Data-driven discovery of Fokker-Planck equation for the Earth's radiation belts electrons using Physics-Informed Neural Networks

Enrico Camporeale[1], George J Wilkie[2], Alexander Yurievich Drozdov[3], and Jacob Bortnik[3]

[1]University of Colorado
[2]Princeton Plasma Physics Laboratory
[3]University of California Los Angeles

November 26, 2022

Abstract

We use the framework of Physics-Informed Neural Network (PINN) to solve the inverse problem associated to the Fokker-Planck equation for radiation belts' electron transport, using four years of Van Allen Probes data. Traditionally, reduced models have employed a diffusion equation based on the quasilinear approximation. We show that the dynamics of "killer electrons' is described more accurately by a drift-diffusion equation, and that drift is as important as diffusion for nearly-equatorially trapped $\sim$1 MeV electrons in the inner part of the belt. Moreover, we present a recipe for gleaning physical insight from solving the ill-posed inverse problem of inferring model coefficients from data using PINNs.
Furthermore, we derive a parameterization for the diffusion and drift coefficients as a function of L only, which is both simpler and more accurate than earlier models. Finally, we use the PINN technique to develop an automatic event identification method that allows to identify times at which the radial transport assumption is inadequate to describe all the physics of interest.

Data-driven discovery of Fokker-Planck equation for the Earth's radiation belts electrons using Physics-Informed Neural Networks

E. Camporeale[1,2], George J. Wilkie[3], Alexander Drozdov[4], Jacob Bortnik[4]

[1]CIRES, University of Colorado, Boulder, CO, USA
[2]NOAA, Space Weather Prediction Center, Boulder, CO 80305
[3]Princeton Plasma Physics Laboratory, Princeton, NJ, USA
[4]University of California Los Angeles, CA, USA

Key Points:

- A drift mechanism is often comparable with diffusion: we analyze their relative importance, with varying L, geomagnetic activity, and phase space density values.
- We derive a simple and interpretable parameterization of drift and diffusion coefficients as functions of L only.
- We use the PINN framework for automatically identify events for which the one-dimensional radial approximation does not hold.

Corresponding author: Enrico Camporeale, `enrico.camporeale@noaa.gov`

Abstract

We use the framework of Physics-Informed Neural Network (PINN) to solve the inverse problem associated to the Fokker-Planck equation for radiation belts' electron transport, using four years of Van Allen Probes data. Traditionally, reduced models have employed a diffusion equation based on the quasilinear approximation. We show that the dynamics of "killer electrons" is described more accurately by a drift-diffusion equation, and that drift is as important as diffusion for nearly-equatorially trapped $\sim$1 MeV electrons in the inner part of the belt. Moreover, we present a recipe for gleaning physical insight from solving the ill-posed inverse problem of inferring model coefficients from data using PINNs. Furthermore, we derive a parameterization for the diffusion and drift coefficients as a function of L only, which is both simpler and more accurate than earlier models. Finally, we use the PINN technique to develop an automatic event identification method that allows to identify times at which the radial transport assumption is inadequate to describe all the physics of interest.

1 Introduction

The mechanisms that regulate the acceleration, transport, and loss of energetic particles in the Earth's radiation belts have long been investigated, both from the standpoint of fundamental research, and for practical space weather applications (Horne et al., 2005). In this region, so-called "killer electrons" can be accelerated to relativistic energies in just a few days, or even minutes, posing a dangerous threat to satellites (Horne, 2007). The radiation belts are composed of a collisionless, tenuous plasma that obeys Maxwell's equations and whose distribution can be described by the first-principle Vlasov equation. However, due to the massive temporal and spatial separation of the leading physical processes, the customary approach to study radiation belt electrons is to use a model reduction known as quasi-linear theory, introduced in the seminal paper (Kennel & Engelmann, 1966), and soon adopted in radiation belt physics (Lyons et al., 1972; Summers et al., 1998). The motion of charged particles in a dipolar magnetic field can be decomposed into three quasi-periodic orbits and corresponding adiabatic invariants. In the quasi-linear procedure one can expand particle orbits around their unperturbed trajectories in the Vlasov-Maxwell equations, and derive a diffusion equation in adiabatic invariant space (Schulz & Lanzerotti, 2012). The scattering due to resonant wave-particle interactions violates the conservation of adiabatic invariants and it is responsible for most of the particle dynamics (since collisions are absent in this tenuous plasma environment). These effects can be described by the diffusion coefficients, hence dramatically reducing the complexity of the model. Furthermore, given the different timescales associated to the three adiabatic invariants, one can decouple the diffusion in the radial direction from the one in energy and pitch angle, ending up with a one-dimensional diffusion equation, valid for particles at a constant value of the first and second adiabatic invariants. Alternatively, one can describe the time evolution of the particles' Phase Space Density (PSD) as a stochastic process due to small random changes in the variables, which leads to the one-dimensional Fokker-Planck equation (Chandrasekhar, 1943):

$$\frac{\partial f(\Phi,t)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial \Phi^2}(D_\Phi f(\Phi,t)) - \frac{\partial}{\partial \Phi}(C_\Phi f(\Phi,t)) \tag{1}$$

where f is the particles' PSD, Φ is the third adiabatic invariant (magnetic flux enclosed by a drift shell), t is time, and Eq.(1) is understood to be valid for constant values of first and second adiabatic invariants. The drift and diffusion coefficients (C_Φ and D_Φ, respectively) have the physical meaning of mean displacement and mean square displacement per unit time. Typically, Eq.(1) is further simplified by assuming a simple relationship between C_Φ and D_Φ, which can be derived in the case of a dipole field (Fälthammar, 1966) or in absence of source or sinks (Roederer & Zhang, 2016): $C_\Phi = 1/2(\partial D_\Phi/\partial \Phi)$ so that, upon transforming Φ to the normalized equatorial radial distance L we get the familiar expression:

$$\frac{\partial f(L,t)}{\partial t} = L^2 \frac{\partial}{\partial L}\left(\frac{D_{LL}}{L^2}\frac{\partial f(L,t)}{\partial L}\right). \tag{2}$$

Eq.(2) has constituted the backbone of a large part of radiation belt research for the past 60 years, and even though it is now understood that energy and pitch angle diffusion are crucial ingredients for an accurate description of electrons dynamics (Y. Y. Shprits et al., 2009; Thorne, 2010; Xiao et al., 2010; Tu et al., 2013), the relative importance of radial diffusion is still vigorously debated (Lejosne & Kollmann, 2020). Although the radial diffusion coefficient D_{LL} can be calculated from first-principles (Liu et al., 2016), as well as for event-specific cases (Tu et al., 2012; Ripoll et al., 2016; L.-F. Li et al., 2020) (keeping in mind the several assumptions built in the quasi-linear approximation (Camporeale, 2015a)), its specification requires detailed knowledge about the power spectrum and distribution of Ultra Low Frequency (ULF) waves that are resonant with electrons (Ozeke et al., 2012; Dimitrakoudis et al., 2015). Hence, most of the research focus has been centered on finding an efficient and accurate empirical parameterization of the diffusion coefficient D_{LL}, possibly as a function of quantities that are available in real-time. The parameterizations most used in the literature use the geomagnetic index Kp as the main driver. The model by Brautigam and Albert (2000) (henceforth BA) is possibly the most widely used parameterization of D_{LL} as a simple function of Kp and L. More recent works include Ozeke et al. (2014); Lejosne (2019); Ali et al. (2016); Drozdov et al. (2020); Wang et al. (2020). A Bayesian approach that accounts for possible source of uncertainties has been presented in Sarma et al. (2020).

Here, we approach the problem of defining and parameterizing the coefficients of the radial diffusion equations from a pure data-driven standpoint and, for the first time, using machine learning techniques. Since Eq.(2) does not account for any injection or loss due to non-diffusive processes, it is customary to add a source/loss term in the form f/τ. When τ is a general function of L and t, that term is general enough to account for all processes that are not included in the diffusive term. In practice, because we want to be able to distinguish losses (for instance due to particles falling into the loss-cone) from sources (for instance due to scattering in energy and pitch angle) we split the loss/source term as:

$$\frac{\partial f(L,t)}{\partial t} = L^2 \frac{\partial}{\partial L}\left(\frac{D_{LL}}{L^2}\frac{\partial f(L,t)}{\partial L}\right) - \frac{f(L,t)}{\tau} + \frac{f(L,t)}{S} \qquad (3)$$

where both τ and S are defined positive, and have the units of time. Eq. (3), however is not solvable as an inverse problem, being strongly ill-posed: there is no unique solution and, in fact, a trivial solution is one where $D_{LL} = 0$ and all the rate of change in f is accounted for by the source/loss terms. A possible way to alleviate such ill-posedeness is to enforce a given parameterization to the coefficients. That approach has successfully been followed in Sarma et al. (2020); however, it inevitably restricts the functional form of the free parameters and it possibly misses more general and insightful solutions. Here, we follow a different strategy to alleviate the problem of ill-posedeness. We generalize Eq.(2) to an advection-diffusion Fokker-Planck equation of the form:

$$\frac{\partial f(L,t)}{\partial t} = L^2 \frac{\partial}{\partial L}\left(\frac{D_{LL}}{L^2}\frac{\partial f(L,t)}{\partial L}\right) - \frac{\partial}{\partial L}(Cf(L,t)), \qquad (4)$$

with $C(L,t)$ a positive-definite drift coefficient. The positiveness of C imposes a constraint on the solution, yet still allowing the drift term to effectively act as both a source or a loss term with respect to the diffusive term (i.e., it can be either positive or negative, depending on the sign of the derivative). In other words, we seek a solution of the Fokker-Planck equation in drift-diffusion form (Eq. 1), without assuming any relationship between the drift and diffusion coefficients, since in general $C_\Phi \neq 1/2(\partial D_\Phi/\partial\Phi)$ (Allanson et al., 2022; Lemons, 2012). The additional drift term is physically related to rapid particle injections into the inner magnetosphere which have often been observed by satellites, and which are not a result of a Fick's law type inward diffusive flow, but of a rapid advective flow (see, e.g. (Bortnik et al., 2008a; Z. Li et al., 2021)).

To solve this inverse problem, we use a Physics Informed Neural Network (Raissi et al., 2019) (PINN), that derives f, D_{LL}, and C as general smooth functions of L and t, by enforcing both consistency with data and a small residual of the drift-diffusion equation (4). We use three years of Van Allen Probes data (that we consider 'noiseless') in the inverse-problem. The procedure approximates the phase space density f by means of a neural network (learning from the observed data), and learns D_{LL} and C as the optimal coefficients that solve Eq. (4) for the approximated f. We emphasize that all of the physics of interest and the particle dynamics are encoded in those coefficients, whose analysis then becomes extremely insightful.

We compare our results with the following benchmarks: the BA model (Brautigam & Albert, 2000), the Ozeke et al. (2014), and the Ali et al. (2016) parameterizations for the diffusion coefficients. For each of these, we use the formula presented in Y. Shprits et al. (2005) for the electron lifetime τ (widely used in the literature (Drozdov et al., 2017)). The forward model is computationally very cheap and it is solved with the finite difference method presented in Welling et al. (2012) (slightly adjusted by substituting the advection term $\partial(Cf)/\partial L$ in lieu of the loss term f/τ).

This work has several goals. First, we present the first ever application of the PINN framework to solve an inverse problem and deriving the optimal coefficients for the radial transport problem using real space observations. Although PINN is gaining increasing attention in all fields of applied mathematics and engineering, its potential in space physics is still not fully realized (Bortnik & Camporeale, 2021). Second, we showcase some examples of data mining approaches that can deepen our physical understanding and possibly unveil new processes. We emphasize that all of the physics of interest and the particle dynamics are encoded in the drift and diffusion coefficients, whose analysis is extremely insightful. We regard that as a fine example of data-driven knowledge discovery, which is one of the ultimate goals of using machine learning in physics (Camporeale, 2019). Third, we perform data-driven discovery of the physics which is missing in the traditional quasi-linear diffusion equation, routinely used to study electrons in the radiation belts. We show that the drift term is often comparable with the diffusion one, and we analyze in detail their relative importance, with varying L, geomagnetic activity, and phase space density values. Fourth, we derive what is possibly the simplest and most interpretable parameterization of drift and diffusion coefficients as functions of L only, that is still able to capture most of the dynamics. We show that this parameterization is competitive and often outperforms less interpretable parameterizations presented in the literature. Eventually, we achieve one of the most important and long-standing goals of scientific machine learning: we use a general but opaque ML technique (PINN) to solve an inverse problem and we discover that the free parameters of our Fokker-Planck equation (diffusion and drift coefficients) can be well approximated by a simple, interpretable formula. That is, we perform data-driven, ML-aided model order reduction. Finally, we use the PINN solution for an automatic event identification task, namely to identify events for which the one-dimensional radial approximation does not hold, requiring other physical mechanisms, such as energy and pitch-angle resonant interactions.

2 Methods

2.1 Forward model

Eq.(4) is solved by means of an unconditionally stable, second order accurate, Crank-Nicholson scheme discussed in Welling et al. (2012). For completeness, we report the numerical discretization here:

$$
\begin{aligned}
\frac{f_j^{n+1}-f_j^n}{\Delta t} = \frac{L_j^2}{2\Delta L^2}\Big[& D_{j+\frac{1}{2}}^{n+\frac{1}{2}}(f_{j+1}^n - f_j^n + f_{j+1}^{n+1} - f_j^{n+1}) \\
& -D_{j-\frac{1}{2}}^{n+\frac{1}{2}}(f_j^n - f_{j-1}^n + f_j^{n+1} - f_{j-1}^{n+1})\Big] \\
& - \frac{1}{4\Delta L}\Big[C_{j+1}^{n+\frac{1}{2}}(f_{j+1}^{n+1}+f_{j+1}^n) - C_{j-1}^{n+\frac{1}{2}}(f_{j-1}^{n+1}+f_{j-1}^n)\Big] \quad (5)
\end{aligned}
$$

where indexes n and j represent discretization in time and space, with time steps Δt and ΔL, and $D_j = D_{LL,j}/L_j^2$, respectively. Eq. (5) is a linear equation that can be written in matrix form with tri-diagonal matrices and is solved by a standard LU decomposition. For all the results presented, we use $\Delta t = 1$ (hours) and $\Delta L = 0.05$. Observations at $L = 2.0$ and $L = 5.5$ are used as time-dependent boundary conditions, while initial conditions are interpolated from the data.

2.2 Physics-Informed Neural Networks

Physics-informed Neural Networks (PINN) are a framework for solving forward and inverse problems involving nonlinear partial differential equations (Raissi et al., 2019). The theoretical foundation of PINNs lies on the well-known universal approximation property of neural networks (Hornik et al., 1989) that essentially allows neural networks to accurately approximate a large class of continuous functions. The basic idea of PINNs is rather simple, and it exploits the fact that the output of a neural network is a continuous and differentiable function (almost everywhere). Moreover, PINNs take advantage of the ability of modern neural network libraries to automatically calculate exact derivatives with respect to the input variables, by applying the chain rule of differentiation (this is known as *autodiff* in machine learning jargon (Géron, 2019)). Hence, each term in a partial differential equation (PDE) can be calculated exactly on a set of collocation points within the domain, and the PDE itself can be used as penalization term in the loss function minimized by the neural network. Upon convergence, a PINN outputs a function that approximately solves the PDE and matches the given data on the points where it has been trained.

An interesting feature of PINNs that we use in this work is their ability to solve inverse problems in a mesh-free fashion and with a minimal set of assumptions. However, the possibility of finding general forms for the free parameters of a PDE has the potential drawback of the converged solution not being unique. We approach this issue by employing an ensemble method, namely by solving the inverse problem several times and averaging the top 5 solutions. Because the solution f spans several orders of magnitude in the L domain, we perform the transformation $f = e^g$ and solve for g:

$$\frac{\partial g}{\partial t} = L^2 \frac{\partial}{\partial L}\left(\frac{D_{LL}}{L^2}\frac{\partial g}{\partial L}\right) + D_{LL}\left(\frac{\partial g}{\partial L}\right)^2 - \frac{\partial C}{\partial L} - C\frac{\partial g}{\partial L} \tag{6}$$

The PINN is designed as a combination of three coupled neural networks, each taking a point in (L, t) as input and outputting the value of f, D_{LL}, and C at that point, respectively. Those three outputs are then combined in the loss function, which is the sum of the mean square error with respect to the observations, and the residual of Eq.(6). Boundary conditions (at $L = 2$ and $L = 5.5$) are enforced by neglecting the residual term in the loss function on those points (that is, the function f is forced to converge to the boundary values). The neural network architectures are standard, and have been selected by progressively increasing their complexity and monitoring changes in the converged values of the loss function until a plateau was observed. Other hyper-parameters were not optimized. The networks use a tanh activation function in all the layers. The network that outputs the solution f uses 6 inner layers with $[30, 20, 20, 20, 20, 20]$ neurons, while the two networks outputting the coefficients D_{LL} and C have 3 inner layers with $[30, 20, 10]$ layers. To perform the optimization we use a combination of the Adam optimizer (Kingma & Ba, 2014) and the BFGS (Broyden-Fletcher-Goldfarb-Shanno) method (Zhu et al., 1997), both within the Tensorflow framework (Abadi et al., 2016).

2.3 Data

We use observations from the Magnetic Electron Ion Spectrometer (MagEIS) instruments aboard the Van Allen Probes spacecraft (Blake et al., 2013). Van Allen Probes is a NASA twin satellite mission that was active for 7 years, since its launch on August 30th, 2012. Its primary mission was to address how populations of high energy charged particles are created, lost and dynamically evolve within Earth's magnetic trapping region (Fox & Burch, 2014). Due to the unprecedented quality and quantity of data collected, Van Allen Probes have marked a golden era for radiation belt studies (W. Li & Hudson, 2019). Here, we limit our study to electrons with first adiabatic invariant $\mu = 700$ MeV/G and second adiabatic invariant $K = 0.1\ R_E\ \mathrm{G}^{0.5}$, which corresponds to approximately 1 MeV electron energies in the heart of radiation belt. We used the TS05 magnetic field model (Tsyganenko & Sitnov, 2005) to calculate the adiabatic invariants. The dataset is comprised of ∼570,000 data points spanning the time range 01-Nov-2013 to 30-Sep-2017. The largest interval between consecutive data points is 2:45 hours, and the average interval is about 4.5 minutes.

Figure 1 shows the PSD (log scale) of the whole dataset as a function of L. The vertical dashed line divides the dataset into training set (70% of the whole dataset, from 01-Nov-2013 to 30-Oct-2016) and test set (30% of the whole dataset, from 01-Nov-2016 to 30-Sep-2017). One can notice that the dataset is sparse both in time and space, since it essentially follows from the highly elliptical trajectory of the satellites.

2.4 Metrics and benchmarks

Our quantity of interest, the phase space density f, changes by several orders of magnitude between $L = 2$ and $L = 5.5$. Hence, it is not straightforward to design a single metric for model performance. A through analysis of several metrics often used in radiation belt modeling, can be found in Morley et al. (2018); Liemohn et al. (2021). Here, we are interested in studying the model accuracy at given values of L, rather than averaged over the whole domain. We define and use three different errors. Following Morley et al. (2018), we characterize accuracy by defining the *percentage symmetric accuracy* ζ as:

$$\zeta_k = 100 \cdot \exp(P_k(|\log(f/\hat{f})|)), \tag{7}$$

where $\hat{f}$ and f are the ground-truth values taken by observations and the corresponding values produced by a model, respectively. P_k represents the $k-$th percentile (i.e. P_{50} is the median) calculated over all values at fixed L. This represents a generalization of the median symmetric accuracy (Morley, 2016) for quantiles other than the median, that allows to estimates error bars (that is, ζ_k is monotonically increasing with increasing k). The second metric we employ characterizes bias and is called the *symmetric signed percentage bias* SSPB, again generalized from the definition in Morley et al. (2018):

$$\mathrm{SSPB} = 100 \cdot \mathrm{sgn}(P_{50}(\log(f/\hat{f})))(\exp(|P_{50}(\log(f/\hat{f}))|) - 1) \tag{8}$$

Note that, by taking the absolute value after calculating the percentile, SSPB is not ordered when considering different percentiles P_k (hence it does not allow to estimate error bars). Finally, we define the relative error ε as the median value at fixed L of the relative error of the logarithmic phase space density. That is:

$$\varepsilon(L) = P_{50}\left(\frac{\log_{10} f - \log_{10} \hat{f}}{\log_{10} \hat{f}}\right) \tag{9}$$

We benchmark our results against several parameterization for the diffusion coefficient: the BA model (Brautigam & Albert, 2000), the Ozeke et al. (2014), and the Ali et al. (2016), which are all functions of L and the geomagnetic index Kp only (Rostoker, 1972). Their formula are:

$$
\begin{aligned}
D_{LL}^{BA} &= L^{10} \cdot 10^{(0.506Kp-9.325)} \\
D_{LL}^{Ozeke} &= 2.6 \cdot L^{6} \cdot 10^{(0.217L+0.461Kp-8)} \\
&+ 6.62 \cdot L^{8} \cdot 10^{(-0.0327L^2+0.625L-0.0108Kp^2+0.499Kp-13)} \\
D_{LL}^{Ali} &= \exp(-16.951 + 0.181Kp \cdot L + 1.982L) \\
&+ \exp(-16.253 + 0.224Kp \cdot L + L)
\end{aligned}
$$

The following definition of electron lifetime is employed (Drozdov et al., 2017) for the BA and Ozeke et al. parameterizations:

$$
\begin{aligned}
\tau &= 10 \text{ for } L \leq L_{pp} \\
&= 6/Kp \text{ for } L > L_{pp}
\end{aligned}
$$

where L_{pp} is the plasmapause location, empirically estimated with the formula in Carpenter and Anderson (1992). The Ali et al. parameterization does not use a loss term.

2.5 Ensemble approach

We have solved the PINN described above for 20 different random initializations of the underlying neural networks, each time training for 100,000 epochs (we note that some of the networks might have converged with a smaller number of iterations). The best 5 solutions in terms of the error $\varepsilon(L)$, Eq. (9), computed on the training set are shown in Figure 2 as black lines. Blue, magenta and yellow lines denote the BA, Ozeke et al. and Ali et al. solutions, respectively. Not surprisingly, the PINN solutions consistently outperform those three benchmark solutions. However, it is interesting that the simple approach of averaging the best 5 diffusion and drift coefficients yields a result that also outperforms the benchmarks and indeed is very close to each of the 5 ensemble members. The error of the PINN ensemble mean is shown in Figure 2 as a red line. This is not a trivial result, because from Eq.(6) one can see that averaging the coefficients D_{LL} and C does not yield a solution that is the average of the ensemble members solutions.

Figure 3 shows the best five realizations of the diffusion coefficient D_{LL} (top panels) and the corresponding drift coefficient C (bottom panels) as heat maps in logarithmic scale and as a function of time (horizontal axis) and L (vertical axis). Figure 4 shows the ensemble mean (average of the best five) for D_{LL} (left) and C (right).

3 Results

3.1 Statistical analysis of coefficients

Here we perform a statistical analysis of the optimal coefficients derived with PINN on the training set. First, we show in Figure 5 the distribution of the PSD f as a function of L. The heat map shows the counts in each bin, normalized to the largest number for a constant value of L. The statistics are computed on about 25,000 times instances, spanning 3 years of data (01-Nov-2013 to 30-Oct-2016). One can notice that three regimes naturally appear: one for $L \lesssim 3.2$ where f is approximately constant at levels of 10^{-10}, one for $3.2 < L \lesssim 4.5$ where f rapidly increases and it has a large spread covering the range $10^{-10} < f < 10^{-4}$ and a third regime at larger L where the $L-$dependence is again flattened, even though the spread in values remains relatively large. Figure 6 (left panel) shows the distribution of the diffusion coefficient D_{LL} as function of L. The gray area represents the interval between the 25th and 75th percentile (for a given L), and the orange line denotes the median. One can notice that the spread increases by moving further away from the coordinate $L \sim 3.2$. Also, the slope of the distribution undergoes several regimes. For reference, we overlay the curves L^{10} (yellow) and L^{20} (magenta). The former is adopted in the BA parameterization and is consistent with the distribution of D_{LL} for small L, while for large L the latter $L-$dependence seems more appropriate. A more detailed examination of this distribution is shown in the right panel of Figure 6. Here, we have ranked column-wise (i.e. for constant L) the number of counts in each bin (the bins are uniformly spaced in $\log_{10} D_{LL}$ and L). The heat map shows the top 20 ranks, with black signifying the top rank (i.e. bins with the largest number of counts at constant L, and white the lowest rank (20 or above). In this way we are able to distinguish different *trajectories* for D_{LL}, and in particular a bifurcation of values, particularly at large L. The same bifurcation is even more prominent in the distribution of C, shown with the same format in Figure 7, where one can notice two different regimes being approximately separated at $L \sim 3.5$. Interestingly, for $L > 3.5$, C can vary by one or two orders of magnitude.

The presence of (at least) two distinct regimes confirms that the physics of interest is different within and outside the plasmapause. Here we do not explicitly model the plasmapause location (see, e.g. (Malaspina et al., 2020; Guo et al., 2021; Chu et al., 2017)), hence the change in the distributions slopes between L=3 and L=3.5 should be attributed to a statistically averaged plasmapause location. The spread in the coefficients is harder to interpret physically, although certainly driven by variations in the boundary conditions at $L = 2$ and $L = 5.5$. We note that one of the important aspects of PINN-based insight discovery is identifying regions in parameter space that are poorly constrained or carry greater error, as specific areas that require better understanding and further investigation. Finally, Figure 8 shows the ranked joint distribution of D_{LL} (horizontal axis) and C (vertical axis). Both quantities are in logarithmic scale. While there seems to be an almost linear dependence between the two coefficients for relatively small values ($\lesssim 10^{-2}$), several branches appear for large values, possibly indicating different physical regimes.

3.2 Relative importance of drift and diffusion terms

In order to understand the relative importance of the diffusion and drift terms in Eq. (4) we define their ratio as $r = \left|\frac{1}{L^2}\left(\frac{\partial Cf}{\partial L}\right) / \left[\frac{\partial}{\partial L}\left(\frac{D_{LL}}{L^2}\frac{\partial f}{\partial L}\right)\right]\right|$. Figure 9 shows the distribution of r (in logarithmic scale, vertical axis) as a function of L (horizontal axis). The distribution is normalized to the maximum value of counts per L-value. The black solid line at $\log_{10} r = 0$ indicates equal balance between drift and diffusion, and the region below that line represents a stronger diffusion than drift. One can notice that in the inner magnetosphere ($L \lesssim 4$) the two terms are approximately balanced, while diffusion plays a larger role with increasing L in the outer belt. Figure 9 can be interpreted in the sense of local versus global losses, where the former are captured by the drift term and the latter by the diffusion term. Typically, local diffusion at $\mu = 700$ MeV/G is controlled by the hiss and chorus waves and radial diffusion becomes very low at lower L-shell. On the other hand, hiss waves will more likely be a cause of local losses at low L-shell, providing a steady decay time, shorter than the one due to radial diffusion. It is important to notice that this picture might change for lower μ values, which is something that can be explored in the future using this technique.

We further analyze the relative contribution of the drift and diffusion terms by studying the ratio r as a function of $\log_{10} f$ and L, and for different geomagnetic activity, represented by the Auroral Electrojet index AE, in Figure 10 (left panel: $AE < 100$, middle panel: $100 \leq AE < 300$, right panel: $AE >= 300$). Interestingly, at low L drift is more dominant than diffusion for larger values of PSD. Also, the range of L in which diffusion is dominant slightly shifts to smaller L with increasing geomagnetic activity. This analysis unambiguously shows an unexpected relatively large contribution of non-diffusive drift in the time evolution of the phase space density.

3.3 Effective electron lifetime and sources

As explained above, in this approach electron losses and sources are not included explicitly, and the last two terms of Eq. (3) (f/τ and f/S) are replaced by a drift term. However, *effective* lifetimes associated to losses and sources can be derived at each point in time and space by calculating $f/(\partial C f(L,t)/\partial L)$ and defining this quantity as τ when it is positive, and $-S$ when it is negative. Notice that both τ and S are positive and have the units of days. Their distribution is shown in Figure 11, as functions of L (logarithmic vertical scale). Here, the different shades of gray denote the area covered by [1-99], [10-90], and [25-75] percentiles at a constant L value. Once again, a distinguishing features is the existence of two regimes: for small L both lifetimes are very large (i.e. the corresponding loss/source terms f/τ and f/S are negligible), but their value decreases substantially with increasing L until they plateau at large L. It is interesting that the range of values taken by τ (i.e. the gray area) also increases significantly with larger L, to the point that at $L = 5$, τ can range approximately 3 orders of magnitude. In the left panel of Figure 11, the black line denotes the parameterization by Y. Shprits et al. (2005) used in the BA and Ozeke et al. models. The underestimation of τ at small L might be the cause of the large errors for low L in those models (see Figure 2).

Several mechanisms that locally enhance the phase space density have been investigated in the literature (Boyd et al., 2018a; Hudson et al., 2020; Jaynes et al., 2015a). Figure 12 shows the source term S over the whole training set, in space (vertical axis) and time (horizontal axis). The interesting feature is that local injection of phase space density can sporadically extend to low values of L, down to $L \sim 3.5 - 4$. Although in the majority of cases the timescale associated with such injections are of the order of tens or hundreds of days, there are cases where $S \sim 1$ day, hence comparable with the timescale of local diffusion and losses.

3.4 Feature selection

The PINN method described above derives D_{LL} and C as generic functions of time and space (t, L), spanning the whole training period. In order to understand the relationship between the diffusion and drift coefficients and their physical drivers, here we perform a feature selection analysis. This analysis can be used, in later works, to inform machine learning models that seek to generate D_{LL} and C as function of past known quantities, for space weather forecasting purposes. Feature selection is an extensive topic in the machine learning literature (see, e.g. (J. Li et al., 2018)). Here, we use the *backward elimination* technique based on generalized linear models, which we briefly describe in the following. First, we define a minimal set of features, based on our physical intuition: since the radiation belt is ultimately driven by the solar wind variability, we include solar wind quantities observed at the L1 (first Lagrangian point) and propagated in time to the magnetosphere bow-shock that are well known to be drivers of geomagnetic activity (Wing et al., 2016; Kilpua et al., 2015). Those solar wind quantities are taken from the NASA OMNI dataset. Table 1 lists the 12 features initially considered. A generalized linear model is built using all combinations of those features up to a quadratic order (a total of 91 terms for C and 78 terms for D_{LL}, including the intercept). The linear model naturally provides the standardized coefficients (so-called t-Statistic or Z-score) for each term, defined as the ratio between the coefficient calculated for that term by solving a least-square problem, and its standard deviation. A large value of the standardized coefficient rejects the hypothesis that the coefficient is zero (null hypothesis). In the backward elimination procedure we iteratively eliminate the coefficient with smallest Z-score (in absolute value) and train a new model with all the terms remaining, until only one term is left. This provide us with a ranking, or selection of the features. Figure 13 illustrates the top ten features for D_{LL} and C, respectively, as a function of the coefficient of determination R^2. The features are ranked from left to right with decreasing importance, and the reported R^2 is intended for a model that uses all features listed to the left (i.e., adding one at the time). The red dashed line represents the largest R^2 achieved when all the features are included. In order to add robustness to the procedure, each model is trained on randomly selected 80% data in the training set. It is interesting to notice that the solar wind features have lower rankings than features that use PSD and boundary conditions. In other words, the solar wind information contained in the PSD and the boundary condition is more informative for D_{LL} and C than using the solar wind directly.

A more comprehensive study of the most efficient time lag between solar wind and diffusion and drift coefficients, following the methodology of Wing et al. (2016) is under way.

Feature	Meaning
L	Spatial coordinate in Eq.(2)
Kp	Geomagnetic index
PSD	log_{10} of the phase space density
Σ PSD	Average of PSD along 10 prior hours
Σ_L PSD	Moving average of PSD along L (10 ΔL)
BC_u	PSD Upper boundary condition ($L = 5.5$)
BC_l	PSD Lower boundary condition ($L = 2.0$)
B_z	z-component of the interplanetary magnetic field
V	solar wind speed
VB_z	product of V and B_z
$Newell$	Newell coupling function (Newell et al., 2007)
D_{LL}	$log_{10}(D_{LL})$, used as a feature for C

Table 1. Features tested in the backward elimination algorithm, and their meaning.

3.5 Interpretable parameterization of drift and diffusion coefficients

Eventually, in the grand scheme of scientific machine learning, one would like to use advanced but often opaque techniques (such as PINN) to extract physical insight from the data, with the final goal of exploiting such new insights to advance our knowledge and possibly derive new interpretable models. In a sense, that follows from Occam's razor argument that suggests that one should seek the most parsimonious yet accurate model. Here, we close the circle of our inquiry by deriving what is possibly the simplest parameterization of D_{LL} and C. The feature selection procedure (Figure 13) demonstrates that most of the variance in both D_{LL} and C can be attributed to changes in L. In other words, L is the best unique predictor for the coefficients, and therefore we aim to describe them as a function of L only, by fitting the PINN-derived values of D_{LL} and C with a cubic interpolator, shown with black lines in the left panels of Figures 6 and 7, respectively.

Not surprisingly, the cubic interpolator is a good approximation of the median values. The derived formulas for the cubic fit are the following:

$$\log_{10} D_{LL} = -0.0593L^3 + 0.7368L^2 - 1.33L - 4.505$$
$$\log_{10} C = 0.0777L^3 - 1.2022L^2 + 6.3177L - 12.6115$$

In order to assess the goodness of this approximation, we use it in a forward model solution (see section Methods) and we compare the results with two benchmarks: a solution derived with the BA diffusion coefficients (Brautigam & Albert, 2000), and another derived by using the diffusion coefficients proposed in Ozeke et al. (2014) (a comparison against the Ali et al. model is not shown since it was found to yield too large errors (Drozdov et al., 2021)). For both cases we solve Eq. (2) with the addition of a loss term ($-f/\tau$), parameterized as in Gu et al. (2012); Orlova et al. (2016), since the inclusion of such term is standard practice to account for wave-particle scattering due to hiss and chorus waves, and it is known to improve accuracy (see section Metrics and Benchmarks). In Figure 14 we show (left) the percentage symmetric accuracy ζ, Eq. 7 and (right) the symmetric signed percentage bias SSPB, Eq. 8 (see Methods) calculated over the whole test set (1 year of data), as a function of L. Blue, red, and black lines denote the results from the baselines by BA and Ozeke et al., and by using the PINN-derived cubic fit, respectively. In the left panel of Figure 14, the solid squares denote the median values ζ_{50} and the error bars are calculated as the spread between ζ_{25} and ζ_{75}. In the right panel, positive values are in solid and negative values in dashed lines. One can notice that the simple cubic approximation of Eqs. (10) yields results comparable or superior to the ones obtained with more sophisticated models (all errors are by definition going to zero at the boundary).

Finally, we present in Figure 15 the PSD resulting from the forward models using the three different parameterizations (BA in red, Ozeke et al. in yellow and PINN-derived cubic fit in purple), compared against the Van Allen Probes data (blue), for the whole period covered in the test set. Top and bottom panels are for $L = 5$ and $L = 4$, respectively. In all cases, the simulations have initial and boundary conditions taken from the data. For $L = 5$, the PSD resulting from the new parameterization presented here is consistently more accurate than the two baseline models, which tend to underestimate the Phase Space Density. At $L = 4$ none of the three models is particularly accurate, although the PINN is often orders of magnitude closer to the observations than the other two models. Note that logarithmic scales are used in vertical axis.

3.6 Automatic event identification

One of the by-products of the PINN approach outlined in this paper is the possibility of studying how well the observational data are consistent with the solution of the underlying PDE. As mentioned in the Introduction, the derivation of the radial diffusion Eq.(4) is based on several assumptions, one of which is the conservation of the first and second adiabatic invariant. Breaking those invariants can cause local diffusion in energy and pitch-angle (Camporeale, 2015b; Tu et al., 2013). By investigating how small the residual of the PDE is on the domain, one can easily identify times when any of the quasi-linear assumptions do not hold and hence Eq.(4) cannot capture some of the physical mechanisms that generate the data. Figure 16 shows the residual of Eq.(4) plotted as a heat map over the whole training period. For ease of visualization, it has been normalized to its maximum value, and the color scale is capped at a value of 0.3. The red dashed lines on the bottom of the figure represent times at which the residual contain values in the 99 percentile of its the distribution. The list of these 'events' is reported in Table 2. Most of these periods are associated to moderate or strong geomagnetic storms, dropout events, or flux enhancements, and have already been studied in the literature. When that is the case, some references that explicitly analyze data from that period are cited in the last column. For other events, we have not found previous studies in the literature, and we encourage the community to analyze them.

4 Conclusion

The process of understanding the mechanisms underlying a physical process, and the ability of describing such mechanisms with the elegant and succinct formalism of partial differential equations (PDEs) lies at the core of scientific discovery. However, the way in which scientists extract information from experiments and observations (*data*) and encodes that information into PDEs has seen dramatic changes over the last decade, when methods originating in machine learning have started playing an increasingly important role. Currently, there is a rich literature on data-driven discovery of PDEs (see, e.g., (Long et al., 2018; Berg & Nyström, 2019; Raissi, 2018; Rudy et al., 2017; Xu et al., 2019; Zhang & Lin, 2018; Boullé et al., 2021; Udrescu & Tegmark, 2020)). The published methods

Table 2. List of events automatically generated. The last column indicates references in case that event has been studied in the literature.

Start time	End time	Previously studied in the literature?
30-Dec-2013	04-Jan-2014	CIR-associated storm (Shen et al., 2017)
10-Feb-2014	10-Feb-2014	
14-Feb-2014	20-Feb-2014	Geomagnetic storm due to multiple interacting ICMEs
		(Kilpua et al., 2019; Vlasova et al., 2020)
25-Jul-2014	7-Aug-2014	
08-Sep-2014	18-Sep-2014	Dropout event
		(Ozeke et al., 2017; Alves et al., 2016; Jaynes et al., 2015b; Ma et al., 2020)
24-Dec-2014	24-Dec-2014	
15-Mar-2015	20-Mar-2015	CME-associated storm
		(Shen et al., 2017; Baker et al., 2016)
15-Apr-2015	17-Apr-2015	
12-May-2015	14-May-2015	CIR-associated storm (Shen et al., 2017)
07-Jun-2015	28-Jun-2015	CIR and CME-associated storms(Shen et al., 2017; Baker et al., 2016);
		Moderate event(Reeves et al., 2020);
		Sudden Particle Enhancements at Low L Shells(Turner et al., 2017)
19-Jul-2015	23-Jul-2015	Sudden Particle Enhancements at Low L Shells (Turner et al., 2017)
17-Aug-2015	31-Aug-2015	Moderate event(Reeves et al., 2020)
05-Oct-2015	09-Oct-2015	Moderate event(Reeves et al., 2020)
03-Nov-2015	06-Nov-2015	
08-Dec-2015	11-Dec-2015	
14-Dec-2015	28-Dec-2015	Moderate and strong storms
		(Boyd et al., 2018b; L.-F. Li et al., 2020; Sotnikov et al., 2019)
27-Jan-2016	07-Feb-2016	Dropout event (Wu et al., 2020)
15-Feb-2016	19-Feb-2016	Moderate event(Reeves et al., 2020); Fast magnetosonic waves(Yu et al., 2021)
01-May-2016	14-May-2016	Moderate event(Reeves et al., 2020; Moya et al., 2017)

can be loosely divided in two classes. On one hand, one can create a large dictionary of terms that contain algebraic, differential and integral operators and search the space of all (or many) combinations of those terms for the optimal PDE that describes the data (i.e., the PDE whose solution is an acceptable approximation of the data). Two seminal examples of this approach are Rudy et al. (2017) (using sparse regression) and Udrescu and Tegmark (2020) (using symbolic regression). On the other hand, one can restrict the search for the optimal PDE to a specific class of functionals, thus setting up the problem of PDE discovery as an inverse problem, where the time and space dependence of free parameters (such as, for instance, drift and diffusion coefficients) needs to be learned. Physics-Informed Neural Network, introduced in (Raissi et al., 2019), falls in this category, and it is the approach used in this paper. Here, we have presented a framework that solves the problem of finding the optimal coefficients for a Fokker-Planck equation (inverse problem) with a Physics-Informed Neural Network, applied to the study of energetic radiation belt's electrons, and using for the first time real space satellite observations (Van Allen Probes). This approach opens several possible avenues for future investigations. In this paper, we have showcased several of them.

Specifically, we have investigated the possibility that the time evolution of the Phase Space Density of electrons in the Earth's radiation belt could be described by the combination of (and the competition between) a diffusion and a drift term. It was found that the data is more consistent with the inclusion of a non-diffusive drift mechanism and it was discovered that the phase space distribution is an important parameter in determining the coefficients. These findings challenge several decades of literature that have exclusively focused on diffusive processes.

The data-driven approach enabled by PINN allows to unambiguously test such hypothesis, by determining the optimal drift and diffusion coefficients that, used in Eq. (4), result in the solution most consistent with observations. Interestingly, we have shown that, at least for the values of first and second adiabatic invariants considered here, drift and diffusion are competing for $L \simeq 4$, while diffusion becomes increasingly dominant for larger values of L. Obviously, as powerful as it is, the PINN method does not solve the issue of ill-posedness of the inverse problem. Namely, there is no guarantee about the uniqueness of the solution. Indeed, we have verified that different realizations of the coefficients are possible and equally valid. Interestingly enough, we have also verified that not only the best 5 coefficients used in this study yield solutions that have comparable errors with respect to the data, but that the average of the coefficients (analyzed in detail in Figures 4-10) also yield a similar level of error.

Furthermore, discovering the optimal diffusion and drift coefficients allows to data-mining them in order to learn their dependence on physical parameters and the statistical behaviour of their profile (Figures 3 - 8). Second, one can re-derive effective loss and source terms, and study their behaviour in space and time (Figures 11, 12). In this way, we have discovered fast sporadic injections of PSD at $L \sim 3.5-4$ that might occur on a ~ 1 day timescale (Figure 12). The analysis has also highlighted a deficiency in modelling the loss term τ at low L in previous works (Figure 11). Third, we have used the PINN-discovered coefficients D_{LL} and C and their learned dependence on L to build a simple and interpretable model that yields an excellent approximation (and forecast) of the PSD (Figure 14), with no free parameters, other than the boundary conditions. In our opinion, this step represents the pinnacle of scientific machine learning, where a simple, analytical, interpretable expression for physical parameters has been discovered by way of using a powerful, yet opaque, ML method such as PINN.

Finally, we have shown a simple way of performing automatic event identification, that is to identify time intervals when the underlying diffusive approximation is not valid (Figure 16). This can be due to a number of physical effects, including non-resonant interactions (Camporeale, 2015a; Camporeale & Zimbardo, 2015), large-amplitude waves (Bortnik et al., 2008b), pitch-angle and energy scattering (Tu et al., 2013), and others. Interestingly, some of the identified events (reported in Table 2) have been well studied in the literature, while others were not and thus deserve further investigation.

Future steps include extending the present study to a range of first and second adiabatic invariants, and eventually to the less approximated diffusion equation in energy and pitch-angle (requiring the specification of a diffusion tensor that includes cross terms, thus increasing the dimensionality of the problem, see, e.g. (Albert & Young, 2005; Camporeale et al., 2013a, 2013b)), and the estimates of uncertainties associated either to the derived coefficients, or directly to PSD solution of the Fokker-Planck equation (Camporeale & Carè, 2021; Chen et al., 2020).

Acknowledgments

This material is based upon work supported by the National Aeronautics and Space Administration under grant 80NSSC20K1580 "SWQU: Ensemble Learning for Accurate and

Reliable Uncertainty Quantification" issued through the Space Weather with Quantified Uncertainty (SWQU) program. A.D. was supported by NASA grant 80NSSC18K0663. J.B. acknowledges subgrant 1559841 to the University of California, Los Angeles, from the University of Colorado Boulder under NASA Prime Grant agreement 80NSSC20K1580, the Defense Advanced Research Projects Agency under U.S. Department of the Interior award D19AC00009, and NASA/SWO2R grant 80NSSC19K0239. E.C. is partially supported by NASA grants 80NSSC20K1275, 80NSSC21K1555.

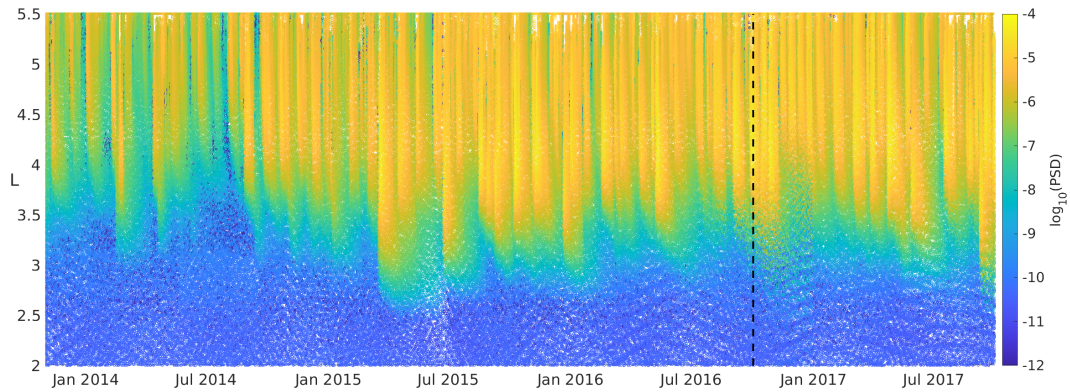

Figure 1. Phase Space Density of the whole dataset, in logarithmic scale, as function of L. The vertical dashed line divides the dataset into training (to the left) and test (to the right) sets.

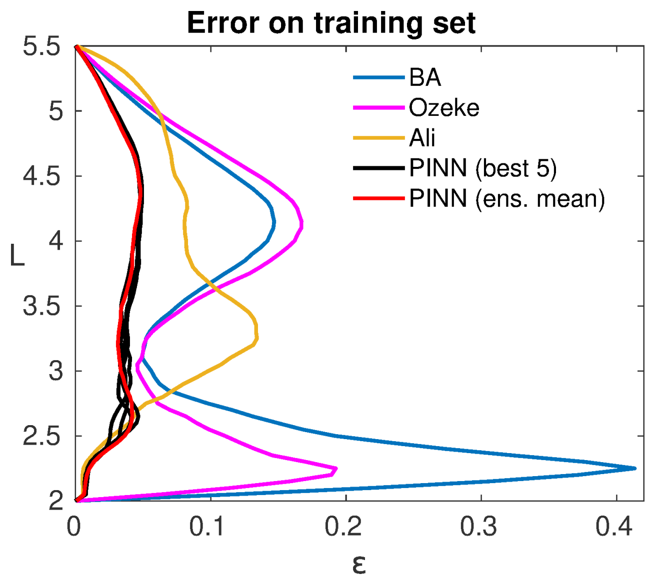

Figure 2. Error ε as a function of L, computed over the whole training set. Blue, magenta and yellow lines are for the baseline models BA, Ozeke et al., and Ali et al., respectively. Five black lines denote the top five solutions from the ensemble run, and the red line represents the solution obtained by using the mean of the top five diffusion and drift coefficients.

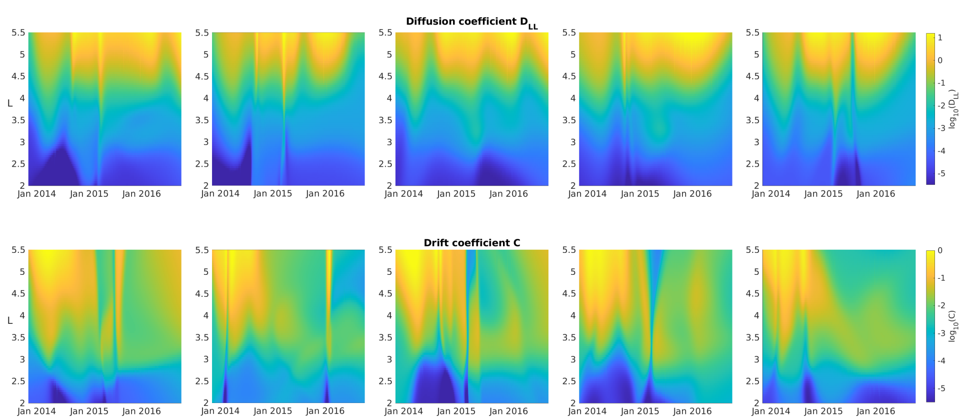

Figure 3. Top 5 diffusion coefficients (top) and corresponding drift coefficients (bottom), in logarithmic scale.

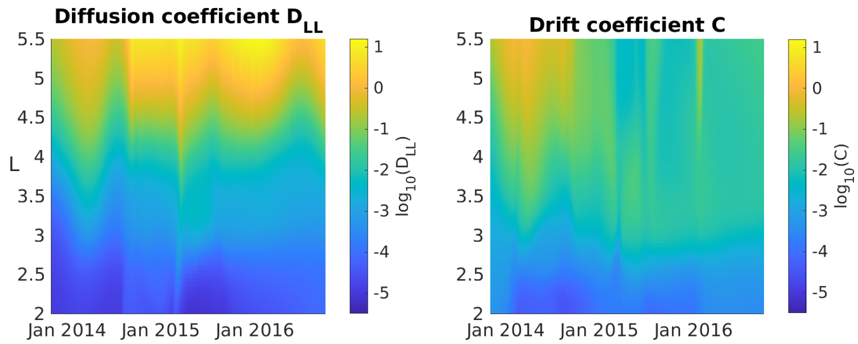

Figure 4. Diffusion (left) and drift (right) coefficients obtained by averaging the top 5 solutions shown in Figure 3

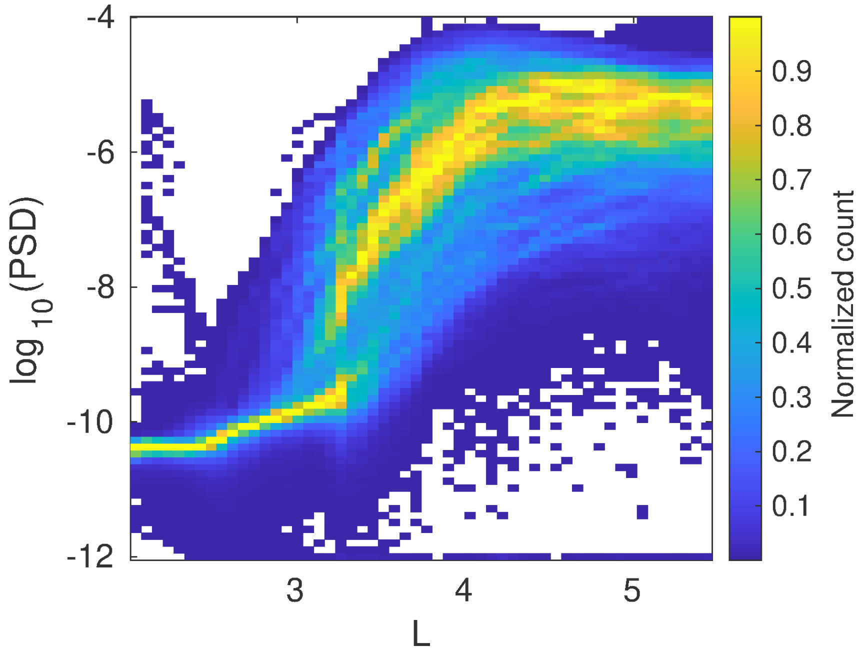

Figure 5. Distribution of the Phase Space Density f as a function of L. The heat map shows the counts in each bin, normalized to the largest number for a constant L. Here and in following Figures, the statistics is computed on about 25,000 times instances, spanning 3 years of data (01-Nov-2013 to 30-Oct-2016).

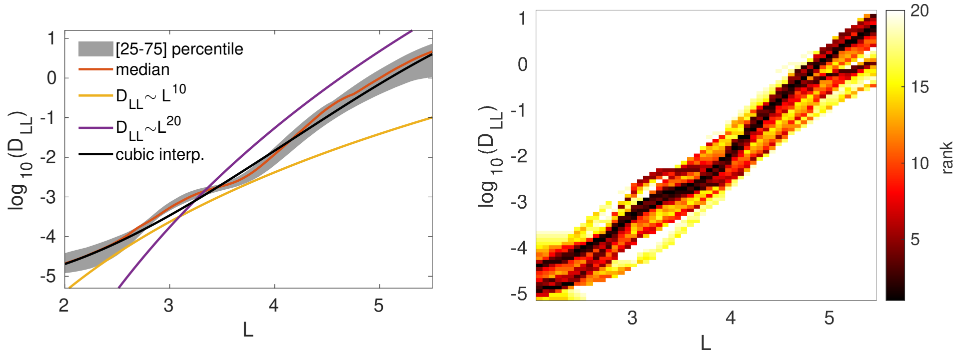

Figure 6. Distribution (left) and rank distribution (right) of the diffusion coefficients D_{LL} as function of L-shell. Left: The gray area represents the interval between the 25th and 75th percentile (for a given L-shell), and the orange line denotes the median. The yellow and magenta lines are shown as a reference for L^{10} and L^{20}, respectively. The black line is a cubic interpolation fit. Right: Dark colors indicate top ranks, and white indicates a rank equal or larger than 20. The ranking is performed by sorting the number of counts in each bin, at a constant L.

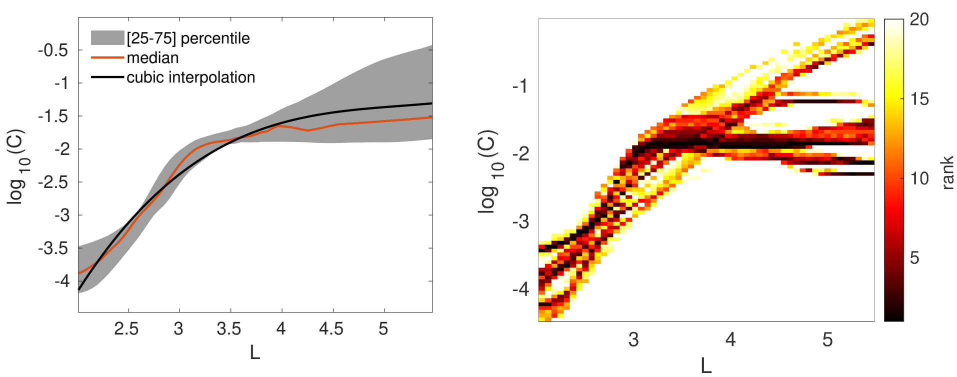

Figure 7. Distribution (left) and rank distribution (right) of the drift coefficient C as function of L. Left: The gray area represents the interval between the 25th and 75th percentile (for a given L), and the orange line denotes the median. The black line is a cubic interpolation fit. Right: dark colors indicate top ranks, and white indicates a rank equal or larger than 20. The ranking is performed by sorting the number of counts in each bin, at a constant L.

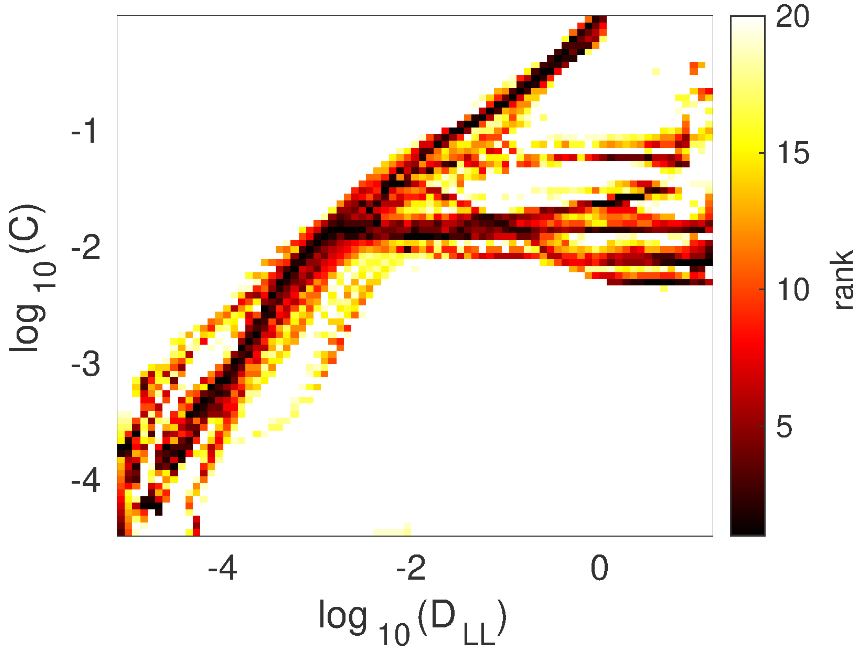

Figure 8. Ranked joint distribution of D_{LL} and C. Dark colors indicate top ranks, and white indicates a rank equal or larger than 20. The ranking is performed by sorting the number of counts in each bin, at a constant L.

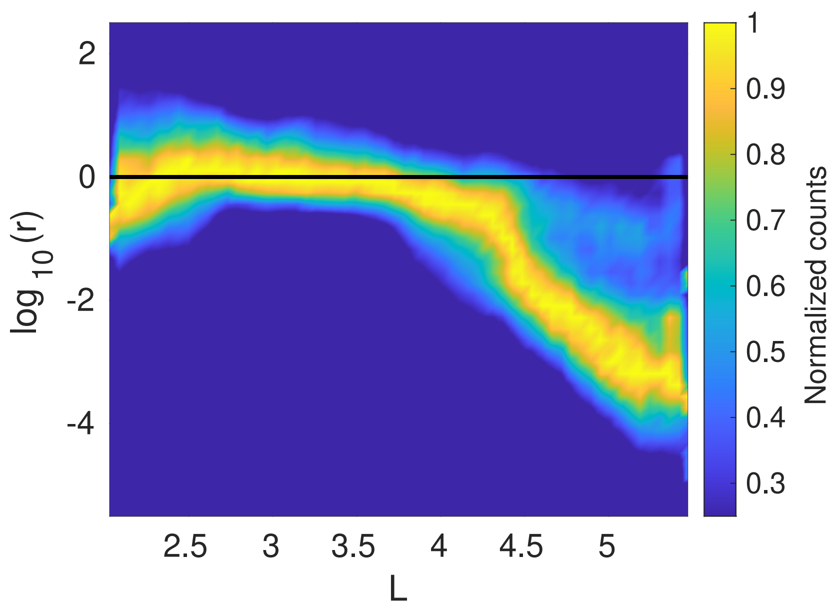

Figure 9. Distribution of r (logarithmic scale) as a function of L. The number of counts is normalized, for each value of L, to its maximum value. The black solid line denotes $r = 1$, that is exact balance between the drift and diffusion terms.

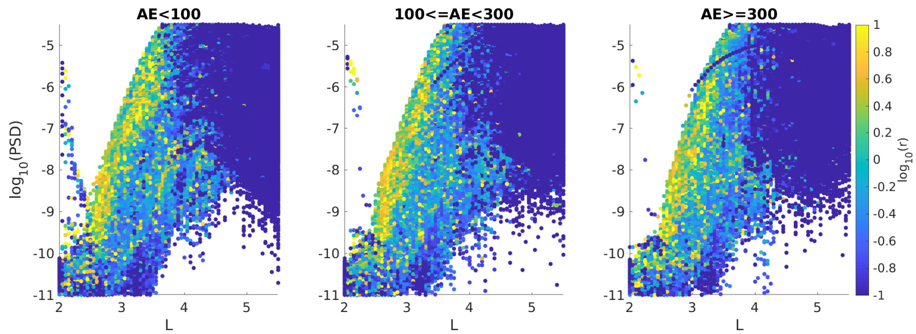

Figure 10. Distribution of r (logarithmic scale) as a function of L and $\log_{10}(PSD)$ for three geomagnetic levels (left panel: $AE < 100$, middle panel: $100 \leq AE < 300$, right panel: $AE >= 300$)

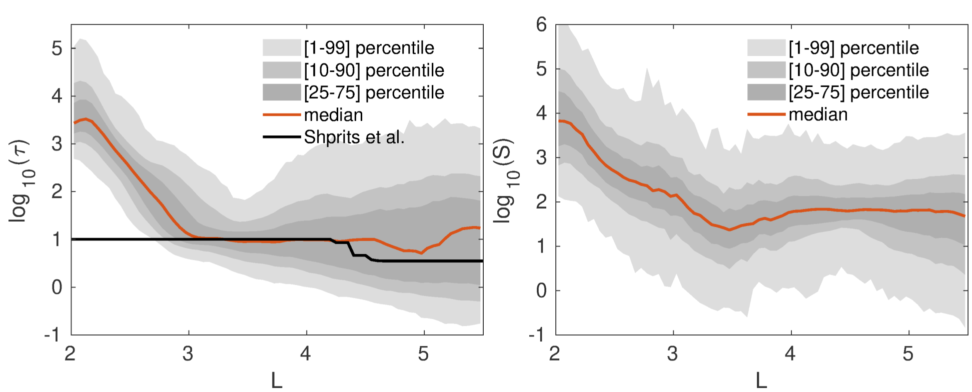

Figure 11. Distribution of the loss term τ (left) and S (right) as a function of L. The gray areas denote different percentiles range and the orange line represents the median value at a given L. Vertical axis in logarithmic scale.

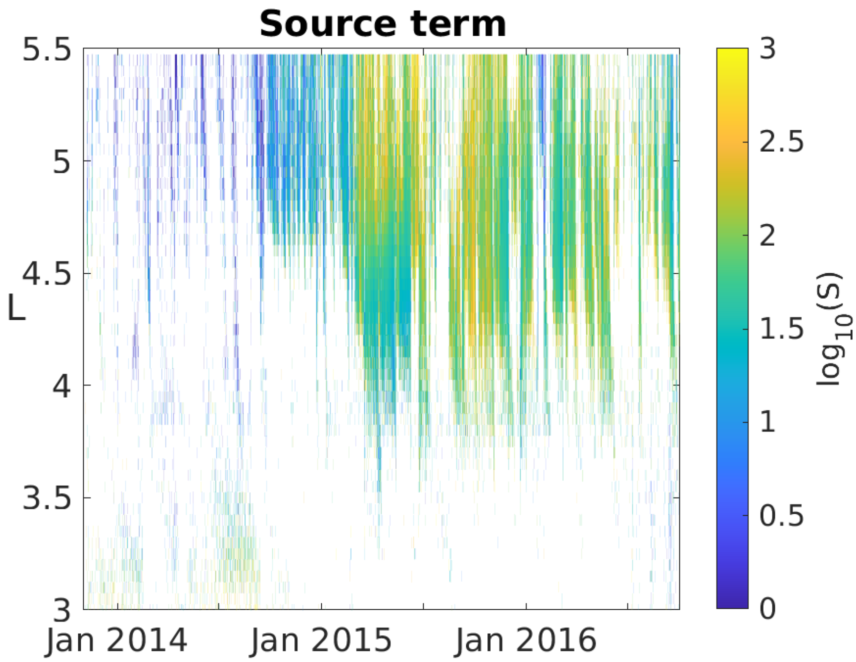

Figure 12. Heat map of the effective source term $\log_{10} S$ in time and L over the whole training set. The colors are saturated at 10^3 and white areas denote regions where there is no source, but a loss term.

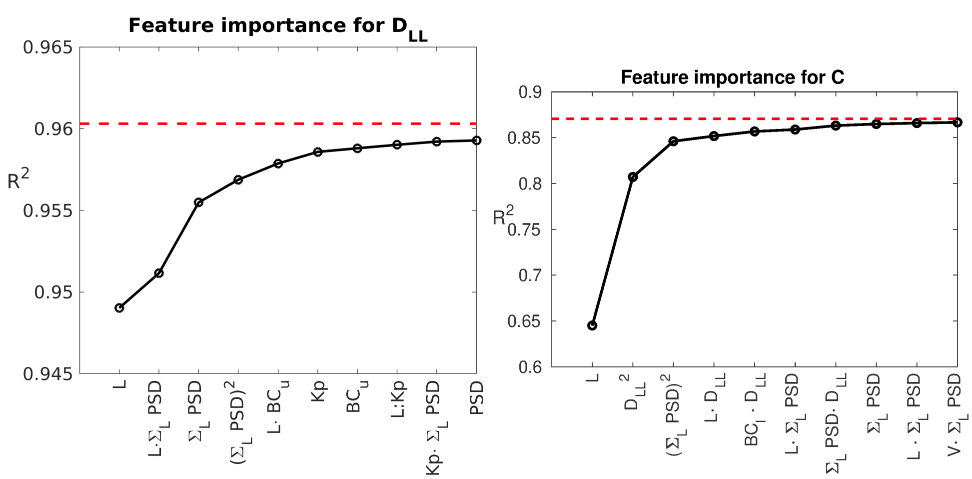

Figure 13. Backward elimination results for D_{LL} (left) and C (right). Each symbol denotes the coefficient of determination R^2 of a linear model that uses only the corresponding feature, in addition to all features shown on its left. The dashed red line represents the upper limit, obtained when all the features are taken into account in a generalized linear model (78 terms in total for D_{LL} and 98 for C). The meaning of each feature is explained in Table 1.

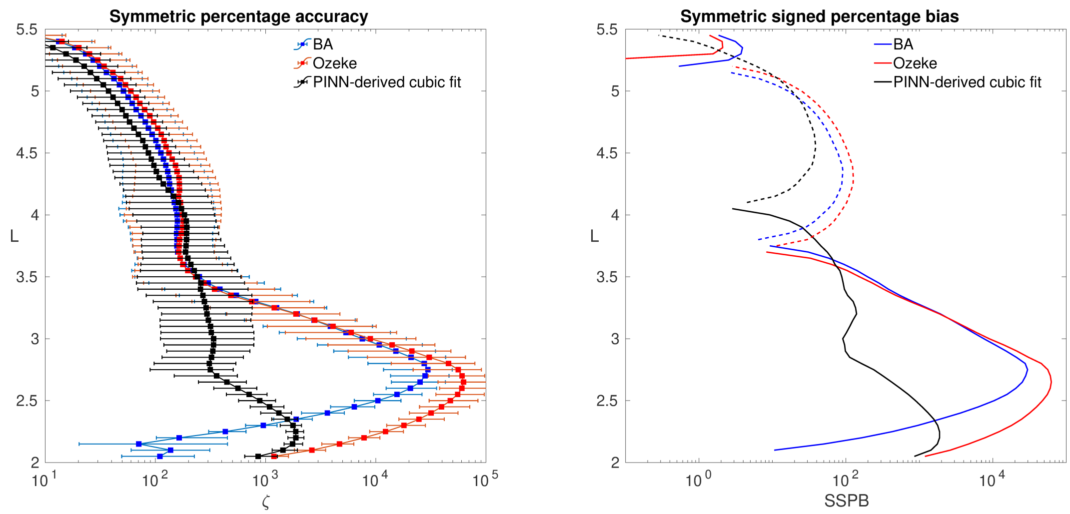

Figure 14. Percentage symmetric accuracy ζ (Eq. 7) (left) and symmetric signed percentage bias SSPB (Eq. 8)(right) calculated over the whole test set (1 year of data), as a function of L. Blue and red lines denotes the BA and Ozeke et al. baseline models, respectively, while the cubic parameterization in Eqs. (10) is shown in black. In the left panel, the solid squares denote the median values ζ_{50} and the error bars are calculated as the spread between ζ_{25} and ζ_{75}. In the right panel, positive values are in solid and negative values in dashed lines.

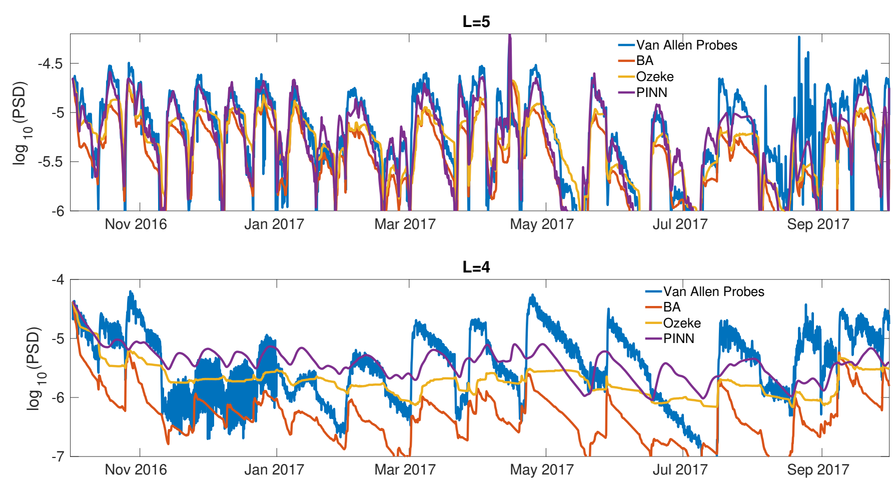

Figure 15. Phase Space Density (PSD) resulting from running the forward model with different coefficient parameterization, for the while test set. Red, yellow and purple lines denote the BA, Ozeke et al. and PINN-derived cubic parameterizations, respectively. The Van Allen Probes data is represented in blue. The vertical axis is in logarithmic scale.

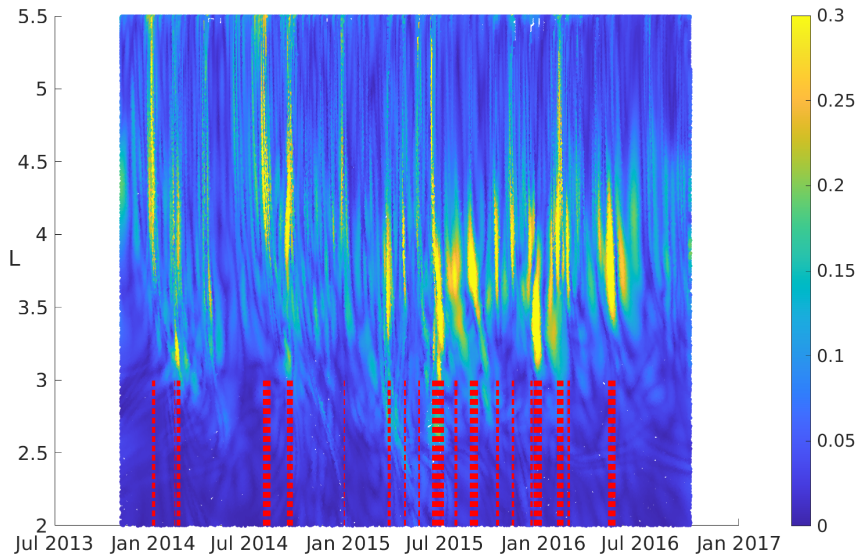

Figure 16. Heat map of the residual of Eq.(4), normalized on its maximum value, over the training set. The red dashed lines denotes time at which the value of the residual is in the 99 percentile of its distribution.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265–283).

Albert, J., & Young, S. (2005). Multidimensional quasi-linear diffusion of radiation belt electrons. *Geophysical Research Letters*, *32*(14).

Ali, A. F., Malaspina, D. M., Elkington, S. R., Jaynes, A. N., Chan, A. A., Wygant, J., & Kletzing, C. A. (2016). Electric and magnetic radial diffusion coefficients using the van allen probes data. *Journal of Geophysical Research: Space Physics*, *121*(10), 9586–9607.

Allanson, O., Elsden, T., Watt, C., & Neukirch, T. (2022). Weak turbulence and quasilinear diffusion for relativistic wave-particle interactions via a markov approach. *Frontiers in Astronomy and Space Sciences*.

Alves, L., Da Silva, L., Souza, V., Sibeck, D., Jauer, P., Vieira, L., ... others (2016). Outer radiation belt dropout dynamics following the arrival of two interplanetary coronal mass ejections. *Geophysical Research Letters*, *43*(3), 978–987.

Baker, D. N., Jaynes, A., Kanekal, S., Foster, J., Erickson, P., Fennell, J., ... others (2016). Highly relativistic radiation belt electron acceleration, transport, and loss: Large solar storm events of march and june 2015. *Journal of Geophysical Research: Space Physics*, *121*(7), 6647–6660.

Berg, J., & Nyström, K. (2019). Data-driven discovery of pdes in complex datasets. *Journal of Computational Physics*, *384*, 239–252.

Blake, J., Carranza, P., Claudepierre, S., Clemmons, J., Crain, W., Dotan, Y., ... others (2013). The magnetic electron ion spectrometer (mageis) instruments aboard the radiation belt storm probes (rbsp) spacecraft. In *The van allen probes mission* (pp. 383–421). Springer.

Bortnik, J., & Camporeale, E. (2021). Ten ways to apply machine learning in the earth and space sciences. In *Eos, 102.* doi: https://doi.org/10.1029/2021EO160257

Bortnik, J., Thorne, R., & Inan, U. S. (2008a). Nonlinear interaction of energetic electrons with large amplitude chorus. *Geophysical Research Letters*, *35*(21).

Bortnik, J., Thorne, R., & Inan, U. S. (2008b). Nonlinear interaction of energetic electrons with large amplitude chorus. *Geophysical Research Letters*, *35*(21).

Boullé, N., Earls, C. J., & Townsend, A. (2021). Data-driven discovery of physical laws with human-understandable deep learning. *arXiv preprint arXiv:2105.00266*.

Boyd, A. J., Turner, D., Reeves, G. D., Spence, H., Baker, D., & Blake, J. (2018a). What causes radiation belt enhancements: A survey of the van allen probes era. *Geophysical Research Letters*, *45*(11), 5253–5259.

Boyd, A. J., Turner, D., Reeves, G. D., Spence, H., Baker, D., & Blake, J. (2018b). What causes radiation belt enhancements: A survey of the van allen probes era. *Geophysical Research Letters*, *45*(11), 5253–5259.

Brautigam, D., & Albert, J. (2000). Radial diffusion analysis of outer radiation belt electrons during the october 9, 1990, magnetic storm. *Journal of Geophysical Research: Space Physics*, *105*(A1), 291–309.

Camporeale, E. (2015a). Resonant and nonresonant whistlers-particle interaction in the radiation belts. *Geophysical Research Letters*, *42*(9), 3114–3121.

Camporeale, E. (2015b). Resonant and nonresonant whistlers-particle interaction in the radiation belts. *Geophysical Research Letters*, *42*(9), 3114–3121.

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

Camporeale, E., & Carè, A. (2021). Accrue: Accurate and reliable uncertainty estimate in deterministic models. *International Journal for Uncertainty Quantification*, *11*(4).

Camporeale, E., Delzanno, G., Zaharia, S., & Koller, J. (2013a). On the numerical simulation of particle dynamics in the radiation belt: 1. implicit and semi-implicit schemes. *Journal of Geophysical Research: Space Physics*, *118*(6), 3463–3475.

Camporeale, E., Delzanno, G., Zaharia, S., & Koller, J. (2013b). On the numerical simulation of particle dynamics in the radiation belt: 2. procedure based on the diagonalization of the diffusion tensor. *Journal of Geophysical Research: Space Physics*, *118*(6), 3476–3484.

Camporeale, E., & Zimbardo, G. (2015). Wave-particle interactions with parallel whistler waves: Nonlinear and time-dependent effects revealed by particle-in-cell simulations. *Physics of Plasmas*, *22*(9), 092104.

Carpenter, D., & Anderson, R. (1992). An isee/whistler model of equatorial electron density in the magnetosphere. *Journal of Geophysical Research: Space Physics*, *97*(A2), 1097–1108.

Chandrasekhar, S. (1943). Stochastic problems in physics and astronomy. *Reviews of modern physics*, *15*(1), 1.

Chen, X., Yang, L., Duan, J., & Karniadakis, G. E. (2020). Solving inverse stochastic problems from discrete particle observations using the fokker-planck equation and physics-informed neural networks. *arXiv preprint arXiv:2008.10653*.

Chu, X., Bortnik, J., Li, W., Ma, Q., Angelopoulos, V., & Thorne, R. (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.

Dimitrakoudis, S., Mann, I. R., Balasis, G., Papadimitriou, C., Anastasiadis, A., & Daglis, I. A. (2015). Accurately specifying storm-time ulf wave radial diffusion in the radiation belts. *Geophysical Research Letters*, *42*(14), 5711–5718.

Drozdov, A., Allison, H. J., Shprits, Y., Elkington, S., & Aseev, N. (2021). A comparison of radial diffusion coefficients in 1-d and 3-d long-term radiation belt simulations. *Journal of Geophysical Research: Space Physics*, *126*(8), e2020JA028707.

Drozdov, A., Shprits, Y., Aseev, N., Kellerman, A. C., & Reeves, G. D. (2017). Dependence of radiation belt simulations to assumed radial diffusion rates tested for two empirical models of radial transport. *Space Weather*, *15*(1), 150–162.

Drozdov, A., Usanova, M., Hudson, M., Allison, H. J., & Shprits, Y. (2020). The role of hiss, chorus, and emic waves in the modeling of the dynamics of the multi-mev radiation belt electrons. *Journal of Geophysical Research: Space Physics*, *125*(9), e2020JA028282.

Fälthammar, C.-G. (1966). On the transport of trapped particles in the outer magnetosphere. *Journal of Geophysical Research*, *71*(5), 1487–1491.

Fox, N., & Burch, J. L. (2014). *The van allen probes mission.* Springer Science & Business Media.

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media.

Gu, X., Shprits, Y. Y., & Ni, B. (2012). Parameterized lifetime of radiation belt electrons interacting with lower-band and upper-band oblique chorus waves. *Geophysical Research Letters*, *39* (15).

Guo, D., Fu, S., Xiang, Z., Ni, B., Guo, Y., Feng, M., ... others (2021). Prediction of dynamic plasmapause location using a neural network. *Space Weather*, *19* (5), e2020SW002622.

Horne, R. B. (2007). Acceleration of killer electrons. *Nature Physics*, *3* (9), 590–591.

Horne, R. B., Thorne, R. M., Shprits, Y. Y., Meredith, N. P., Glauert, S. A., Smith, A. J., ... others (2005). Wave acceleration of electrons in the van allen radiation belts. *Nature*, *437* (7056), 227–230.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2* (5), 359–366.

Hudson, M. K., Elkington, S. R., Li, Z., & Patel, M. (2020). Drift echoes and flux oscillations: A signature of prompt and diffusive changes in the radiation belts. *Journal of Atmospheric and Solar-Terrestrial Physics*, *207*, 105332.

Jaynes, A. N., Baker, D. N., Singer, H. J., Rodriguez, J. V., Loto'aniu, T., Ali, A., ... others (2015a). Source and seed populations for relativistic electrons: Their roles in radiation belt changes. *Journal of Geophysical Research: Space Physics*, *120* (9), 7240–7254.

Jaynes, A. N., Baker, D. N., Singer, H. J., Rodriguez, J. V., Loto'aniu, T., Ali, A., ... others (2015b). Source and seed populations for relativistic electrons: Their roles in radiation belt changes. *Journal of Geophysical Research: Space Physics*, *120* (9), 7240–7254.

Kennel, C., & Engelmann, F. (1966). Velocity space diffusion from weak plasma turbulence in a magnetic field. *The Physics of Fluids*, *9* (12), 2377–2388.

Kilpua, E., Hietala, H., Turner, D., Koskinen, H., Pulkkinen, T. I., Rodriguez, J., ... Spence, H. E. (2015). Unraveling the drivers of the storm time radiation belt response. *Geophysical Research Letters*, *42* (9), 3076–3084.

Kilpua, E., Turner, D., Jaynes, A., Hietala, H., Koskinen, H., Osmane, A., ... others (2019). Outer van allen radiation belt response to interacting interplanetary coronal mass ejections. *Journal of Geophysical Research: Space Physics*, *124* (3), 1927–1947.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv* *preprint arXiv:1412.6980*.

Lejosne, S. (2019). Analytic expressions for radial diffusion. *Journal of Geophysical Research: Space Physics*, *124*(6), 4278–4294.

Lejosne, S., & Kollmann, P. (2020). Radiation belt radial diffusion at earth and beyond. *Space Science Reviews*, *216*(1), 1–78.

Lemons, D. S. (2012). Pitch angle scattering of relativistic electrons from stationary magnetic waves: Continuous markov process and quasilinear theory. *Physics of Plasmas*, *19*(1), 012306.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, *50*(6), 94.

Li, L.-F., Tu, W., Dai, L., Tang, B.-B., Wang, C., Barani, M., … Burch, J. (2020). Quantifying event-specific radial diffusion coefficients of radiation belt electrons with the ppmlr-mhd simulation. *Journal of Geophysical Research: Space Physics*, *125*(5).

Li, W., & Hudson, M. (2019). Earth's van allen radiation belts: From discovery to the van allen probes era. *Journal of Geophysical Research: Space Physics*, *124*(11), 8319–8351.

Li, Z., Elkington, S., Hudson, M., Patel, M., Boyd, A., & Wygant, J. (2021). Modeling advective transport of radiation belt electrons. *Journal of Atmospheric and Solar-Terrestrial Physics*, *214*, 105509.

Liemohn, M. W., Shane, A. D., Azari, A. R., Petersen, A. K., Swiger, B. M., & Mukhopadhyay, A. (2021). Rmse is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, *218*, 105624.

Liu, W., Tu, W., Li, X., Sarris, T., Khotyaintsev, Y., Fu, H., … Shi, Q. (2016). On the calculation of electric diffusion coefficient of radiation belt electrons with in situ electric field measurements by themis. *Geophysical Research Letters*, *43*(3), 1023–1030.

Long, Z., Lu, Y., Ma, X., & Dong, B. (2018). Pde-net: Learning pdes from data. In *International conference on machine learning* (pp. 3208–3216).

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

Ma, X., Xiang, Z., Ni, B., Fu, S., Cao, X., Hua, M., ... others (2020). On the loss mechanisms of radiation belt electron dropouts during the 12 september 2014 geomagnetic storm. *Earth and Planetary Physics*, *4*(6), 598–610.

Malaspina, D. M., Zhu, H., & Drozdov, A. Y. (2020). A wave model and diffusion coefficients for plasmaspheric hiss parameterized by plasmapause location. *Journal of Geophysical Research: Space Physics*, *125*(2), e2019JA027415.

Morley, S. K. (2016). Alternatives to accuracy and bias metrics based on percentage errors for radiation belt modeling applications.

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

Moya, P. S., Pinto, V. A., Sibeck, D. G., Kanekal, S. G., & Baker, D. N. (2017). On the effect of geomagnetic storms on relativistic electrons in the outer radiation belt: Van allen probes observations. *Journal of Geophysical Research: Space Physics*, *122*(11), 11–100.

Newell, P., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research: Space Physics*, *112*(A1).

Orlova, K., Shprits, Y., & 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.

Ozeke, L. G., Mann, I. R., Murphy, K. R., Jonathan Rae, I., & Milling, D. K. (2014). Analytic expressions for ulf wave radiation belt radial diffusion coefficients. *Journal of Geophysical Research: Space Physics*, *119*(3), 1587–1605.

Ozeke, L. G., Mann, I. R., Murphy, K. R., Rae, I. J., Milling, D. K., Elkington, S. R., ... Singer, H. J. (2012). Ulf wave derived radiation belt radial diffusion coefficients. *Journal of Geophysical Research: Space Physics*, *117*(A4).

Ozeke, L. G., Mann, I. R., Murphy, K. R., Sibeck, D. G., & Baker, D. N. (2017). Ultra-relativistic radiation belt extinction and ulf wave radial diffusion: Modeling the september 2014 extended dropout event. *Geophysical Research Letters*, *44*(6), 2624–2633.

Raissi, M. (2018). Deep hidden physics models: Deep learning of nonlinear partial differential equations. *The Journal of Machine Learning Research*, *19*(1), 932–

955.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707.

Reeves, G. D., Vandegriff, E. M., Niehof, J. T., Morley, S. K., Cunningham, G. S., Henderson, M. G., & Larsen, B. A. (2020). Defining radiation belt enhancement events based on probability distributions. *Space Weather*, *18*(8), e2020SW002528.

Ripoll, J.-F., Reeves, G. D., Cunningham, G. S., Loridan, V., Denton, M., Santolík, O., ... others (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.

Roederer, J. G., & Zhang, H. (2016). *Dynamics of magnetically trapped particles.* Springer.

Rostoker, G. (1972). Geomagnetic indices. *Reviews of Geophysics*, *10*(4), 935–950.

Rudy, S. H., Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2017). Data-driven discovery of partial differential equations. *Science Advances*, *3*(4), e1602614.

Sarma, R., Chandorkar, M., Zhelavskaya, I., Shprits, Y., Drozdov, A., & Camporeale, E. (2020). Bayesian inference of quasi-linear radial diffusion parameters using van allen probes. *Journal of Geophysical Research: Space Physics*, *125*(5), e2019JA027618.

Schulz, M., & Lanzerotti, L. J. (2012). *Particle diffusion in the radiation belts* (Vol. 7). Springer Science & Business Media.

Shen, X.-C., Hudson, M. K., Jaynes, A. N., Shi, Q., Tian, A., Claudepierre, S. G., ... Sun, W.-J. (2017). Statistical study of the storm time radiation belt evolution during van allen probes era: Cme-versus cir-driven storms. *Journal of Geophysical Research: Space Physics*, *122*(8), 8327–8339.

Shprits, Y., Thorne, R., Reeves, G., & Friedel, R. (2005). Radial diffusion modeling with empirical lifetimes: Comparison with crres observations. In *Annales geophysicae* (Vol. 23, pp. 1467–1471).

Shprits, Y. Y., Subbotin, D., & Ni, B. (2009). Evolution of electron fluxes in the outer radiation belt computed with the verb code. *Journal of Geophysical Re-*

search: Space Physics, *114*(A11).

Sotnikov, N., Antonova, E., Riazantseva, M., Ovchinnikov, I., Rubinstein, I., Barinova, V., et al. (2019). Spectra and pitch-angular distributions of relativistic electrons near the outer radiation belt maximum during the magnetic storm of december 19–22, 2015. *Geomagnetism and Aeronomy*, *59*(6), 651–659.

Summers, D., Thorne, R. M., & Xiao, F. (1998). Relativistic theory of wave-particle resonant diffusion with application to electron acceleration in the magnetosphere. *Journal of Geophysical Research: Space Physics*, *103*(A9), 20487–20500.

Thorne, R. M. (2010). Radiation belt dynamics: The importance of wave-particle interactions. *Geophysical Research Letters*, *37*(22).

Tsyganenko, N. A., & Sitnov, M. I. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *J. Geophys. Res.*, *110*(A3), 7737. doi: 10.1029/2004JA010798

Tu, W., Cunningham, G., Chen, Y., Henderson, M., Camporeale, E., & 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., Elkington, S. R., Li, X., Liu, W., & Bonnell, J. (2012). Quantifying radial diffusion coefficients of radiation belt electrons based on global mhd simulation and spacecraft measurements. *Journal of Geophysical Research: Space Physics*, *117*(A10).

Turner, D. L., O'Brien, T., Fennell, J., Claudepierre, S., Blake, J., Jaynes, A., … others (2017). Investigating the source of near-relativistic and relativistic electrons in earth's inner radiation belt. *Journal of Geophysical Research: Space Physics*, *122*(1), 695–710.

Udrescu, S.-M., & Tegmark, M. (2020). Ai feynman: A physics-inspired method for symbolic regression. *Science Advances*, *6*(16), eaay2631.

Vlasova, N., Kalegaev, V., Nazarkov, I., & Prost, A. (2020). Magnetic field variations and dynamics of the outer electron radiation belt of the earth's magnetosphere in february 2014. *Geomagnetism and Aeronomy*, *60*, 7–19.

Wang, D., Shprits, Y. Y., Zhelavskaya, I. S., Effenberger, F., Castillo, A. M., Drozdov, A. Y., … Cervantes, S. (2020). The effect of plasma boundaries on the dynamic evolution of relativistic radiation belt electrons. *Journal of Geophysical Research: Space Physics*, *125*(5), e2019JA027422.

Welling, D., Koller, J., & Camporeale, E. (2012). Verification of spacepy's radial diffusion radiation belt model. *Geoscientific Model Development*, *5*(2), 277–287.

Wing, S., Johnson, J. R., Camporeale, E., & 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.

Wu, H., Chen, T., Kalegaev, V., Panasyuk, M., Vlasova, N., Duan, S., … Wang, C. (2020). Long-term dropout of relativistic electrons in the outer radiation belt during two sequential geomagnetic storms. *Journal of Geophysical Research: Space Physics*, *125*(10), e2020JA028098.

Xiao, F., Su, Z., Zheng, H., & Wang, S. (2010). Three-dimensional simulations of outer radiation belt electron dynamics including cross-diffusion terms. *Journal of Geophysical Research: Space Physics*, *115*(A5).

Xu, H., Chang, H., & Zhang, D. (2019). Dl-pde: Deep-learning based data-driven discovery of partial differential equations from discrete and noisy data. *arXiv preprint arXiv:1908.04463*.

Yu, X., Yuan, Z., Ouyang, Z., & Yao, F. (2021). Effects of the plasmapause on the radial propagation of fast magnetosonic waves: An analytical approach. *Journal of Geophysical Research: Space Physics*, *126*(3), e2020JA028330.

Zhang, S., & Lin, G. (2018). Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *474*(2217), 20180305.

Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, *23*(4), 550–560.