways to achieve this is to use, for example, first 80 percent of data for training, the next 10 percent for validation and then test the model on the last 10 percent of the data. However, in case of magnetospheric phenomena we have to take into account the solar cycle evolution of the processes we model. Indeed, the radiation belts dynamics during the descending and quiet phases of the solar cycle are vastly different. Therefore, the model will be trained on some part of the solar cycle and validated on another, which would not yield an adequate estimate of its performance. In order to take the solar cycle dependence of the radiation belt dynamics into account, we perform a 10-fold cross validation (CV), described below.

We first reserve $>$1.5 years of data (March 2016 - January 2018) for testing the model. This is done due to the fact that on one hand, this interval had numerous solar wind enhancement events and also had enough data points (∼1.100.000) for such evaluation. The entire data set consists of ~ 5.5 million data points, and therefore approximately 20% are reserved for test set.

The rest of the dataset is used for the K-fold CV, illustrated in Figure 3. The data are divided into K roughly equal parts (in our case, K=10). At every such split the model is fitted on the (K-1) parts and validated on 1 part that was left out. For instance, during the first split we use the first fold for validation and fit the model on the rest of the data. This splitting process is repeated K times, each time selecting a different interval for validation. The K-fold allows one to utilize all observations for training and evaluating the model, and each of the data points is used for validation only once. The K-fold CV is used to optimize the hyperparameters and also to retrieve the accuracy of the model, averaged over different phases of the solar cycle.

3.3 Model Set Up

The present study is based on 17 years of GPS/CXD electron flux measurementsat energies 120-600 keV. The flux values were cleaned using the flags, and also outliers in the data were removed by setting the minimum allowed flux values to 1 cm^{-2}kev^{-1}s^{-1}sr^{-1}. We use data from 21 satellites (ns53−ns73) carrying the CXD detectors [for details see Carver et al., 2018]. We train the model on data with the original cadence of measurements equal 240 seconds. Before fitting the model, we applied the base 10 logarithm to the GPS electron flux, as the data variance can be up to several orders of magnitude. In this study, we use the `LightGBMRegressor` method as implemented in the Python version of `lightgbm` library [Ke et al., 2017].

LightGBM has a variety of algorithm parameters, also called hyperparameters, that can have a non-negligible effect on the model performance. Several of them have to remain fixed, while others need to be optimized to achieve higher accuracy. Some of the parameters that do not change, include the `objective` (in our case, `regression`), the booster method (we use `gbdt`, although we note that other possibilities are available using the novel DART and GOSS methods [Ke et al., 2017]), `metrics` of the loss function values (we select both L1 and L2 metrics, representing the MSE and MAE, respectively), and the `early_stopping_rounds` parameter which is used to stop the model training once overfitting occurs. The `learning_rate` controls how much the model is adjusted on each iteration. In case of the high learning rate, the algorithm makes faster fits that can cause overfitting, while in case of extremely low values the training speed drops sufficiently and more iterations are needed to reach convergence. We select the `learning_rate` of 0.05 and keep it fixed throughout further tuning. Other parameters need to be optimized. The most sensitive hyperparameter is the maximum number of leaves in a tree, or `num_leaves`. Deeper trees have better learning capabilities, but since the gradient boosting model represents an ensemble of weak learners, the `num_leaves` is usually not very high, in our case ranging from 15 to 250. Another important parameter is the minimum number of data points in leaf (`min_data_in_leaf`), which has regularization effect and stops the model from learning the noise. However, the large values can cause decrease in model accuracy. One can also use the subset of the input features for training each individual tree by setting the `colsample_by_tree` value. To optimize the said hyperparameters we use the `hyperopt` Python library [Bergstra et al., 2013] which employs the Tree of Parzen Estimator (TPE) approach. The resulting hyperparameters values, as well as their search domains are given in Table 3.3.

4 Results and discussion

4.1 Feature importances

One of the advantages of the tree-based machine learning methods, LightGBM included, lies in the fact that the resulting model can be easily interpreted. Every tree comprising the model can be visualized and can give direct information about the individual split gains, internal values in each leaf and the decision making process in general. In practice, it is often not possible

Table 1. LightGBM Hyperparameters used for model set up. First 5 parameters were kept fixed, while the next ones were optimized using HyperOpt.

Name	Search Range	Value
'objective'	fixed	'regression'
'boosting_type'	fixed	'gbdt'
'metric'	fixed	'L1' & 'L2'
'learning_rate'	fixed	0.05
'early_stopping_rounds'	fixed	50
'num_leaves'	$15-250$	69
'reg_alpha'	$0-1$	0.95
'reg_lambda'	$0-1$	0.03
'min_data_in_leaf'	$10-1000$	18
'colsample_by_tree'	$0.7-1$	0.99

to analyze the model in this way due to the large number of trees in an ensemble (in our case, $\sim$300). The insight on the model construction can then be obtained indirectly, for example, by analyzing the importance scores of each variable, also called feature importances. They are computed for each GBDT and then averaged across all of the trees forming the model. There are multiple ways to retrieve the feature importances. LightGBM utilizes the so-called `split` or `gain` methods. The `split` method counts the number of times each variable was used to make a split. The `gain` method summarizes all gains of splits which use each of the features. It has been well established that the two methods can, in fact, give different results, and also that feature importances estimated this way only carry information about how the particular model was constructed, rather than physical meaning. Furthermore, removing one of the features can potentially redistribute its feature importance between several other variables and yield a different result altogether. Other methods, which are more stable, include the mutual information criterion, permutation method [Auret and Aldrich, 2011] and the recently developed Shapley values technique [Lundberg and Lee, 2017]. These methods require an in-depth analysis which is beyond the scope of the present paper and will be evaluated in future studies. In the present section we confine to describing of the key attributes of the MERLIN model. Feature importances estimated using the `split` and `gain` methods are shown in Figure 4. MERLIN uses the satellite position (Lshell, latitude, MLT), as well as values of the SYM-H, Kp, and AE indices and solar wind density (n_sw), IMF, electric field (El_field) and solar wind velocity. Plasmapause location is denoted as Lpp. Furthermore, the time history of velocity is incorporated into the model in the form of averages over the certain time intervals. We use the progressively increasing time steps, from 1 hour (1h) up to 3 days (3d) and also the averages over 1 and 2 weeks (2w).

We find that the most important features by gain are the L-shell and maximum of AE over 36 hours, which is a proxy of the plasmapause location (Figure 4). Indeed, the values of electron flux depend on these quantities to a large extent, due to the fact that the electron intensities are higher in the heart of the outer belt, and then decrease to the outer edge of the belt and also in the opposite direction towards the slot region, which is reflected in the L-values. The importance of the plasmapause location has been discussed in [Li et al., 2006], and it was shown that Lpp correlated with the inner edge of the outer belt. The flux values drop significantly below the Lpp, although it is of note that the relationship between the innermost location of the outer belt and Lpp is energy dependent [e.g., Reeves et al., 2016; Ripoll et al., 2016]. Among the instantaneous values of solar wind and geomagnetic indices, the SYM-H index, which is a proxy of geomagnetic storms, shows the most importance in both split and gain methods. We also note that the time history of solar wind velocity plays an important role. For the 210 keV electrons, the average velocity over 6 hours and also over 1 and 2 weeks have the most importance based on impurity reduction (Fig. 4). The importance of the 2 weeks average of v_{sw} likely comes from the fact that following the flux enhancement events, velocity drops to quiet time values faster than the electron intensities which remain at elevated levels for longer periods of time. Using CRRES data, [Meredith et al., 2006] demonstrated that the flux values elevated by the substorm or storm processes decay to their pre-storm values gradually, and that the flux can fall by 2 orders of magnitude over a period of approximately 20 days. Therefore, an indicator of the past events is needed in order to correctly reproduce the dynamics of the flux decay. In the MERLIN model we do not use any previous values of flux and hence it is the history of velocity that the model uses as such an indicator. We note that AE and Kp indices exhibit very close feature importance values, because being highly correlated with one another, they produce equal reduction in variance when making splits.

Figure 5 shows the influence of the solar wind history on the model performance on the example of 425 keV electrons. The MSE of the model with no solar wind history employed is shown as black dots. We further add the history of solar wind velocity of up to 14 days with gradually increasing time steps. The MSE gradually decreases as more history is included, both on training and validation data. On the other hand, including also the time history of solar wind density decreases the training but not the validation error. In Figure 5 it can be seen that while the training MSE is lower at every time step when density is included, the validation MSE does not change as compared to using velocity history alone and remains within the error bar, indicating overfitting. Therefore, we only use the history of velocity as inputs.

4.2 Results of 10-fold cross validation

Multiple metrics can be used to evaluate the model accuracy [for details see Morley et al., 2018; Liemohn et al., 2018]. The LightGBM library offers a variety of metrics implemented for model analysis. The default metrics is the mean squared error (MSE), which is computed at each training iteration. It should be noted that the electron flux can exhibit strong depletions, up to several orders of magnitude, over short times [Ganushkina et al., 2019]. The mean squared error is susceptible to outliers , and therefore we also evaluate the median of the squared error. In Table 4.2 it can be seen that both for training and validation sets, median of the squared error is ~3 times lower than the MSE. This means that while some of the rapid depletions/enhancements are not completely reproduced, the value of the median squared error of 0.05 shows that the model predictions are very close to the observed data.

Table 2. Metrics evaluated on the train and validation sets during 10-fold CV, and on the test data for 425 keV electrons. The standard error for the 10-fold CV are shown in brackets.

Metrics	Train	Validation	Test
Mean SE	**0.16** ($\pm$0.011)	**0.24** ($\pm$0.026)	**0.23**
Median SE	**0.05** ($\pm$0.004)	**0.08** ($\pm$0.009)	**0.07**
MAE	**0.21** ($\pm$0.009)	**0.26** ($\pm$0.014)	**0.26**
NRMSD	**0.05** ($\pm$0.006)	**0.08** ($\pm$0.009)	**0.08**
PE	**0.76** ($\pm$0.015)	**0.59** ($\pm$0.018)	**0.55**
Spearman ρ	**0.88** ($\pm$0.006)	**0.81** ($\pm$0.012)	**0.79**
Ratio	**1**	**1**	**1**

Another useful metric, implemented in the LightGBM library, is the median absolute error, denoted as MAE. Evaluating the median error allows us to pay less attention to the outliers but gives a good proxy of model performance otherwise. In our case, MAE values are close for the train (0.21) and validation sets (0.26). These values of MAE mean that in general, the predicted values differ from observations by a factor of $\sim$1.5, which is considered acceptable for radiation belt modeling. The scale-dependent metrics can be difficult to interpret, since the model predicts base 10 logarithms of flux and also because the level of flux is generally different for different energies. We evaluate the normalized root mean squared deviation (NRMSD), defined in [Denton et al., 2019, eq. 1]. The zero value of NRMSD corresponds to the perfect prediction, and the values of < 1 generally indicate a good match between the model and data [Denton et al., 2019]. From Table 4.2 one can see that the values of NRMSD are close to 0 (0.05 for train and 0.08 for the validation set), indicating that the model reproduces the observations very well.

The metrics mentioned above mainly quantify how far the predictions deviate from observations at stationary points. In order to evaluate how well the model reproduces the dynamical behaviour of the electron flux, we employ the following metrics. First, we evaluate the correlation between the predictions and observations. The standard Pearson linear correlation coefficient is susceptible to outliers, and therefore we use the Spearman rank correlation coefficient (ρ). From Table 4.2 it can be seen that the correlations are rather high both for train and validation sets, based on the 10-fold CV. The average correlation on the train set is 0.88 and is slightly higher than that on the validation (0.79), which is as expected since the model was fitted on the train data. The high values of the correlation coefficient show that the model captures the behavior of flux at energies 120-600 keV, which are known to be very dynamic and can exhibit drastic changes on the order of several minutes.

Another popular metric widely used in machine learning is the r^2-score. This indicator is used to quantify the fraction of variance explained by the model [e.g., Morley et al., 2018]. The average r^2 for the train data is 0.76, and 0.59 for the validation set. While these values appear low, we conduct a more detailed analysis of the r^2 on the synthetic data, shown in the Supplement (Figure S1). We first generate a harmonic function - a sinusoid, for 20 periods of 2π each, and then add random noise of maximum magnitude equal 0.2. The electron flux are known to exhibit rapid depletions of up to several orders of magnitude, and to account for this we add 80 dropouts where we subtract 2 units from the synthetic signal. We compute the values of the r^2 for the case with and without outliers. The initial sinusoid signal compared to the data with no outliers yields prediction efficiency of $\sim$0.9, while the data in presence of outliers give a lower r^2 value of $\sim$0.6, which is approximately equal to the value we have for MERLIN on the validation set. This means that the model can adequately reproduce the behaviour of flux but miss some of the dropout magnitudes, which results in the lower prediction efficiency. This will be further discussed while analyzing the performance on the test data.

Figure 6 demonstrates the model performance, averaged over the 10-fold CV, for individual energy channels. Fig. 6a shows the median absolute error. It can be seen that the best performance is achieved for 425 keV electrons, although we note that the MAE for other channels is only slightly different. We find that the accuracy improves with higher energies (MAE for 120 keV on the validation set is 0.33 compared with $\sim$0.26 for 425 keV). The accuracy then slightly decreases from 425 keV to 600 keV electrons. Same can be seen in terms of the correlation (Figure 6b) - the averaged correlation on the validation sets is $\sim$0.76 for 120 keV and $\sim$0.82 for 425 keV. The ρ value then slightly decreases to 0.8 for 600 keV. A slightly lower $\rho \sim 0.75$ for 120 keV electrons likely comes from the more dynamic nature of this low-energy population, with processes on timescales of shorter than 1 minute having a non-negligible effect [e.g., Ganushkina et al., 2019]. It should be noted that the values of MAE and the correlation for each individual channel show that the MERLIN model well predicts both amplitudes and flux dynamics throughout the considered energy range.

4.3 Performance on test data

The train set is used to learn the model, and the model is constructed so as to minimize the error on the validation set. On each iteration, the MSE and MAE values should decrease for the train set as the gradient boosting model adjusts the residuals to fit the train set better, and then the updated model is evaluated on the validation set. If the validation error reduces, the training continues to a new iteration (i.e., grows a new GBDT). The model never learns from the validation set, however, it is still being used on every training iteration. We therefore need another totally independent set of values to check that the model generalizes well onto the unseen data. It is of note that this test set can only be used once to evaluate the performance, after the model training has been completed, and no further changes to the model can be made. More details on the train-validation-test splitting of the data can be found in Section 3 and in Figure 3.

The values of different metrics discussed above are given in Table 4.2. We find that the values on the test set are close to those on the average of the validation sets. MAE values are identical and equal to 0.26 on the two sets. MERLIN yields NRMSD values of 0.08 for both test and validation data, which show that the model performs well on both sets as the NRMSDs are close to zero. The values of the Spearman correlation coefficient are very close for the test data (0.79) and the average validation (0.81). The mean and median squared errors also yield almost identical values for validation and test sets, and their differences are within the error bar. In general, all of the metrics exhibit values which are sufficiently close on test and validation sets and are slightly lower than on training data. This means that the model successfully learnt the underlying relationships between the input parameters and the resulting electron flux on the training data, yields good accuracy on the validation intervals, and generalizes well onto the unseen data.

Figure 7a shows the GPS electron flux from all 21 spacecraft for March 2016-December 2017 for 300 keV. Figure 7b gives the flux values from the MERLIN model, and the difference between the predicted and observed flux is shown in Figure 7c. Figure 7e demonstrates the solar wind velocity for the test set. It should be noted that several solar wind velocity enhancements happened during the test interval with v_{sw} rising up to >700 km/s. These events generally correspond to increases in electron flux, but it can be seen that the after velocity drops to its quiet time values the flux remains elevated for longer periods of time. For instance, v_{sw} increased up to 700 km/s at the beginning of September 2016 (detailed illustration is in the Figure S2) and caused a rapid increase in electron flux by over an order of magnitude. The velocity then started decreasing and within 1 week dropped to 500 km/s while the flux remained at the elevated levels. Within a few days, velocity continued decreasing until it reached <300 km/s but the electron flux had a much longer decay, and even after ~1 week of very low velocities the flux did not fall to the pre-event values.

In general, the model adequately reproduces all of the major flux enhancement events and also reproduces well the flux decay due to consideration of the velocity history. We note that for L-shells lower than 5, the flux enhancements are sometimes followed by periods of the intense data variance (Fig. 7a). The most likely explanation lies in the fact that the GPS electron flux data are a derived quantity. As such, the fluxes are computed using a forward model combining 3 relativistic Maxwellians (in energy) and a Gaussian (in log of momentum) [see Morley et al., 2016, for details]. The electron energy spectra inside the plasmapause can have local peaks that may not be well-captured by the forward model [see Section 3.2 of Morley et al., 2016], and intense plasmaspheric hiss can generate a reverse spectrum in the energy range of hundreds of keV [Zhao et al., 2019]. These spectra cannot be properly represented by the forward model used for calculating the GPS flux and can lead to ill-determined fits, giving the observed variance in flux at lower energies. This explanation is also supported by the fact that these events are not visible in the higher GPS energy channels. As expected, these periods are not reproduced by the MERLIN model.

The Spearman correlation between the observed and predicted flux is approximately 0.77. It is, however, important that the model reproduces not only the flux along the GPS orbits but is also capable of giving reasonable prediction at fixed L-values. Figure 7d shows the GPS observations and the MERLIN output for L of 5.2. The correlation between the two is 0.75 and the mean squared error is 0.06. One can therefore conclude that the model generalizes well on the unseen data, both at a fixed L and along the orbit. The same conclusion can be drawn for the 600 keV population, demonstrated in Figure 8.

Figure 9a-e shows the occurrence density plots of the observed versus predicted flux for all 5 energies considered. Solid white lines show the one-to-one relationship between observations and predictions, and the dashed likes represent the flux deviating by a factor of 5. In general, the occurrence maxima follow the trend and most of the points are within the factor of 5 from the trend line. Figure 9f provides an example of the histogram of model residuals for the 210 keV population. From the figure it is evident that the model has very low bias of $\sim$ 0.05, depicted by a green line. Furthermore, the errors are normally distributed and the 10-th and 90-th percentiles are $\sim \pm$-0.52 and 0.53, respectively, meaning that 80% of the model residuals are within a range of $\pm$0.5. Therefore, we can conclude that MERLIN predicts the electron flux in 120-600 keV energy range well, has low bias and captures all of the general trends represented in the data.

5 Conclusions

A new data-driven model of the electron flux in the outer radiation belt is presented. The model uses satellite position and a combination of geomagnetic indices and solar wind parameters (with time history of velocity) in order to predict the flux values at energies 120-600 keV. The model has been trained and validated on more than 15 years of GPS electron flux data, and tested on $>$1.5 years of observations. The 10-fold cross validation shows that the MERLIN model predicts the MEO radiation environment well, capturing both the dynamics and amplitudes of electron flux. The results of the 10-fold cross validation agree well with the evaluation on the test data meaning that the model is able to generalize well onto the unseen events. Predicted values of flux exhibit high correlation with the observations ($\sim$0.8) and low values of error.

The MERLIN model can have wide Space Weather applications. It can be used by the scientific community to analyze specific events as well as to reconstruct the long-term radiation belt dynamics at the 120-600 keV energy range, for which there is generally a lack of models. Furthermore, it is of use for satellite operators for the nowcast of the MEO environment and can provide information for the surface charging analysis. We note that the model was trained on the data from the GPS constellation, which follows an inclined orbit. Therefore, at higher magnetic latitudes the satellites sample only a part of the equatorial pitch angle distribution. However, using the appropriate pitch angle models it is possible to reconstruct the values of equatorial flux from MERLIN predictions [e.g., using a methodology of Allison et al., 2018]. Further directions for the present study include, first, a more refined analysis of the feature importances using the appropriate permutation and Shapley values methods and the corresponding feature selection. Second, as GPS continues to probe the outer radiation belt, more data can be incorporated into the model.

Acknowledgments
The resulting model files in `.txt` format can are available in open access at: https://doi.org/10.5281/zenodo.3783757. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870452 (PAGER). The data of solar wind parameters and geomagnetic indices were taken from OMNIWeb database: [omniweb.gsfc.nasa.gov]. Contributions by S.K. Morley and M.R. Carver were performed under the auspices of the US Department of Energy and supported by the Laboratory Directed Research and Development (LDRD) program, award 20190262ER. We acknowledge the CXD team at Los Alamos National Laboratory (LANL), operating the CXD detectors aboard the GPS constellation. The GPS particle data can be obtained in open access at [https://www.ngdc.noaa.gov/stp/space-weather/satellite-data/satellite-systems/gps/data/]. The analysis in this study was conducted

using gradient boosting with regression trees, as implemented in the LightGBM Python library: [lightgbm.readthedocs.io].

References

Allison, H. J., Horne, R. B., Glauert, S. A., and Del Zanna, G. (2018). Determination of the equatorial electron differential flux from observations at low earth orbit. *Journal of Geophysical Research: Space Physics*, 123(11):9574–9596.

Aseev, N., Shprits, Y., Drozdov, A., and Kellerman, A. (2016). Numerical applications of the advective-diffusive codes for the inner magnetosphere. *Space Weather*, 14(11):993–1010.

Auret, L. and Aldrich, C. (2011). Empirical comparison of tree ensemble variable importance measures. *Chemometrics and Intelligent Laboratory Systems*, 105(2):157–170.

Baker, D., McPherron, R., Cayton, T., and Klebesadel, R. (1990). Linear prediction filter analysis of relativistic electron properties at 6.6 re. *Journal of Geophysical Research: Space Physics*, 95(A9):15133–15140.

Bala, R., Reiff, P., and Landivar, J. (2009). Real-time prediction of magnetospheric activity using the boyle index. *Space Weather*, 7(4):1–17.

Balikhin, M. A., Boynton, R. J., Walker, S. N., Borovsky, J. E., Billings, S. A., and Wei, H.-L. (2011). Using the narmax approach to model the evolution of energetic electrons fluxes at geostationary orbit. *Geophysical Research Letters*, 38(18).

Bergstra, J., Yamins, D., and Cox, D. D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures.

Bortnik, J., Chu, X., Ma, Q., Li, W., Zhang, X., Thorne, R. M., Angelopoulos, V., Denton, R. E., Kletzing, C. A., Hospodarsky, G. B., et al. (2018). Artificial neural networks for determining magnetospheric conditions. In *Machine learning techniques for space weather*, pages 279–300. Elsevier.

Boyd, A., Spence, H., Huang, C.-L., Reeves, G., Baker, D., Turner, D., Claudepierre, S., Fennell, J., Blake, J., and Shprits, Y. (2016). Statistical properties of the radiation belt seed population. *Journal of Geophysical Research: Space Physics*, 121(8):7636–7646.

Boyd, A. J., Spence, H. E., Claudepierre, S., Fennell, J. F., Blake, J., Baker, D., Reeves, G., and Turner, D. (2014). Quantifying the radiation belt seed population in the 17 march 2013 electron acceleration event. *Geophysical Research Letters*, 41(7):2275–2281.

Boynton, R., Balikhin, M., Sibeck, D., Walker, S., Billings, S., and Ganushkina, N. (2016). Electron flux models for different energies at geostationary orbit. *Space Weather*, 14(10):846–860.

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, 17(8):1166–1207.

Carver, M. R., Sullivan, J. P., Morley, S. K., and Rodriguez, J. V. (2018). Cross calibration of the gps constellation cxd proton data with goes eps. *Space Weather*, 16(3):273–288.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, Y., Reeves, G. D., Fu, X., and Henderson, M. (2019). Premeve: New predictive model for megaelectron-volt electrons inside earth's outer radiation belt. *Space Weather*, 17(3):438–454.

Colak, T. and Qahwaji, R. (2009). Automated solar activity prediction: a hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. *Space Weather*, 7(6).

de Lima, R. P., Chen, Y., and Lin, Y. (2020). Forecasting megaelectron-volt electrons inside earth's outer radiation belt: Premeve 2.0 based on supervised machine learning algorithms. *Space Weather*, 18(2).

Denton, M., Henderson, M. G., Jordanova, V. K., Thomsen, M. F., Borovsky, J. E., Woodroffe, J., Hartley, D., and Pitchford, D. (2016). An improved empirical model of electron and ion fluxes at geosynchronous orbit based on upstream solar wind conditions. *Space Weather*, 14(7):511–523.

Denton, M., Taylor, M., Rodriguez, J., and Henderson, M. (2019). Extension of an empirical electron flux model from 6 to 20 earth radii using cluster/rapid observations. *Space Weather*, 17(5):778–792.

Denton, M., Thomsen, M., Jordanova, V., Henderson, M., Borovsky, J., Denton, J., Pitchford, D., and Hartley, D. (2015). An empirical model of electron and ion fluxes derived from observations at geosynchronous orbit. *Space Weather*, 13(4):233–249.

Ding, C., Wang, D., Ma, X., and Li, H. (2016). Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustainability*, 8(11):1100.

Druzhkov, P., Erukhimov, V., Zolotykh, N. Y., Kozinov, E., Kustikova, V., Meerov, I., and Polovinkin, A. (2011). New object detection features in the opencv library. *Pattern Recognition and Image Analysis*, 21(3):384.

Fennell, J. F., Koons, H. C., Chen, M. W., and Blake, J. (2000). Internal charging: A preliminary environmental specification for satellites. *IEEE transactions on plasma science*, 28(6):2029–2036.

Fok, M.-C. and Moore, T. (1997). Ring current modeling in a realistic magnetic field configuration. *Geophysical research letters*, 24(14):1775–1778.

Fok, M.-C., Wolf, R., Spiro, R., and Moore, T. (2001). Comprehensive computational model of earth's ring current. *Journal of Geophysical Research: Space Physics*, 106(A5):8417–8424.

Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Ganushkina, N. Y., Dandouras, I., Shprits, Y. Y., and Cao, J. (2011). Locations of boundaries of outer and inner radiation belts as observed by cluster and double star. *Journal of Geophysical Research: Space Physics*, 116(A9).

Ganushkina, N. Y., Sillanpää, I., Welling, D., Haiducek, J., Liemohn, M., Dubyagin, S., and Rodriguez, J. (2019). Validation of inner magnetosphere particle transport and acceleration model (imptam) with long-term goes maged measurements of kev electron fluxes at geostationary orbit. *Space Weather*, 17(5):687–708.

Garrett, H. B. (1981). The charging of spacecraft surfaces. *Reviews of Geophysics*, 19(4):577–616.

Ginet, G., O'Brien, T., Huston, S., Johnston, W., Guild, T., Friedel, R., Lindstrom, C., Roth, C., Whelan, P., Quinn, R., et al. (2013). Ae9, ap9 and spm: New models for specifying the trapped energetic particle and space plasma environment. In *The Van Allen Probes Mission*, pages 579–615. Springer.

Glauert, S. A., Horne, R. B., and Meredith, N. P. (2014). Three-dimensional electron radiation belt simulations using the bas radiation belt model with new diffusion models for chorus, plasmaspheric hiss, and lightning-generated whistlers. *Journal of Geophysical Research: Space Physics*, 119(1):268–289.

Glauert, S. A., Horne, R. B., and Meredith, N. P. (2018). A 30-year simulation of the outer electron radiation belt. *Space Weather*, 16(10):1498–1522.

Horne, R. B., Glauert, S. A., Meredith, N. P., Koskinen, H., Vainio, R., Afanasiev, A., Ganushkina, N. Y., Amariutei, O. A., Boscher, D., Sicard, A., et al. (2013). Forecasting the earth's radiation belts and modelling solar energetic particle events: Recent results from spacecast. *Journal of Space Weather and Space Climate*, 3:A20.

Horne, R. B. and Pitchford, D. (2015). Space weather concerns for all-electric propulsion satellites. *Space Weather*, 13(8):430–433.

Jaynes, A. N., Baker, D. N., Singer, H. J., Rodriguez, J. V., Loto'aniu, T., Ali, A., Elkington, S. R., Li, X., Kanekal, S. G., Claudepierre, S. G., 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.

Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., and Rehman, M. U. (2019). A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting. *IEEE Access*, 7:28309–28318.

Kavanagh, A. J., Cobbett, N., and Kirsch, P. (2018). Radiation belt slot region filling events: Sustained energetic precipitation into the mesosphere. *Journal of Geophysical Research: Space Physics*, 123(9):7999–8020.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, pages 3146–3154.

Kellerman, A. and Shprits, Y. (2012). On the influence of solar wind conditions on the outer-electron radiation belt. *Journal of Geophysical Research: Space Physics*, 117(A5).

Kitamura, K., Nakamura, Y., Tokumitsu, M., Ishida, Y., and Watari, S. (2011). Prediction of the electron flux environment in geosynchronous orbit using a neural network technique. *Artificial Life and Robotics*, 16(3):389–392.

Lanzerotti, L., LaFleur, K., Maclennan, C., and Maurer, D. (1998). Geosynchronous spacecraft charging in january 1997. *Geophysical research letters*, 25(15):2967–2970.

Li, X. (2004). Variations of 0.7–6.0 mev electrons at geosynchronous orbit as a function of solar wind. *Space Weather*, 2(3).

Li, X., Baker, D., O'Brien, T., Xie, L., and Zong, Q. (2006). Correlation between the inner edge of outer radiation belt electrons and the innermost plasmapause location. *Geophysical research letters*, 33(14).

Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., Tobiska, W. K., Wintoft, P., Ganushkina, N. Y., Welling, D. T., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, 16(12):2079–2102.

Ling, A., Ginet, G., Hilmer, R., and Perry, K. (2010). A neural network-based geosynchronous relativistic electron flux forecasting model. *Space Weather*, 8(9):1–14.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pages 4765–4774.

Lyons, L. R. and Thorne, R. M. (1973). Equilibrium structure of radiation belt electrons. *Journal of Geophysical Research (1896-1977)*, 78(13):2142–2149.

Lyons, L. R., Thorne, R. M., and Kennel, C. F. (1972). Pitch-angle diffusion of radiation belt electrons within the plasmasphere. *Journal of Geophysical Research (1896-1977)*, 77(19):3455–3474.

Maloney, K. O., Schmid, M., and Weller, D. E. (2012). Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. *Methods in Ecology and Evolution*, 3(1):116–128.

Meredith, N. P., Horne, R. B., Glauert, S. A., Thorne, R. M., Summers, D., Albert, J. M., and Anderson, R. R. (2006). Energetic outer zone electron loss timescales during low geomagnetic activity. *Journal of Geophysical Research: Space Physics*, 111(A5).

Morley, S. K., Brito, T. V., and Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, 16(1):69–88.

Morley, S. K., Sullivan, J. P., Henderson, M. G., Blake, J. B., and Baker, D. N. (2016). The global positioning system constellation as a space weather monitor: Comparison of electron measurements with van allen probes data. *Space Weather*, 14(2):76–92.

O'Brien, T. and Moldwin, M. (2003). Empirical plasmapause models from magnetic indices. *Geophysical Research Letters*, 30(4).

Olifer, L., Mann, I. R., Morley, S. K., Ozeke, L. G., and Choi, D. (2018). On the role of last closed drift shell dynamics in driving fast losses and van allen radiation belt extinction. *Journal of Geophysical Research: Space Physics*, 123(5):3692–3703.

Orlova, K., Shprits, Y., and Spasojevic, M. (2016). New global loss model of energetic and relativistic electrons based on van allen probes measurements. *Journal of Geophysical Research: Space Physics*, 121(2):1308–1314.

Osthus, D., Caragea, P., Higdon, D., Morley, S., Reeves, G., and Weaver, B. (2014). Dynamic linear models for forecasting of radiation belt electrons and limitations on physical interpretation of predictive models. *Space Weather*, 12(6):426–446.

Paulikas, G. and Blake, J. (1979). Effects of the solar wind on magnetospheric dynamics: Energetic electrons at the synchronous orbit. *Quantitative Modeling of Magnetospheric Processes, Geophys. Monogr. Ser*, 21:180–202.

Pinto, V., Zhang, X.-J., Mourenas, D., Bortnik, J., Artemyev, A., Lyons, L., and Moya, P. (2020). On the confinement of ultrarelativistic electron remnant belts to low l-shells. *Journal of Geophysical Research: Space Physics*.

Reeves, G., McAdams, K., Friedel, R., and O'Brien, T. (2003). Acceleration and loss of relativistic electrons during geomagnetic storms. *Geophysical Research Letters*, 30(10).

Reeves, G., Morley, S., and Cunningham, G. (2013). Long-term variations in solar wind velocity and radiation belt electrons. *Journal of Geophysical Research: Space Physics*, 118(3):1040–1048.

Reeves, G. D., Chen, Y., Cunningham, G., Friedel, R., Henderson, M. G., Jordanova, V., Koller, J., Morley, S., Thomsen, M., and Zaharia, S. (2012). Dynamic radiation environment assimilation model: Dream. *Space Weather*, 10(3):1–25.

Reeves, G. D., Friedel, R. H., Larsen, B. A., Skoug, R. M., Funsten, H. O., Claudepierre, S. G., Fennell, J. F., Turner, D. L., Denton, M. H., Spence, H. E., et al. (2016). Energy-dependent dynamics of kev to mev electrons in the inner zone, outer zone, and slot regions. *Journal of Geophysical Research: Space Physics*, 121(1):397–412.

Reeves, G. D., Morley, S. K., Friedel, R. H., Henderson, M. G., Cayton, T. E., Cunningham, G., Blake, J. B., Christensen, R. A., and Thomsen, D. (2011). On the relationship between relativistic electron flux and solar wind velocity: Paulikas and blake revisited. *Journal of Geophysical Research: Space Physics*, 116(A2).

Ren, C. X., Dorostkar, O., Rouet-Leduc, B., Hulbert, C., Strebel, D., Guyer, R. A., Johnson, P. A., and Carmeliet, J. (2019). Machine learning reveals the state of intermittent frictional dynamics in a sheared granular fault. *Geophysical Research Letters*, 46(13):7395–7403.

Ripoll, J.-F., Reeves, G. D., Cunningham, G. S., Loridan, V., Denton, M., Santolík, O., Kurth, W., Kletzing, C., Turner, D., Henderson, M., et al. (2016). Reproducing the observed energy-dependent structure of earth's electron radiation belts during storm recovery with an event-specific diffusion model. *Geophysical Research Letters*, 43(11):5616–5625.

Sakaguchi, K., Nagatsuma, T., Reeves, G. D., and Spence, H. E. (2015). Prediction of mev electron fluxes throughout the outer radiation belt using multivariate autoregressive models. *Space Weather*, 13(12):853–867.

Shprits, Y. Y., Kellerman, A. C., Drozdov, A. Y., Spence, H. E., Reeves, G. D., and Baker, D. N. (2015). Combined convective and diffusive simulations: Verb-4d comparison with 17 march 2013 van allen probes observations. *Geophysical Research Letters*, 42(22):9600–9608.

Shprits, Y. Y., Vasile, R., and Zhelavskaya, I. S. (2019). Nowcasting and predicting the kp index using historical values and real-time observations. *Space Weather*, 17(8):1219–1229.

Simms, L. E., Engebretson, M. J., Pilipenko, V., Reeves, G. D., and Clilverd, M. (2016). Empirical predictive models of daily relativistic electron flux at geostationary orbit: Multiple regression analysis. *Journal of Geophysical Research: Space Physics*, 121(4):3181–3197.

Simms, L. E., Pilipenko, V., Engebretson, M. J., Reeves, G. D., Smith, A., and Clilverd, M. (2014). Prediction of relativistic electron flux at geostationary orbit following storms: Multiple regression analysis. *Journal of Geophysical Research: Space Physics*, 119(9):7297–7318.

Smirnov, A., Kronberg, E. A., Latallerie, F., Daly, P. W., Aseev, N., Shprits, Y., Kellerman, A., Kasahara, S., Turner, D., and Taylor, M. (2019). Electron intensity measurements by the cluster/rapid/ies instrument in earth's radiation belts and ring current. *Space Weather*, 17(4):553–566.

Smirnov, A. G., Kronberg, E. A., Daly, P. W., Aseev, N. A., Shprits, Y. Y., and Kellerman, A. C. (2020). Adiabatic invariants calculations for cluster mission: A long-term product for radiation belts studies. *Journal of Geophysical Research: Space Physics*, page e2019JA027576.

Subbotin, D. and Shprits, Y. (2009). Three-dimensional modeling of the radiation belts using the versatile electron radiation belt (verb) code. *Space Weather*, 7(10).

Summers, D., Ma, C., and Mukai, T. (2004). Competition between acceleration and loss mechanisms of relativistic electrons during geomagnetic storms. *Journal of Geophysical Research: Space Physics*, 109(A4).

Tsutai, A., Mitsui, C., and Nagai, T. (1999). Prediction of a geosynchronous electron environment with in situ magnetic field measurements. *Earth, planets and space*, 51(3):219–223.

Tu, W., Cunningham, G., Chen, Y., Henderson, M., Camporeale, E., and Reeves, G. (2013). Modeling radiation belt electron dynamics during gem challenge intervals with the dream3d diffusion model. *Journal of Geophysical Research: Space Physics*, 118(10):6197–6211.

Tu, W., Cunningham, G., Chen, Y., Morley, S., Reeves, G., Blake, J., Baker, D., and Spence, H. (2014). Event-specific chorus wave and electron seed population models in dream3d using the van allen probes. *Geophysical Research Letters*, 41(5):1359–1366.

Vampola, A. (1997). Outer zone energetic electron environment update. In *Conference on the High Energy Radiation Background in Space. Workshop Record*, pages 128–136. IEEE.

Van Allen, J. A. and Frank, L. A. (1959). Radiation around the earth to a radial distance of 107,400 km. *Nature*, 183(4659):430–434.

Wing, S., Johnson, J. R., Camporeale, E., and Reeves, G. D. (2016). Information theoretical approach to discovering solar wind drivers of the outer radiation belt. *Journal of Geophysical Research: Space Physics*, 121(10):9378–9399.

Zhang, J., Mucs, D., Norinder, U., and Svensson, F. (2019). Lightgbm: An effective and scalable algorithm for prediction of chemical toxicity–application to the tox21 and mutagenicity data sets. *Journal of Chemical Information and Modeling*, 59(10):4150–4158.

Zhao, H., Ni, B., Li, X., Baker, D., Johnston, W., Zhang, W., Xiang, Z., Gu, X., Jaynes, A., Kanekal, S., et al. (2019). Plasmaspheric hiss waves generate a reversed energy spectrum of radiation belt electrons. *Nature Physics*, 15(4):367–372.

Zhelavskaya, I. S., Shprits, Y. Y., and Spasojević, M. (2017). Empirical modeling of the plasmasphere dynamics using neural networks. *Journal of Geophysical Research: Space Physics*, 122(11):11–227.

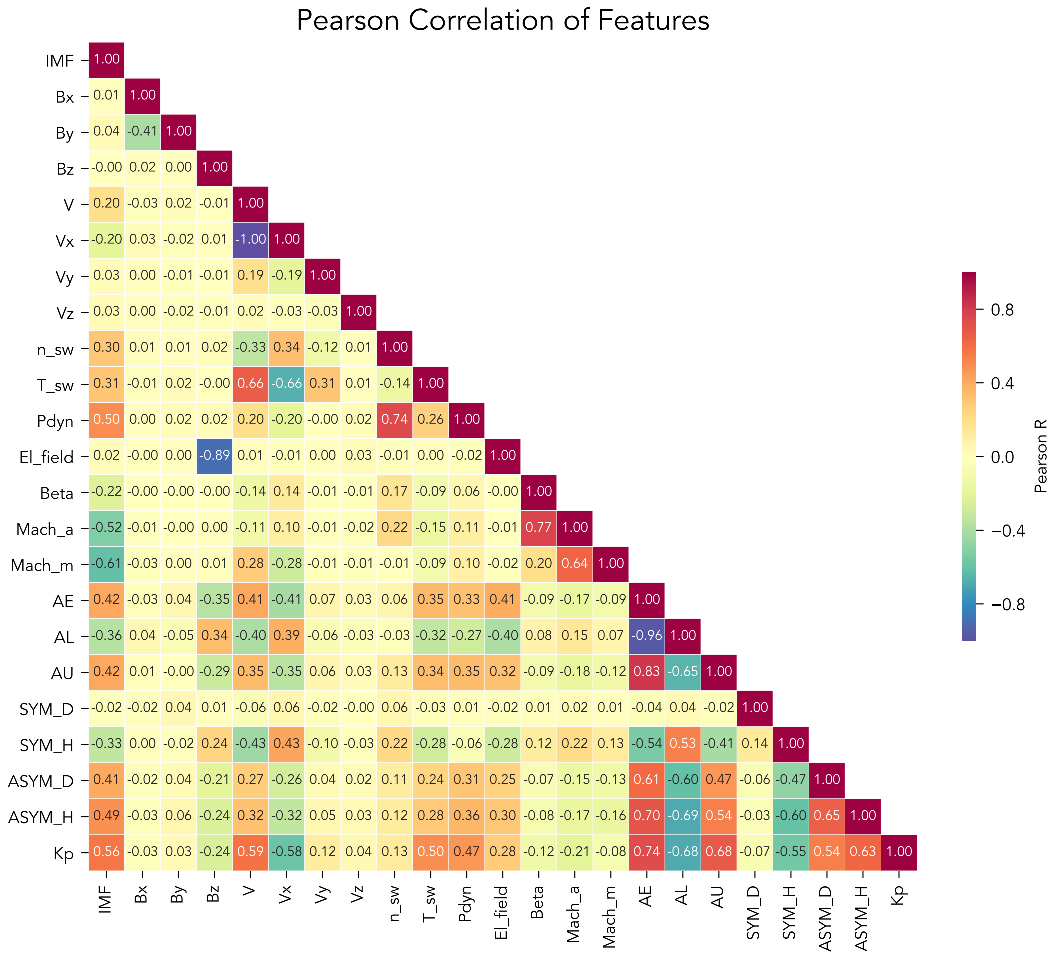


Figure 1. Correlation between different solar wind and geomagnetic parameters based on 2001-2016 OMNIWeb data.

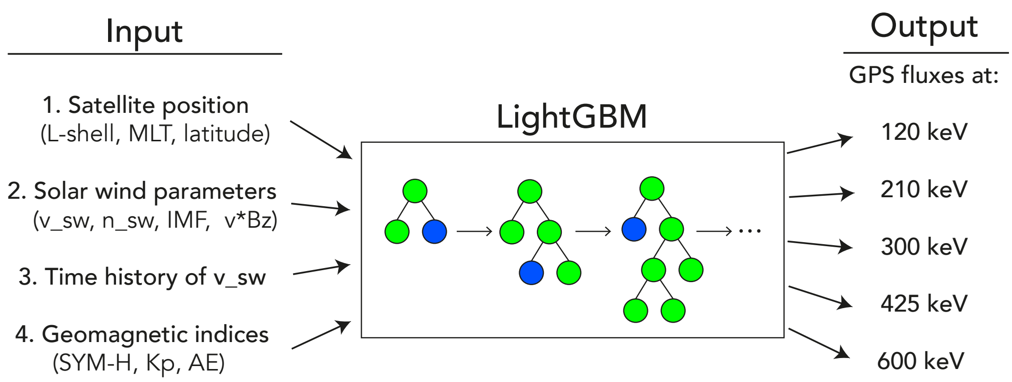


Figure 2. Schematic representation of the model workflow. The input parameters include the satellite position in L-MLT-latitude frame, solar wind parameters with history of velocity, and geomagnetic indices. The inputs are supplied to the LightGBM algorithm in order to return the flux values at energies 120-600 keV.

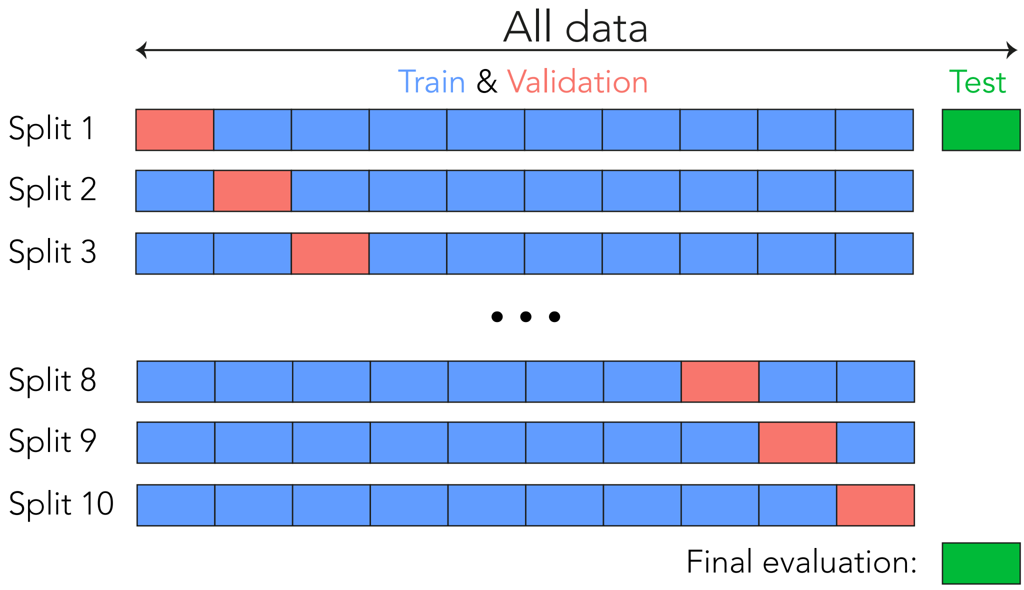


Figure 3. Schematics of the K-fold cross validation (CV). The last 1.5 years of data are reserved as the test set, never to be used during training and validation. The rest of the data are then divided into K equal parts (here, K=10) and at every split the model is trained on 9 parts and validated on 1 part. The training process is repeated 10 times, each time withholding a different set for validation. Thus, the model uses all observations for training and validation, and each of the data points is used for validation only once. The final evaluation is performed on the test set.

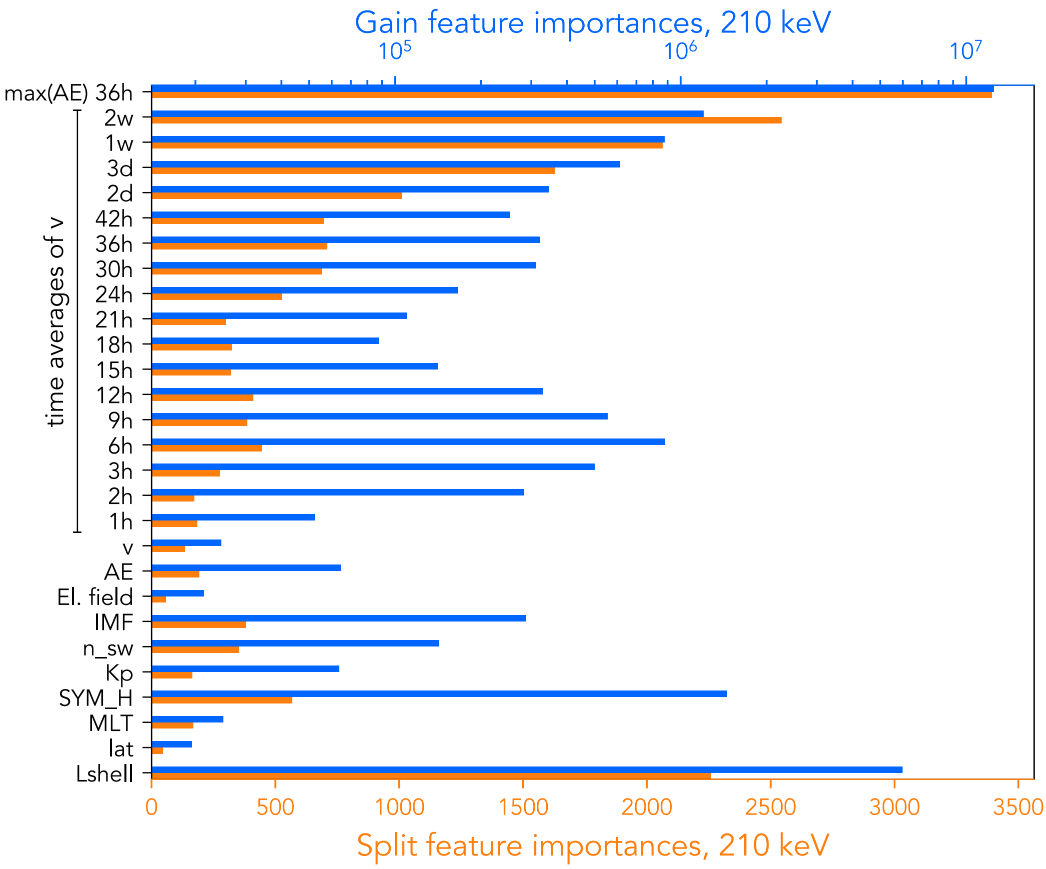


Figure 4. Feature importances estimated using the intrinsic LightGBM gain and split methods for 210 keV electron flux.

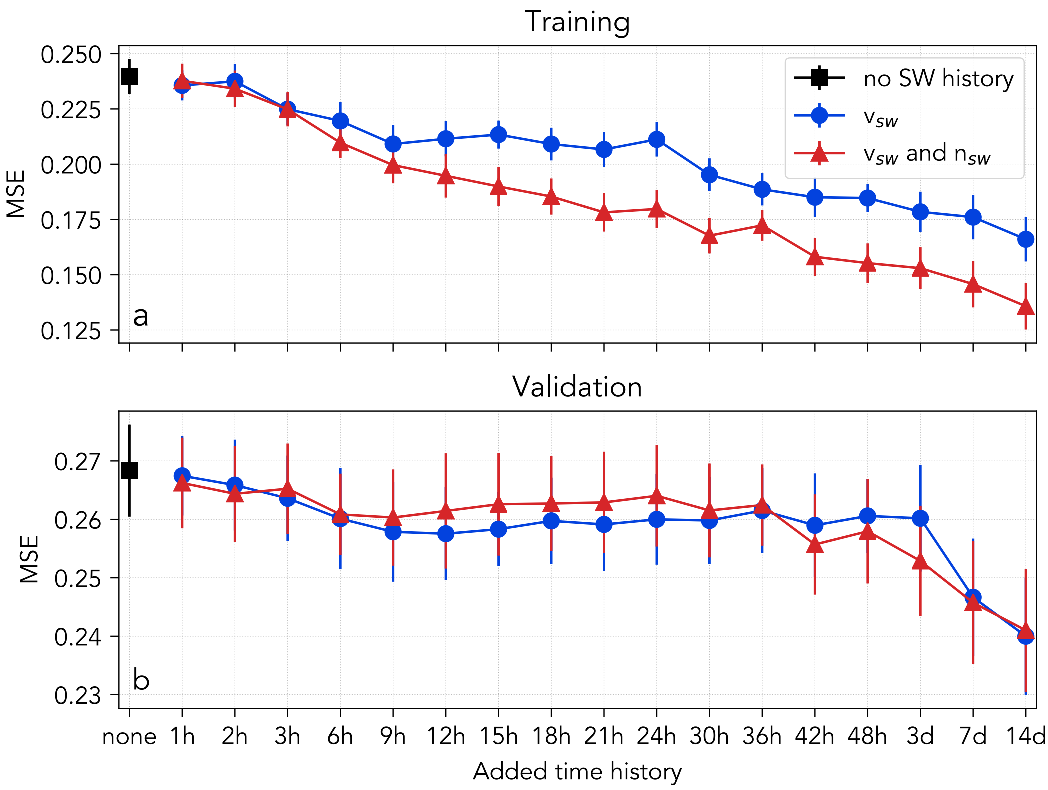


Figure 5. MSE on the training (a) and validation (b) data for 425 keV electron flux depending on solar wind history. The black dot represents the model with no solar wind history employed. Averages of solar wind velocity (blue curve) are added as model inputs and reduce both train and validation error. Including also the history number density reduces the train but not the validation error and leads to overfitting. The vertical dashes represent the errors of the 10-fold CV.

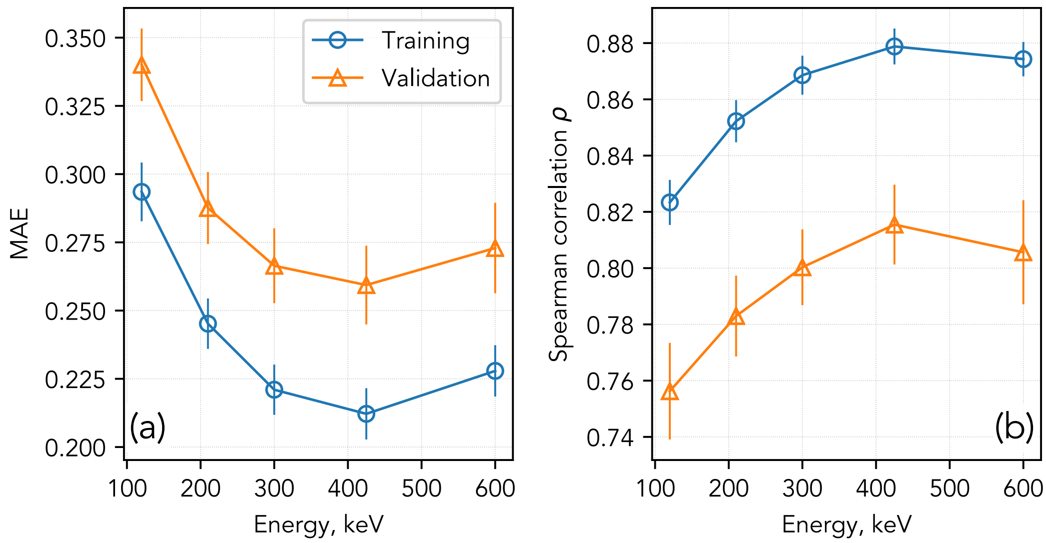


Figure 6. Model performance on training and validation sets averaged over the 10-fold cross validation. Median absolute error values are shown in (a), and Spearman correlation coefficients are given in (b).

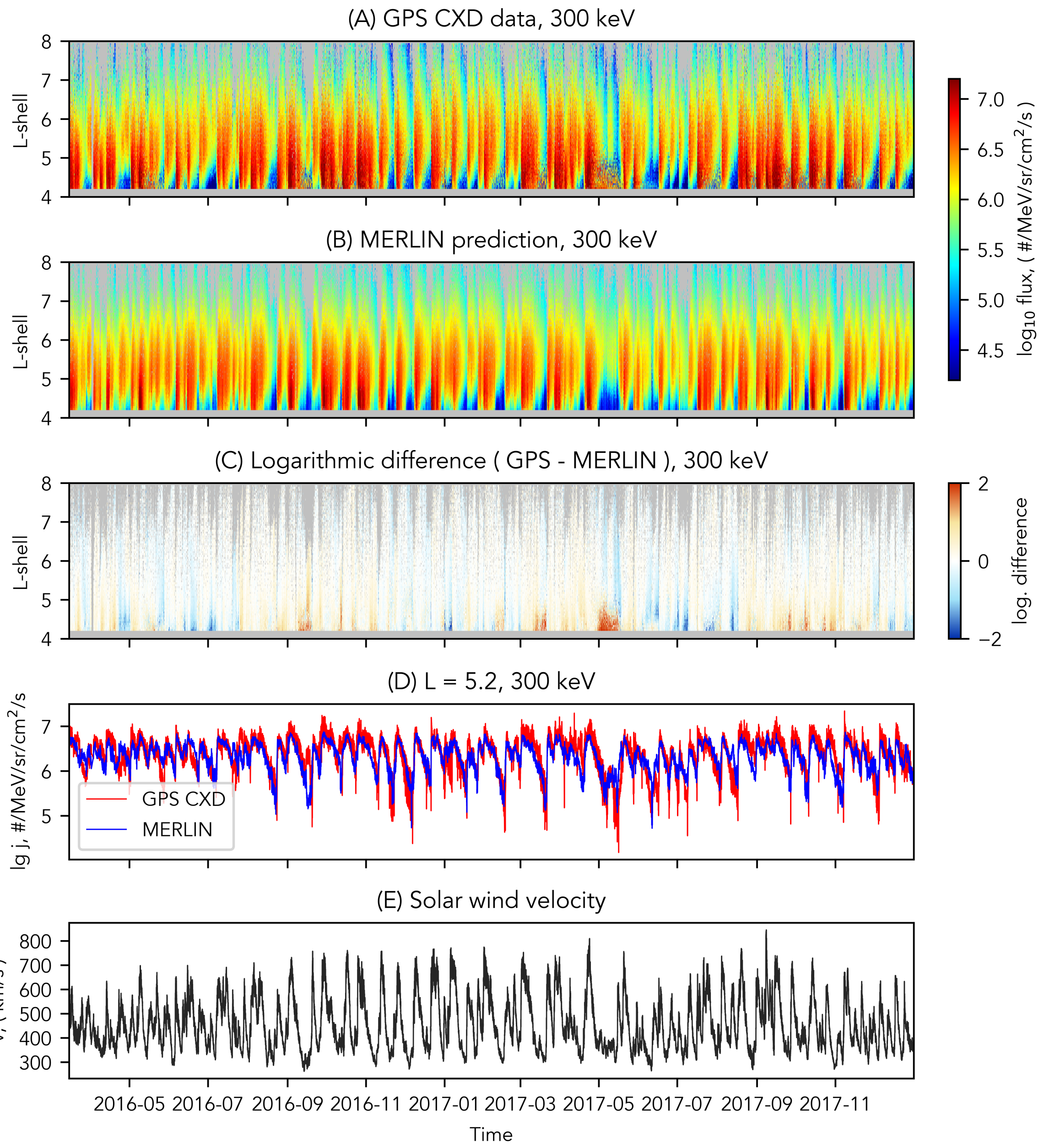
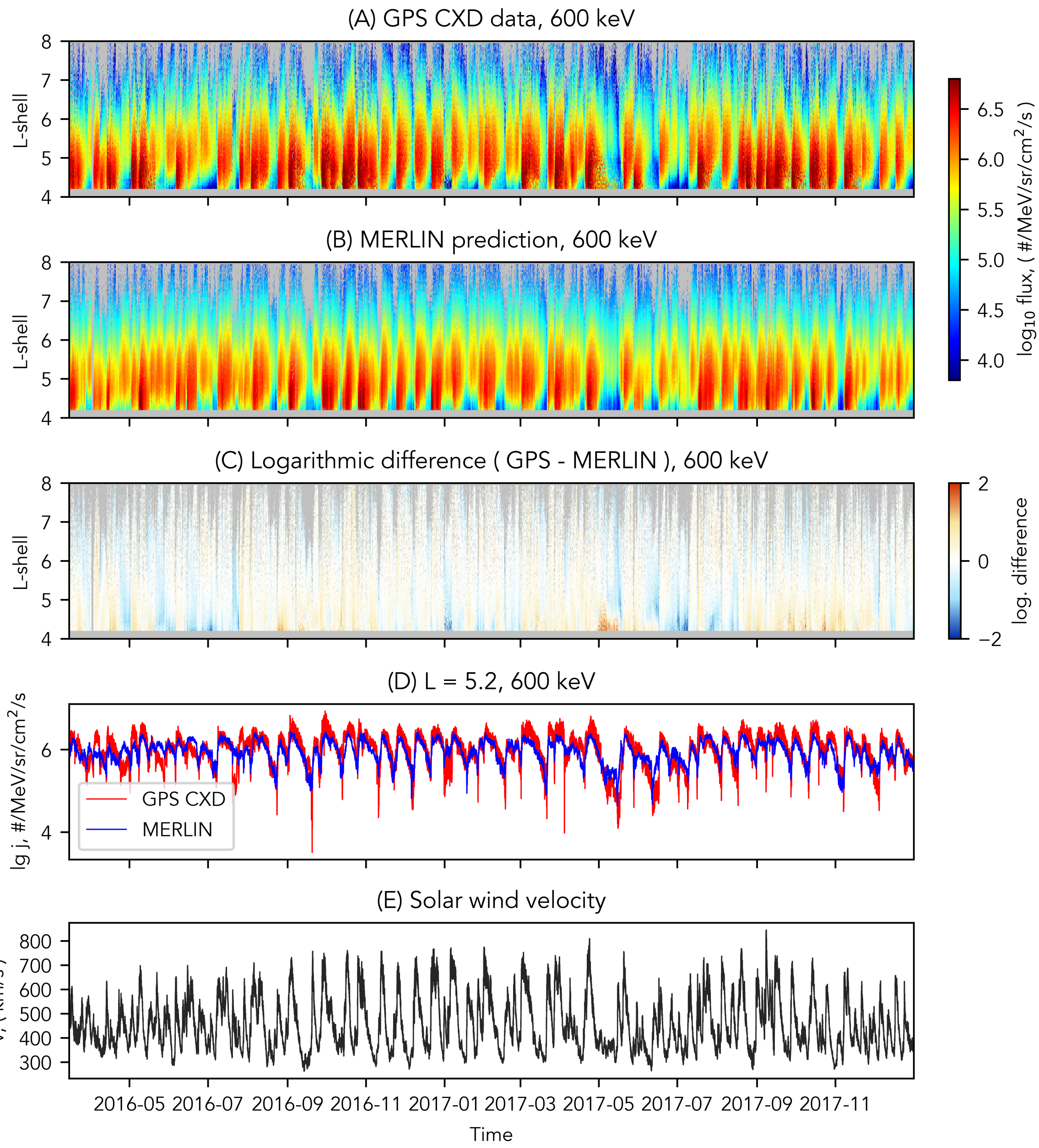


Figure 7. Model performance on the test set for 300 keV electron flux. (a) GPS CXD measurements; (b) prediction using the MERLIN model; (c) logarithmic difference between the observed and predicted flux; (d) comparison of observed (red) and predicted (blue) flux at the fixed L-shell of 5.2; and (e) solar wind velocity over the test time interval.



933 **Figure 8.** Same as Figure 7 but for 600 keV electron flux.

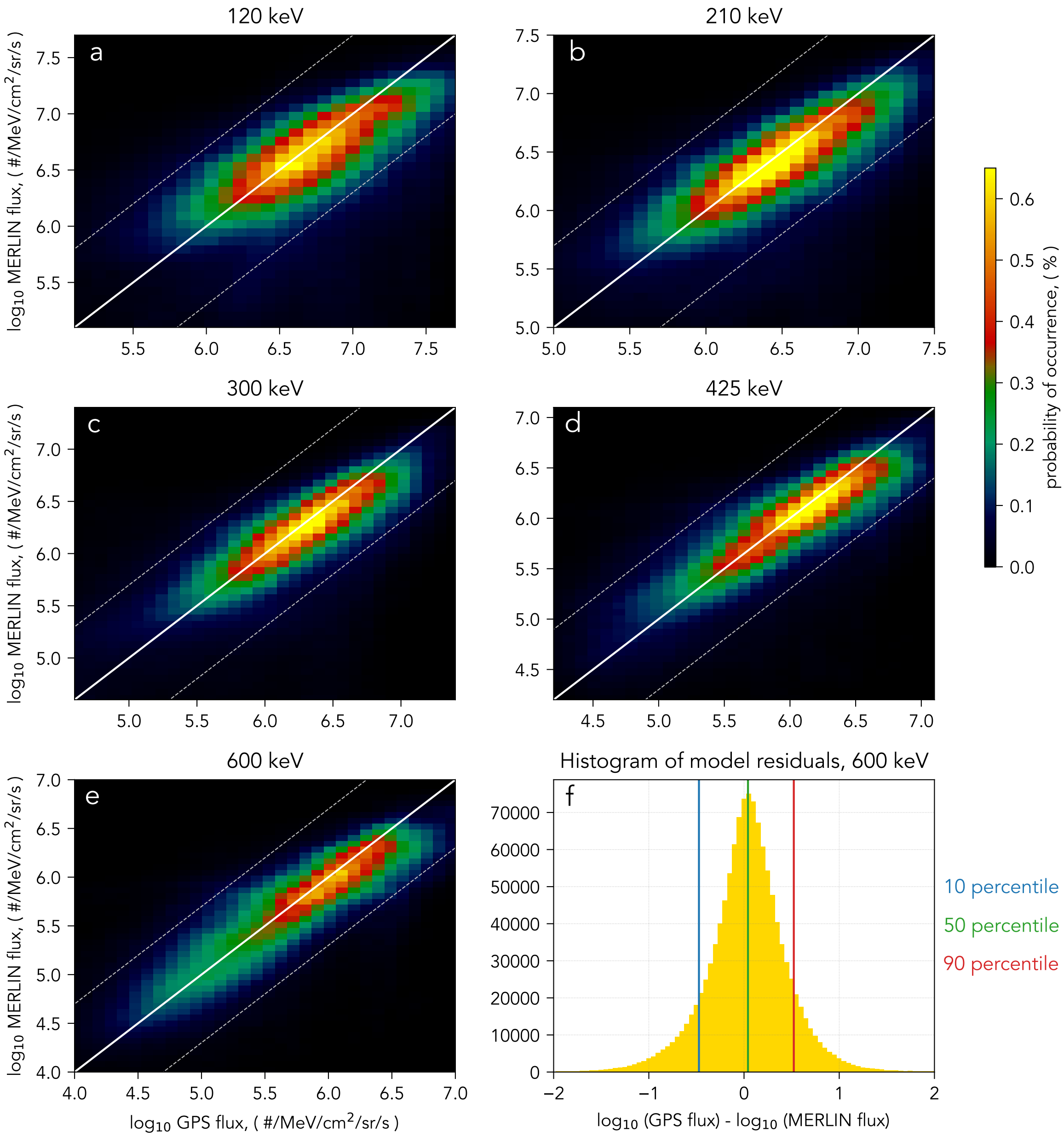


Figure 9. Probability of occurrence of the observed (on x-axis) versus predicted (on y-axis) electron flux for (a) 120, (b) 210, (c) 300, (d) 425, (e) 600 keV for test data. The white lines show the one-to-one ratio between the observed and predicted flux. The silver dashed lines give the threshold within a factor of 5. (f) shows an example of the histogram of the model residuals for 600 keV electron flux.

Supporting Information for ”Medium Energy Electron Flux in Earth’s Outer Radiation Belt (MERLIN): A Machine Learning Model”

A.G. Smirnov[1,2], M. Berrendorf[3], Y.Y. Shprits[1,2,5], E.A. Kronberg[4,6], H.J. Allison[1], N.A. Aseev[1,2], I.S. Zhelavskaya[1,2], S.K. Morley[7], G.D. Reeves[7], M.R. Carver[7], and F. Effenberger[1]

[1]Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany.

[2]Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany.

[3]Department of Database Systems and Data Mining, Ludwig-Maximilians University of Munich, Munich, Germany.

[4]Department of Earth and Environmental Sciences, Ludwig Maximilians University of Munich, Munich, Germany.

[5]Department of Earth, Planetary and Space Sciences, University of California, Los Angeles, California, USA.

[6]Max Planck Institute for Solar System Research, Göttingen, Germany.

[7]Los Alamos National Laboratory, New Mexico, USA.

Contents of this file

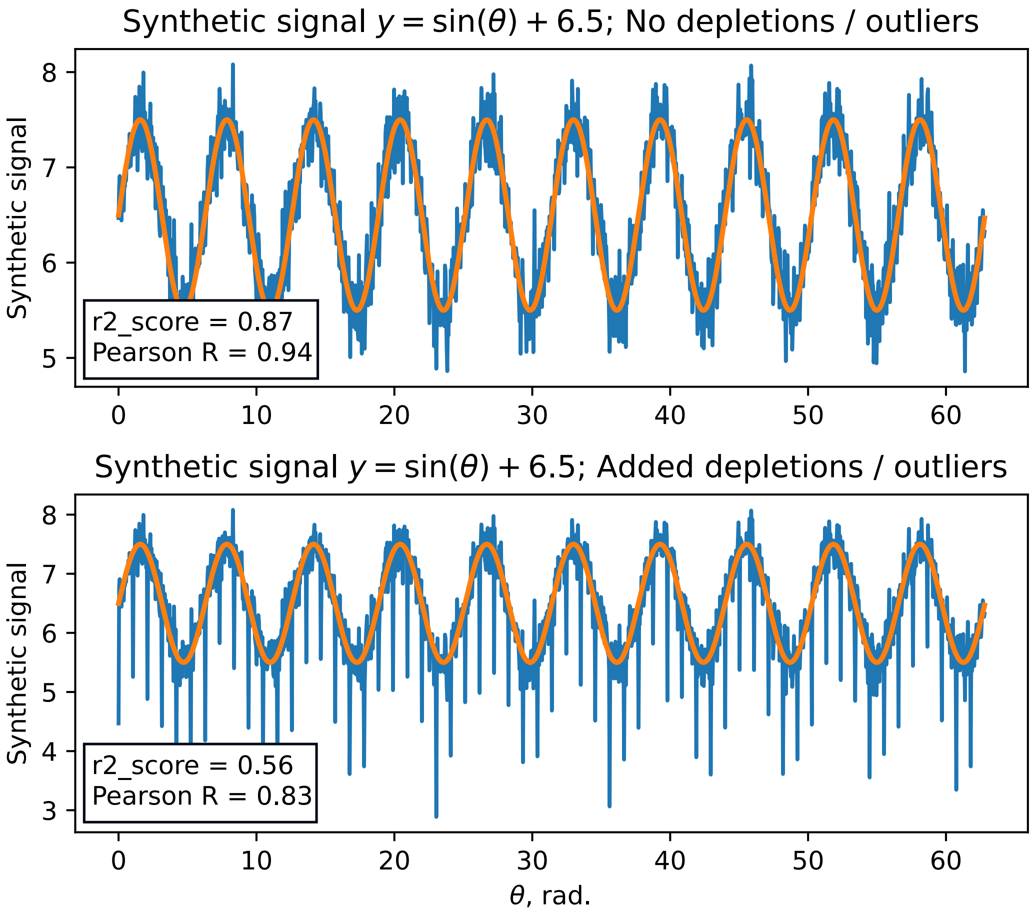


Figure S1. (a) Sinusoid signal with added random noise (blue), and the pure sinusoid signal (orange); r2-score value is 0.87. (b) Sinusoid signal with random noise and 60 added outliers (blue), and the original sinusoid (orange); r2-score value is 0.56.

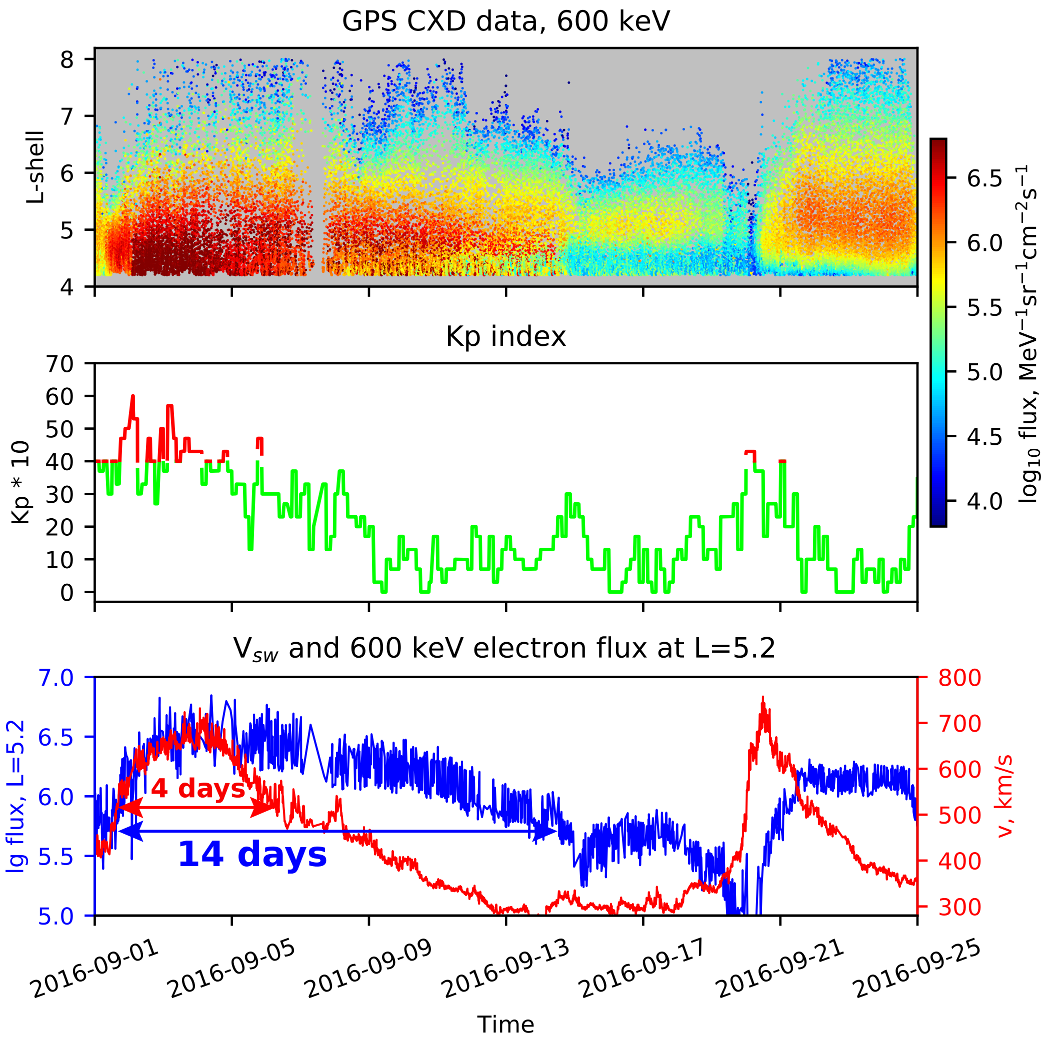


Figure S2. (a) GPS electron flux data merged for 21 satellites with CXD detectors in September 2016. (b) Kp-index during this interval. (c) 600 keV electron flux at a fixed L=5.2 (blue), and solar wind velocity (red). After the flux enhancement on the 1-2 September 2016, the velocity dropped to the quiet-time value in ∼4-5 days and continued decreasing up to $<$ 300 km/s, while the flux took approximately 14 days to decay to the pre-storm value.