Key Points:

- Developed a surrogate model EPREM-S yielding the same output as EPREM, with low predictive uncertainty while thousands of times faster
- A proof-of-concept study of EPREM-S and EPREM revealed that the surrogate models achieved generalization beyond the training set
- In an event-analysis, as a Bayesian inference problem, all input parameters' ground truth values were recovered as mode of the posterior

Correspondence to:
B. Poduval,
bala.poduval@unh.edu

Author Contributions:
Conceptualization: Bala Poduval
Data curation: Bala Poduval
Formal analysis: Atilim Guneş Baydin, Bala Poduval
Funding acquisition: Bala Poduval
Investigation: Bala Poduval
Methodology: Atilim Guneş Baydin, Bala Poduval
Project Administration: Bala Poduval
Resources: Bala Poduval
Software: Atilim Guneş Baydin, Bala Poduval, Nathan A. Schwadron
Supervision: Bala Poduval
Validation: Atilim Guneş Baydin, Bala Poduval
Visualization: Bala Poduval
Writing – original draft: Atilim Guneş Baydin, Bala Poduval
Writing – review & editing: Bala Poduval

A Surrogate Model for Studying Solar Energetic Particle Transport and the Seed Population

Atilim Guneş Baydin[1], Bala Poduval[2] , and Nathan A. Schwadron[2]

[1]Department of Computer Science, University of Oxford, Oxford, UK, [2]Space Science Center, University of New Hampshire, Durham, NH, USA

Abstract The high energy particles originating from the Sun, known as solar energetic particles (SEPs), contribute significantly to the space radiation environment, posing serious threats to astronauts and scientific instruments on board spacecraft. The mechanism that accelerates the SEPs to the observed energy ranges, their transport in the inner heliosphere, and the influence of suprathermal seed particle spectrum are open questions in heliophysics. Accurate predictions of the occurrences of SEP events well in advance are necessary to mitigate their adverse effects but prediction based on first principle models still remains a challenge. In this scenario, adopting a machine learning approach to SEP modeling and prediction is desirable. However, the lack of a balanced database of SEP events restrains this approach. We addressed this limitation by generating large data sets of synthetic SEP events sampled from the physics-based model, Energetic Particle Radiation Environment Module (EPREM). Using this data, we developed neural networks-based surrogate models to study the seed population parameter space. Our models, EPREM-S, run thousands to millions of times faster (depending on computer hardware), making simulation-based inference workflows practicable in SEP studies while providing predictive uncertainty estimates using a deep ensemble approach.

Plain Language Summary Space weather refers to the spatio-temporal variations in the near-Earth environment constituted by the solar wind. These variations cause disruptions in the technological infrastructure on Earth and in space. Outside the Earth's protective envelope, spacecraft are exposed to a harsh radiation environment constituted by the galactic cosmic rays and solar energetic particles (SEPs), posing serious threats to the astronauts and the electronics on board. Mitigation of space radiation hazard includes shielding but radiation exceeding 100 MeV produces secondary penetrating particles and nuclear fragments in the shielding material. The alternate solution is to predict the occurrences of SEPs well in advance. Since predicting the arrival time and location of SEPs based on first principle models remains a challenge, adopting a machine learning approach to SEP modeling and prediction is desirable. However, the lack of a balanced database of SEP events restrains this approach. We addressed this limitation by generating synthetic SEP events sampled from the physics-based model Energetic Particle Radiation Environment Module and developed neural networks-based surrogate models to study the seed population parameter space. Our models run thousands to millions of times faster, making simulation-based inference workflows practicable in SEP studies while providing predictive uncertainty estimates using a deep ensemble approach.

1. Introduction

The near-Earth environment is constituted by the outward expanding solar corona known as the solar wind. Solar activities such as solar flares and coronal mass ejections (CMEs) trigger the spatio-temporal variabilities in the solar wind that lead to adversities in the *Space Weather*, a term referring to the changing conditions in the near-Earth space causing disturbances in the geomagnetic field which in turn cause disruptions and damages to our technological infrastructure on Earth and space (see Hundhausen, 1972; Richardson, 2018; Temmer, 2021; Vidotto, 2021, for reviews on solar wind and space weather). The solar wind is comprised of magnetized plasma and particles such as electrons, protons, He, O, N, Fe and other heavy elements with a wide range in energies and abundances. These particles change their number densities, energy levels and speed as the solar activity progresses, constituting the energetic particle space radiation environment. These variations or enhancements well above the energy levels and abundances in the ambient solar wind, often reaching relativistic speeds, are known as solar energetic particle (SEP) events or solar radiation storm (see Martucci et al., 2023; Reames, 2013, 2021, 2022; Schwadron et al., 2010, 2017, 2018, for comprehensive reviews and recent observations).
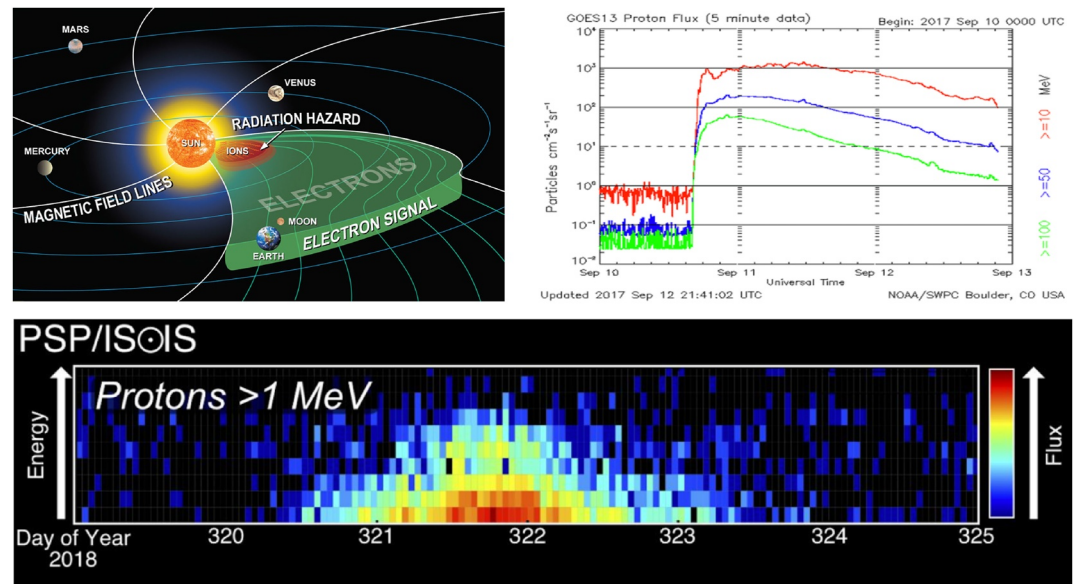
Figure 1. Illustration of the radiation environment (top left panel—(Posner et al., 2009)) and the SEP events observed by GOES-13 (top right panel—Courtsey: NOAA/SWPC, Boulder, CO) and PSP (bottom panel—(McComas et al., 2019)).

First observed in the 1940s as a ground level enhancement (GLE: Forbush, 1946), SEPs consist mainly of electrons and protons. SEPs can reach Earth (at a heliocentric distance of 150 million kilometers or 8 light minutes) in an hour or even less. Their energy levels range from several keVs (kilo electron volts) to a few GeVs (giga electron volts). Measurements of SEPs are available from space missions such as Parker Solar Probe (PSP), Solar Orbiter, Geostationary Operational Environmental Satellite (GOES), Advanced Composition Explorer (ACE), and Solar Terrestrial Relations Observatory (STEREO) (Kaiser et al., 2008; McComas et al., 2016; Rodríguez-Pacheco et al., 2020; Stone et al., 1998). Figure 1 depicts a graphical representation of the radiation environment (top left panel), GOES - 13 observation of SEPs associated with the 2017 September 10 solar flare (top right) and the observation of the 2018 November proton event by the Integrated Science Investigation of the Sun (IS\astrosunIS) instrument on board PSP (bottom panel).

The observed SEP events have been broadly classified into two categories, impulsive and gradual but recent works suggest there must be four different types based on particle abundances (Reames, 2013, 2021, 2022). The impulsive events are electron-rich with significant (~1,000 times) enhancements in ^{3}He/^{4}He and are associated with solar flares and Type III radio bursts. Gradual events, occurring in connection with CMEs and/or CME-driven shocks and Type II radio bursts (Cane & Lario, 2006; Reames, 2013, 2022; Schwadron et al., 2020; Winter & Ledbetter, 2015), are proton-rich with no significant ^{3}He/^{4}He enhancements. The observed characteristics and spatio-temporal profiles of SEPs are influenced by various physical phenomena at the solar sources including the suprathermal seed particle spectrum or *seed population* (the low-energy particles often produced in solar flares and accelerated by coronal mass ejection shocks), the heliospheric magnetic field configuration, and the particle acceleration mechanism (e.g., CME speed and energy, the microphysics such as turbulence in the local interplanetary medium). Further, the typical radial dependence in CME speeds leads to changes in the strength of the CME-driven shocks associated with gradual SEP events throughout the heliosphere (Brooks & Yardley, 2021; Reames, 2013, 2021, 2022; Schwadron et al., 2010, 2017, 2018, 2020; Winter & Ledbetter, 2015). Therefore, an in depth analysis of the SEP profiles along with observations of other solar and solar wind variabilities will provide information on their energy spectra, ionization states, elemental and isotope abundances and, most importantly, the plasma properties of these sources and the physical mechanisms of their release and acceleration.

The SEPs form a significant component of the radiation exposure risk for space exploration. They cause a wide range of disruptions and damages to our technological infrastructure both on Earth and the near-Earth environment. To mention a few, these include aircraft navigation systems, radio and wireless communications and the electronics on board spacecraft. In addition, they can damage the DNA of astronauts in space and pose a threat of radiation exposure to the crew and passengers on board high-latitude commercial airlines. Further,

SEPs are, possibly, the single major hurdle in the future lunar and interplanetary missions (e.g., Schwadron et al., 2010, 2017, 2018, and the references therein). Therefore, addressing the mitigation of space weather hazards due to SEPs is in the forefront of heliophysics research, both from operational and research perspectives. Providing adequate spacecraft shielding is one way to mitigate space radiation hazards but the secondary particles such as neutrons and nuclear fragments generated in the shielding material when the impinging radiation carries energies exceeding 100 MeVs, worsen the radiation hazard (see Schwadron et al., 2010, 2017, 2018, and the references therein). A more tangible solution is to accurately predict with sufficient lead time when and where these space weather events will occur. However, SEP prediction based on first principle models continues to be challenging and the role of the seed population in the acceleration and transport of SEPs is still an open question in heliophysics. In this scenario, adopting a machine learning (ML) approach to SEP modeling and prediction is desirable and we present in this article the results of an ML approach known as surrogate modeling or emulation where we developed a neural network (NN) model named as EPREM-S, a fast surrogate of the physics-based particle transport and acceleration model known as the Energetic Particle Radiation Environment Module (EPREM) developed by Schwadron et al. (2007, 2010).

This article is organized as follows: Section 2 describes the data, models and the methods used and Section 3 presents the results. We conclude this article with a discussion of the results and the methods adopted, along with the scope of the present work and our plans for future work using EPREM-S in Section 4.

2. Methods and Data

The main challenge in developing ML-based models for SEP prediction is the lack of a large enough and balanced data set of observed SEP events for training and validation of modern NN architectures. When the approach is formulated as a classification task (e.g., detection of SEP events vs. background), the problem is exacerbated even further because of class imbalance, implying that SEP events are rare in existing data. To circumvent this difficulty, we generated the first large data set of physically accurate synthetic SEP events using EPREM, a sub-module within the Earth-Moon-Mars Radiation Environment Module (EMMREM) framework, developed by Schwadron et al. (2007, 2010) for predicting the radiation exposure on Earth, Moon, Mars, and anywhere in the heliosphere between Earth and Mars caused by solar energetic particles and cosmic rays. The EPREM model, the synthetic data generation using EPREM and the development of the NN surrogate model are described in detail in the coming sections.

2.1. Energetic Particle Radiation Environment Model (EPREM)

EPREM is a kinetic transport code that solves for the propagation and acceleration of energetic particles in the evolving magnetic fields of the inner heliosphere (Kóta et al., 2005; K. Kozarev et al., 2010; K. A. Kozarev et al., 2010; Schwadron et al., 2007; Schwadron et al., 2010) in 3-dimension and is publicly available through NASA community Coordinated Modeling Center (CCMC). EPREM solves the focused transport equation in the Lagrangian frame of reference (i.e., comoving and field aligned), incorporating parallel transport of particles, cross-field diffusion and particle drifts. EPREM is designed to easily modify the energy range, step size and the boundary conditions in energy space, and is parallelized. EPREM has been validated using well-known events such as the Halloween event and using spacecraft data such as GOES and Ulysses (Schwadron et al., 2010, 2015, 2017, 2018). In addition, EPREM has been used in various contexts such as modeling radiation doses at 1 AU during an extreme SEP event (Schwadron et al., 2014) and solving for SEP distribution based on focussed transport equation (Schwadron et al., 2010); studying the time-dependent effects of SEP acceleration in the low corona during a CME evolution (K. A. Kozarev et al., 2010) and solving for pick up ion distributions (Hill et al., 2009), to mention a few.

The EPREM grid is situated on nested cubes whose surfaces are regularly subdivided into square arrays of square cells with grid nodes at their centers. The shell of grids at the inner boundary corotates with the Sun at each time step and the nested shells advance radially outward with the solar wind. The displacement of each node at each time step Δt is given by: $\Delta x = \Delta t \times V_x$, where V_x is the solar wind velocity and the nodes propagate at the speed of the solar wind (Schwadron et al., 2010). Due to the frozen-in nature of the coronal plasma, the magnetic field is carried along by the outward propagating solar wind giving rise to the interplanetary magnetic field, with an average orientation of ~45° to the radial direction at 1 AU (known as the Parker spiral; see Hundhausen, 1972,

for a review) caused by the solar rotation (the spirals in Figure 4 in Schwadron et al., 2010). *Observer nodes* are special class of nodes where the energetic particle distributions projected at a given observer such as Earth, Moon or spacecraft. The magnetic fields connected to these nodes are known as *observer connected field lines* (the thick red and black dashed line in Figure 4 in Schwadron et al., 2010). Given a source of particles such as SEPs or pickup ions, EPREM computes the distribution function as a function of space (heliocentric distance), time and energy ranges anywhere in the heliosphere at individual nodes advecting with the speed of solar wind, naturally tracing the Parker spirals.

2.2. Data Generation

The EPREM simulator consists of more than 10,000 lines of C++ code, and it uses a human-readable configuration file format describing the physical characteristics of SEP events including parameters defining the grid, simulation time, background solar wind, seed particle source function, and shock parameters.

For this study, we selected five core parameters of the seed function, namely, boundary function amplitude (Amplitude), energy spectrum power-law index γ (Gamma), radial scaling index β (Beta), boundary function cut-off energy, also called roll over energy or knee (Cut-off), and the mean free path (λ_0 or LAMO) as the input parameters that vary, that is, parameters which EPREM-S takes as input. EPREM utilizes a seed particle spectrum that is a function of energy and heliocentric distance from the Sun. The Cut-off energy or knee signifies the contribution of particles that are deeply penetrating and is of great significance in the prediction of radiation hazard in the context of manned and robotic space explorations beyond the protective envelope of terrestrial atmosphere and magnetosphere. We defined prior probability distributions over these five parameters using continuous uniform distributions $\mathcal{U}(0.01, 80)$, $\mathcal{U}(0, 2)$, $\mathcal{U}(0, 2)$, $\mathcal{U}(1, 10)$, $\mathcal{U}(10^{-4}, 1)$ respectively. The selection of the upper and lower limits of each of these five parameters were based on the numerous event analyses carried out by various groups that are available in the literature and our current understanding of SEPs and their origin, acceleration and transport in the heliosphere. The values in these ranges are expected to yield simulated events that are physically meaningful and close to the realistic scenario. However, these ranges are not conclusive or final but rather a first step toward the better understanding of the seed population parameter space. For the chosen ranges, we select uniform distributions because we would like EPREM-S to be trained over the whole parameter space in an unbiased manner. Future analyses using observed SEP events is likely to update the current ranges chosen and can also suggest more informative prior distributions over these parameters that reflect real SEP events.

The rest of EPREM parameters are fixed as follows. Though EPREM can be coupled with MHD models such as MAS and Enlil capable of simulating the global solar wind, we used the uncoupled version where a uniform solar wind is assumed throughout the simulation domain into which the shock is injected. For the spatial grid, we chose 1 row and 4 columns for the nested cubes so that there are 24 streams, each stream is a linked sequence of nodes representing the trajectory of fluid particles advecting with the solar wind. Along each stream, there were 288 nodes. The inner boundary of the simulation domain was set at ~22.0 $R_\odot$ and we selected a time step of 20 min for a 4 day simulation duration. For the background solar wind, through which the nodes are advected forming the Parker spiral (see Section 2.1), a uniform velocity of 450 km s^{-1} has been used, a reasonable value since the shock is injected near the equatorial region where the slow solar wind is mostly confined to except during solar maxima. The density was chosen to fall off as inverse square of radial distance from the Sun. Simulation has been done for 20 energy steps from 0.01 to 200 MeV/nucleon. A shock of width ~70° placed in the equatorial region, at ~60° from the face center, and a speed of 1,600 km s^{-1} was injected 0.25 days after the simulation started.

For data generation we implemented a Python wrapper that allows us to call EPREM as a simple function $\phi = \text{EPREM}(\psi)$, which takes the parameter vector ψ as a PyTorch tensor, writes the values in the corresponding fields of a config file generated on-the-fly, executes EPREM inside a Docker container, parses the computed values from EPREM output files, and returns the output fluxes ϕ in a PyTorch tensor. This allows us to use EPREM in a flexible way in our ML data generation, training, and inference setup in Python.

We generated 32,000 unique events using EPREM and the high-performance computing (HPC) resources at the San Diego Supercomputer Center (SDSC: part of the National Science Foundation's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support—ACCESS). Typically, EPREM takes tens of minutes to hours, depending on the specific input configuration, to simulate a single event which is considered computationally
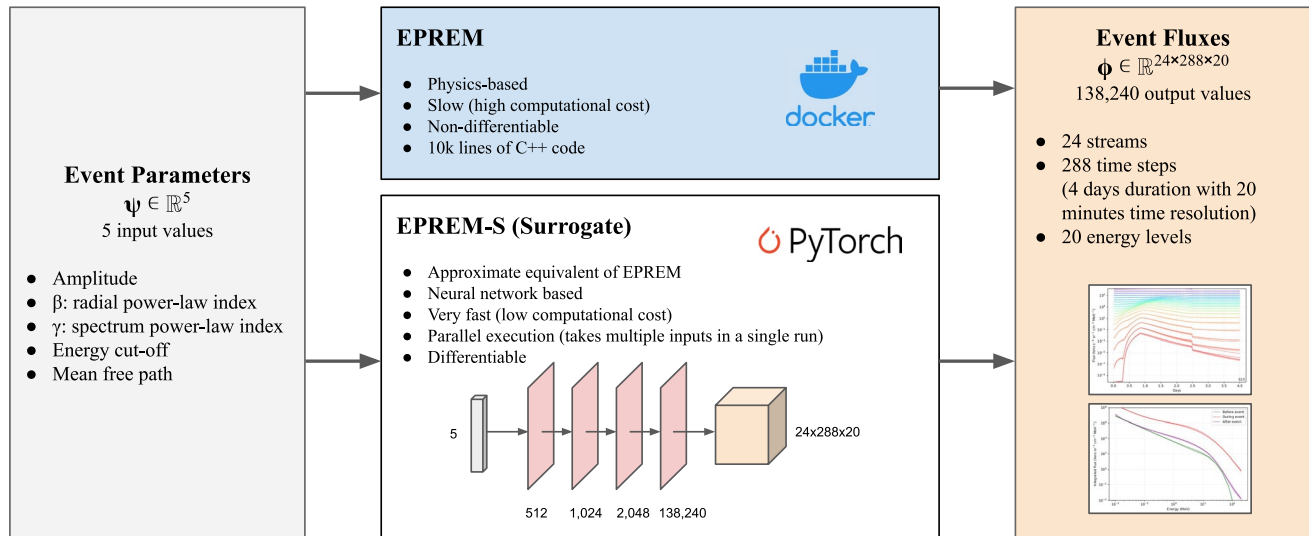
Figure 2. Overview of EPREM and EPREM-S models. Both map input parameters ψ to output fluxes ϕ. EPREM is a physics-based model for generating SEP events implemented in C++. EPREM-S is a neural network trained with a data set $\mathcal{D} = \{\psi_i, \phi_i = \text{EPREM}(\psi_i)\}_{i=1}^{N}$ obtained by running EPREM with inputs $\psi_i \sim p(\psi)$ sampled from a continuous uniform prior distribution.

expensive in the context of data generation sufficient for NN training. Using this data set, we developed and trained *EPREM-S*, a surrogate model several times faster than EPREM while producing the same output as EPREM, as described in the next section.

2.3. Surrogate Modeling

Surrogate modeling, also known as emulation, refers to the creation of fast and simple models that approximate the behavior of complex analytical models that are computationally expensive to evaluate (Queipo et al., 2005; Sobester et al., 2008). The low computation cost and differentiability (if available) of surrogate models enable tasks that might be infeasible using the original expensive model, including sensitivity analysis (e.g., using differentiability), optimization of parameters (e.g., using gradient descent), and uncertainty quantification (e.g., via efficient sampling). Surrogate models have been used in settings as diverse as particle physics (Shirobokov et al., 2020), exoplanet studies (Himes et al., 2022), probabilistic programming (Munk et al., 2022), and computer vision (Behl et al., 2020).

Recent advances in ML, especially in the field of NNs, make the defining and training of surrogate models relatively straightforward using off-the-shelf model components such as linear, recurrent, convolutional, or transformer modules available in frameworks such as PyTorch (Paszke et al., 2019) and TensorFlow (Abadi et al., 2016). Crucially, modern NN techniques require that a sufficiently large set of data representing the behavior of the original expensive model is available, usually in the form of input-output pairs. The generation of this training data is usually the most demanding and computationally costly part of a surrogate modeling approach, involving the construction of a code base to run the original simulator (usually in a distributed computing setting) with input parameters supplied via a sampling scheme.

2.4. Surrogate Training and Uncertainty Estimation

Figure 2 provides an overview of the EPREM and EPREM-S models, mapping SEP event parameters (inputs) ψ to event fluxes (outputs) ϕ. The code for the EPREM-S model is written in Python using the PyTorch framework (Paszke et al., 2019). After running hyperparameter tuning with a range of feed-forward and convolutional architectures, we base our results presented in this paper on a feed-forward NN with four hidden layers of sizes 512, 1,024, 2,048, and 138,240, and ReLU nonlinearities after each layer except the last one. This corresponds to a total number of 285,881,344 trainable parameters θ. The output of the last layer is reshaped into a cube with shape $24 \times 288 \times 20$, where the dimensions correspond to 24 streams, 288 time steps, and 20 energy levels. Given a data set of input–output pairs of

$$\mathcal{D} = \{\psi_i, \phi_i = \mathrm{EPREM}(\psi_i)\}_{i=1}^{N},$$

we train the surrogate model with a mean squared error loss

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left((\phi_i - \hat{\phi}_i)^\top (\phi_i - \hat{\phi}_i) \right),$$

where, $\hat{\phi}_i = S_\theta(\psi_i)$ are predictions of the surrogate S_θ parameterized by θ which we learn by minimizing $\mathcal{L}(\theta)$. For gradient-based optimization of $\mathcal{L}(\theta)$ we use the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 2×10^{-4}.

We applied data scaling to both the inputs ψ and prediction targets ϕ which we observed to significantly improve the results we obtained. We standardized the inputs to

$$\bar{\mu} = \frac{\psi - \mu_\psi}{\sigma_\psi}$$

where, μ_ψ and σ_ψ are the mean and standard deviation, respectively, of all ψ_i values in the training data. We applied thresholding and log-scaling to the prediction targets to obtain

$$\bar{\phi} = \log\left(\min\left(\max\left(\phi, 10^{-10}\right), 10^{10}\right)\right).$$

At inference time this scaling is undone after executing EPREM-S, giving flux values with the correct units (Figure 3).

In order to provide uncertainty quantification when running trained EPREM-S models, we used a deep ensemble (Lakshminarayanan et al., 2017) approach, where we worked with multiple independently trained EPREM-S instances. Following standard practice, we trained these model instances with the same training data, but using a different random number seed leading to different model weight initialization and course of stochastic optimization for each instance. Given the set of pretrained surrogate models S_i, $i = 1, \ldots, M$, and a new event parameter ψ, the mean and standard deviation of the flux predictions were estimated as

$$\mu_\phi = \frac{1}{M} \sum_i S_i(\psi)$$

and

$$\sigma_\phi = \sqrt{\frac{1}{M} \sum_i \left(S_i(\psi) - \mu_\phi\right)^2}.$$

In our experiments we used an ensemble size of $M = 5$.

3. Results

We trained an ensemble of five independently trained EPREM-S models where each model is a feed-forward NN with four layers and approximately 286 million learnable parameters. During each training we used a minibatch size of 64, a learning rate of 2×10^{-4}, and a weight decay (L2 regularization) coefficient of 10^{-6}. We used a random split of the data into a training set containing 90% of data and a hold-out validation set containing the rest. We ran the training for 10 epochs, and the lowest mean squared error (MSE) loss achieved is 0.07 for validation and 0.006 for training. We took the model with the lowest validation loss across the 10 epochs as the final model, to eliminate the possibility of producing a model that might overfit to the training data. Further details of the NN architecture and training are described in § 2.

Using the EPREM-S ensemble, we carried out a proof-of-concept study by generating several thousand SEP events for a range of values of physical parameters that are relevant to the seed population in the physics-based model EPREM. Figure 3 shows a comparison of the outputs of EPREM-S and EPREM for a sample of four events. Here, out of the 24 streams (see Section 2.2 for details of EPREM grid and other parameters), we identified stream 10 as the *observer connected stream* for an observer at 0.48 AU where the energetic particle distributions
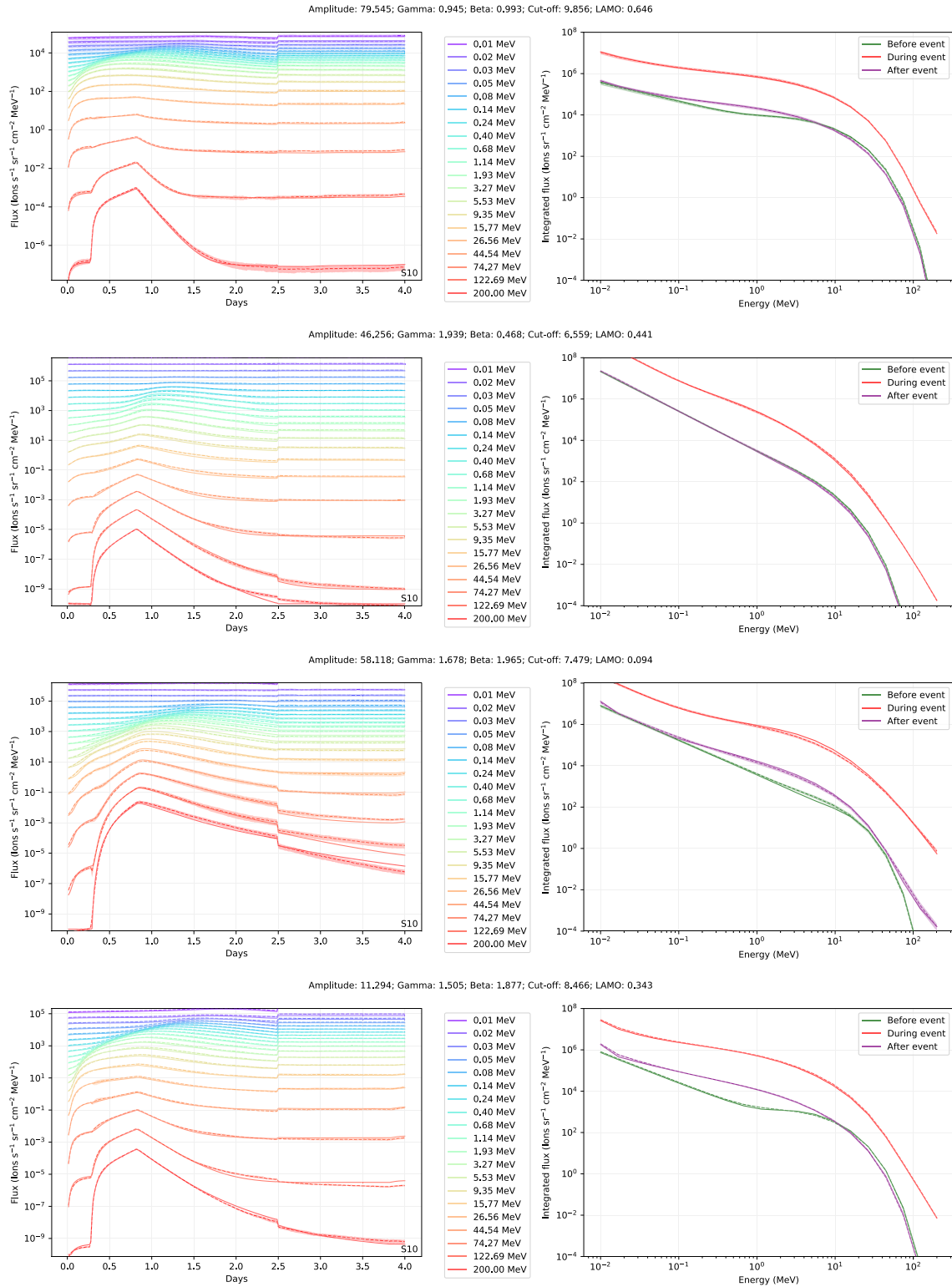
Figure 3. Four SEP events from the test set (data unseen during training), showing the outputs of EPREM (solid lines) and EPREM-S (dashed lines), and the predictive uncertainty in EPREM-S output (shaded regions representing one standard deviation). Out of the 24 streams (see Section 2.2 for details of EPREM grid and other parameters), we identified stream 10 as the *observer connected stream* for an observer at 0.48 AU where the energetic particle distributions are projected and shown here. The panels on the left hand side show the differential flux in different energy levels from 0.01–200 MeV. The panels on the right hand side depict the flux during (event-integrated flux represented by red), before (green) and after the event (purple) for EPREM (solid lines) and EPREM-S (dashed lines) for the events shown on the left hand panels.

are projected. The figures on the left hand side depict a comparison of the EPREM and EPREM-S outputs in flux space for different energy levels as a function of time, and the figures on the right hand side describe the flux as a function of energy level before, during, and after the event. Here, the curve showing the flux "before the event" (green) was obtained by integrating the flux over a few hours before the start of the event, indicated by the sudden increase in the flux. Similarly, for the curve for "after the event" (purple), the flux has been integrated over a few hours after the event has occurred, which is marked by the time around which the fluxes dropped to the background. The curve depicting "during the event" (red) was obtained by integrating the flux over the entire duration of the event. In all the plots, solid lines depict the EPREM simulation output while dashed lines represent the mean EPREM-S predictions of the ensemble, with shaded regions showing the standard deviation representing predictive uncertainty. We noted that the EPREM and EPREM-S outputs are in remarkable agreement, the MSE being 0.07 as we found during validation of the surrogate model. Achieving this level of accuracy with previously unseen input data illustrates that the models achieved generalization beyond the training set. We also noted, for example, in the top row, that uncertainty in the predictions increase when the mean EPREM-S prediction deviates from the EPREM output.

In addition to simulating SEP events in the forward-in-time direction, the low run-time cost of EPREM-S allows the solution of inverse problems using simulation-based inference (Cranmer et al., 2020), where we express the problem as inference over the input parameter space conditioned on an observed event output. In Figures 4 and 5 we present analyses for six different events that we performed using Markov chain Monte Carlo (MCMC) sampling over EPREM parameter space. We formulated event analysis as a Bayesian inference problem where we infer the posterior distribution $p(\psi|\phi)$ describing how parameters ψ are distributed in order to match a given data observation ϕ, based on the prior distribution $p(\psi)$ over parameters and the likelihood $p(\phi|\psi)$ measuring the closeness of match. The prior $p(\psi)$ is defined by the continuous uniform distributions used in EPREM-S training data generation (Section 2.2), and the likelihood

$$p(\phi|\psi) = \mathcal{N}\left(\phi|\mu_\phi(\psi), \sigma_\phi(\psi)\right)$$

is constructed using the EPREM-S ensemble output, representing the probability of flux data ϕ given parameters ψ. In order to confirm correctness, we start with a ground truth event (ψ^*, ϕ^*) with both the parameters and fluxes known. We then take only the fluxes ϕ^* as input for inference and check if the obtained posterior $p(\psi|\phi^*)$ includes the known correct parameters ψ^*. With an MCMC chain of 10,000 iterations, we obtain the posterior distributions shown in Figures 4 and 5, where the dashed vertical lines show known ground truth values of the parameters for each event. We see that all parameters are inferred correctly as their ground truth values (unknown to the inference algorithm) are contained within the resulting posterior distributions. We also see that the value for the roll-over energy or knee (discussed in Section 2) can be inferred with lower uncertainty (or more constrained given the observed data) whereas the remaining parameters have higher uncertainty (less constrained given the observed data), depending on the event.

We measured the runtime costs of EPREM and EPREM-S and present a detailed comparison in Table 1. We take the single-threaded CPU execution of EPREM as the baseline duration with respect to which we compute speed-up factors of EPREM-S. A single execution of EPREM, computing one SEP event on the CPU (EPREM has no GPU support), takes 712,630 ms (11.87 min), whereas the computation of one SEP event by EPREM-S takes 9.82 ms on CPU and 0.32 ms on GPU, corresponding to speed-up factors of 7.25×10^4 and 2.23×10^6 respectively *per execution*. Furthermore, EPREM-S allows multiple inputs to be processed in parallel in a single forward execution using "minibatching," which is the typical way of running many inputs through NNs simultaneously. For example, in the rightmost column of Table 1 we see that we are able to run 10,000 SEP events in parallel in a single execution of EPREM-S, taking 2,495.77 ms on CPU and 0.71 ms on GPU. Finally, we can consider the compound effect of a single execution of EPREM-S being faster than EPREM and the fact that multiple SEP events can be computed simultaneously with a single execution of EPREM-S. For the example of 10,000 events computed simultaneously with EPREM-S, this gives us speed-up factors of 2.85×10^6 on CPU and 1.01×10^{10} on GPU *per event*. All run-time figures presented are the mean of 10 independent repetitions of each experiment. In our experiments the CPU used is an Intel Core i5-12400 and the GPU is an NVIDIA A100 80 GB Tensor Core.

In order to generate our new data set of 32,000 SEP events, we used a simulation duration of four days, using the full configuration for EPREM described in Section 2.2. The configuration consists of parameters that are fixed
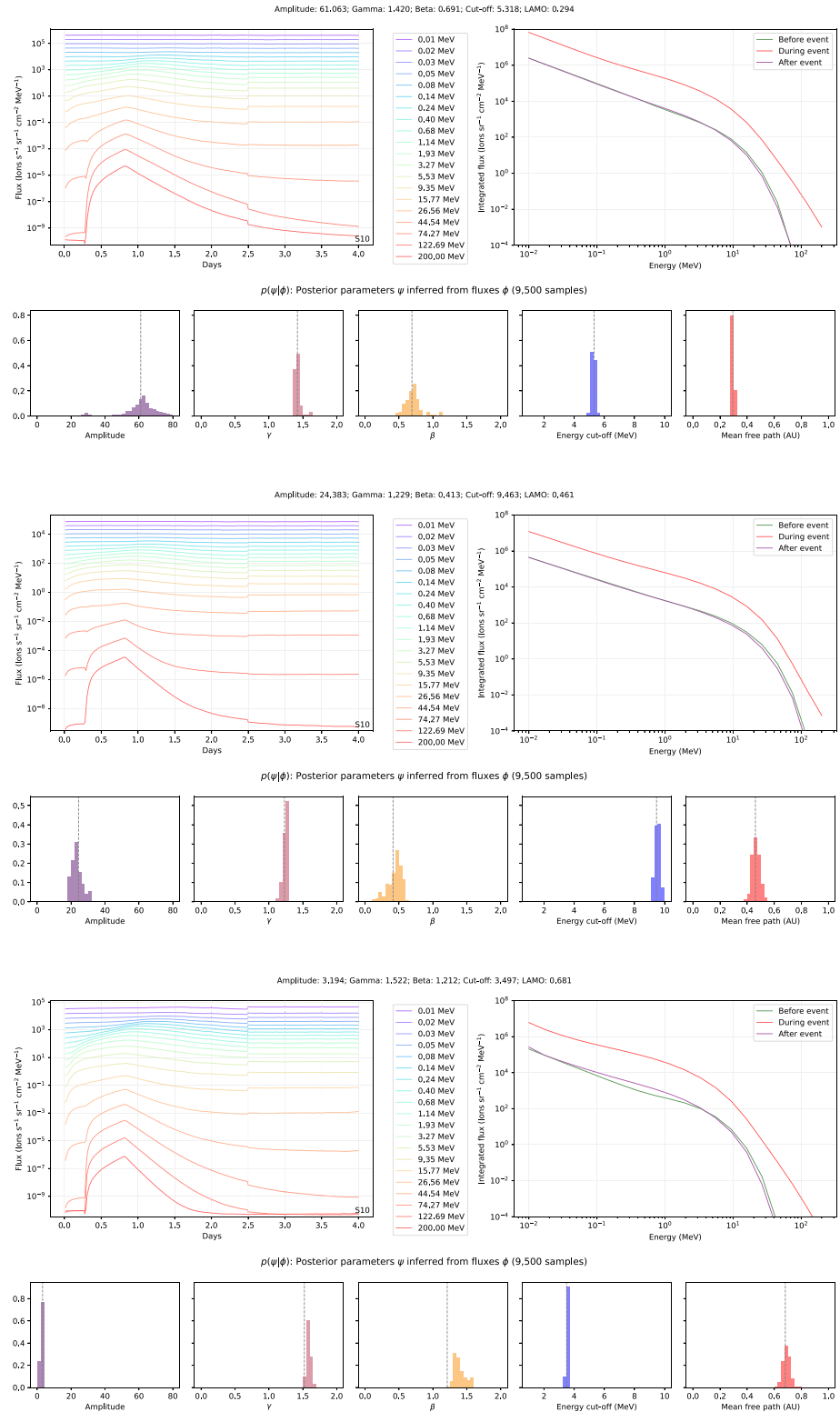
Figure 4. Three events, unseen during EPREM-S training, analyzed via MCMC sampling of EPREM-S outputs conditioned on observed flux data. For each event: *top row* shows the fluxes observed; *bottom row* shows posterior distributions over EPREM parameters inferred via MCMC, conditioned on the fluxes observed. Dashed vertical lines represent the ground truth parameter values for each event.
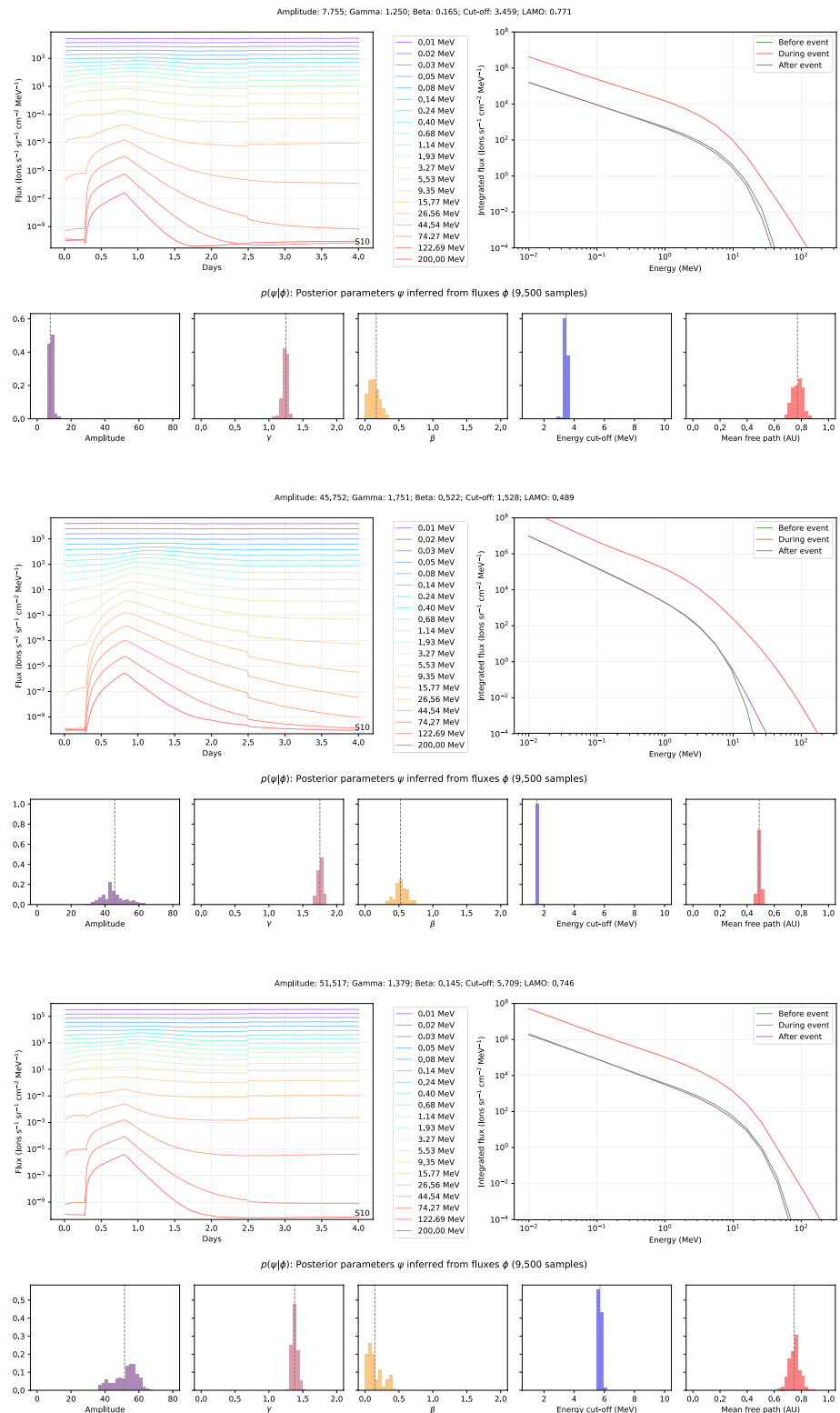
Figure 5. Same as 4 but for another three events unseen during EPREM-S training.

during all the runs (such as the total event duration and time step) and physical parameters that are sampled from the continuous uniform distributions we defined in the parameter space. We executed the EPREM simulator, together with our data sampling code, in a distributed setting on the Expanse supercomputer at SDSC. Given the

Table 1
Runtime Results of EPREM and EPREM-S, Based on the Mean of Ten Independent Runs in Each Case

Model	Run time (milliseconds) per execution				
	Single event	10 events	100 events	1,000 events	10,000 events
EPREM	712,630.00	-	-	-	-
EPREM-S (CPU)	9.82	14.87	35.06	296.42	2,495.77
EPREM-S (GPU)	0.32	0.33	0.36	0.36	0.71
	Speed-up factor per execution relative to EPREM				
EPREM-S (CPU)	7.25×10^4	4.79×10^4	2.33×10^4	2.40×10^3	2.85×10^2
EPREM-S (GPU)	2.23×10^6	2.18×10^6	1.96×10^6	1.97×10^6	1.00×10^6
	Speed-up factor per event relative to EPREM				
EPREM-S (CPU)	7.25×10^4	4.79×10^5	2.03×10^6	2.40×10^6	2.85×10^6
EPREM-S (GPU)	2.23×10^6	2.17×10^7	1.96×10^8	1.98×10^9	1.01×10^{10}

Note. The GPU figures are obtained on the NVIDIA A100 80 GB unit.

simulation configuration we used, EPREM has a single-threaded run time of 712.6 s (11.9 min) per SEP event, and therefore our overall data generation run corresponds to 6,334.5 hr (263.9 days) of CPU computation. We save the data using a gzip-compressed format holding PyTorch (Paszke et al., 2019) tensors where each file, representing an individual SEP event, takes approximately 1.6 MB on disk. The total size of the generated SEP data is 50 GB. We expect this new data set to be a useful resource enabling further ML-based work in this domain and we have made this data publicly available to the research community (Poduval et al., 2023a, 2023b).

4. Discussion and Concluding Remarks

Implementation of artificial intelligence (AI) and machine learning (ML) techniques is increasingly recognized by the scientific community as a powerful tool in all areas of Space Science, particularly in Space Weather, to better extract information from the enormous volume of data and to improve the analytic and predictive performances of models (e.g., Azari et al., 2021; Camporeale, 2019; Poduval, McPherron, et al., 2023; Poduval et al., 2022). Techniques such as NNs have already been implemented in space weather predictions in the 1980s (e.g., Lundstedt, 1996; Lundstedt et al., 2002; Lundstedt, 2005, 2006, and the references therein). These are essentially precursors of the recent *deep learning* techniques that represent a revival of NNs with very large data set, better optimization algorithms, and the availability of modern training hardware such as GPUs. ML techniques including support vector machines (Bobra & Couvidat, 2015), logistic regression (Winter & Ledbetter, 2015), ensemble models (decision trees) (FDLTeam2017, 2017), principal component analysis (Winter & Ledbetter, 2015), long short-term memory (Chen et al., 2019; Tan et al., 2018), Gaussian process (Gruet et al., 2018) and deep NNs (Nishizuka et al., 2018) have been used in the prediction of solar flares, SEPs and the geomagnetic indices such as Dst and Kp.

Lack of sufficient and balanced data set of observed SEP events for training a NN makes it challenging to apply ML methods in the prediction of SEPs. To overcome this limitation to a large extent, we utilised a first principle model EPREM to generate synthetic events. Our final data set contained 32,000 unique events represented as input–output pairs of latent physical parameters and observable SEP event data. Using this data set, we developed and trained the surrogate generative model EPREM-S, which approximates the behavior of the underlying EPREM simulator with high fidelity. Our work on the fast EPREM-S surrogate has two main objectives: (a) making SEP simulations computationally cheap (ideally running in a fraction of a second) on average computer hardware, and (b) therefore making it feasible to develop fast and systematic event analyses based on observed SEP data, using simulation-based inference techniques. Using a deep ensemble approach (Lakshminarayanan et al., 2017) our models also provide an estimate of the predictive uncertainty associated with each simulated SEP event. Importantly, EPREM-S is thousands to millions of times faster (depending on hardware) than EPREM as shown in Table 1 and capable of simulating many events in parallel in a single run, enabling tractable simulation-based inference (Baydin et al., 2019; Cranmer et al., 2020) techniques in this problem domain.

This is the first time such a fast model has been developed that has applications in space weather. With EPREM-S, we would be able to generate thousands of events for studying the influence of the seed particle spectrum in the acceleration and transport of energetic particles in the heliosphere. An example of such an analysis is presented in Figures 4 and 5 where the top panel depicts a ground truth event that is not used for EPREM-S training, the parameter space for which are given on the top and also represented by the dashed line in the bottom panels. Posterior distribution over these parameters, taking only the ground truth fluxes as input, are depicted in the bottom panels. Encouraged by the accuracy with which the parameters are recovered by inference as evident from Figures 4 and 5, we plan to carry out similar Bayesian inference over input parameters using EPREM-S and event-analysis of observed events in the near future. Having developed a fast surrogate, EPREM-S, that is parallelizable is a first step toward simulation-based inference in SEP studies.

The simulated data we generated using EPREM for selected ranges of the five parameters discussed in Section 2.2 will be of great value to the scientific community engaged in SEP studies as they represent various types and levels of synthetic events, analyzing which one can infer useful characteristics of the source region, and the acceleration and propagation of SEPs. Moreover, the surrogate we developed is a generative model using which we can generate large volume of synthetic data to develop a larger database of simulated events. This is another future work we plan.

We expect EPREM-S to serve multiple purposes. In addition to being an ideal tool for forecasting SEPs, it can be used for studying the influence of the model parameters on the SEP profiles, determining the parameter space for accurate SEP simulations using EPREM, and sampling more synthetic data from the fast generative model learned by EPREM-S. We have released our simulated EPREM data set and the EPREM-S model (Poduval et al., 2023a, 2023b) to support a new class of SEP studies by the wider heliophysics community.

Data Availability Statement

The new data set of 32,000 SEP events that we generated using EPREM is an important component of our work which we believe will enable ML-based approaches to SEP modeling and prediction, in addition to significantly accelerating existing workflows in SEP studies. This unique data, the trained model and the software used are publicly available (Poduval et al., 2023a, 2023b).

References

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

Azari, A., Biersteker, J. B., Dewey, R. M., Doran, G., Forsberg, E. J., Harris, C. D. K., et al. (2021). Integrating machine learning for planetary science: Perspectives for the next decade. *Bulletin of the AAS*, *53*(4). https://doi.org/10.3847/25c2cfeb.aa328727

Baydin, A. G., Shao, L., Bhimji, W., Heinrich, L., Meadows, L. F., Liu, J., et al. (2019). Etalumis: Bringing probabilistic programming to scientific simulators at scale. In *Proceedings of the international conference for high performance computing, networking, storage, and analysis (SC19), November 17-22, 2019*. Association for Computing Machinery. https://doi.org/10.1145/3295500.3356180

Behl, H. S., Baydin, A. G., Gal, R., Torr, P. H. S., & Vineet, V. (2020). Autosimulate: (quickly) learning synthetic data generation. In *16th European conference on computer vision (ECCV)*.

Bobra, M. G., & Couvidat, S. (2015). Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *The Astrophysical Journal*, *798*(2), 135–146. https://doi.org/10.1088/0004-637X/798/2/135

Brooks, D. H., & Yardley, S. L. (2021). The source ofthe major solar energetic particle events from super active region 11944. *Science Advances*, *7*(10), eabf0068. https://doi.org/10.1126/sciadv.abf0068

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, *17*(8), 1199–1207. https://doi.org/10.1029/2018SW002061

Cane, H. V., & Lario, D. (2006). An introduction to CMEs and energetic particles. *Space Science Reviews*, *123*(1–3), 45–56. https://doi.org/10.1007/s11214-006-9011-3

Chen, Y., Manchester, W. B., Hero, A. O., Toth, G., Fumier, B. D., Zhou, T., et al. (2019). Identifying solar flare precursors using time series of *SDO*/HMI images and SHARP parameters. *Space Weather*, *17*(10), 1404–1426. https://doi.org/10.1029/2019sw002214

Cranmer, K., Brehmer, J., & Louppe, G. (2020). *The frontier of simulation-based inference*. In (Vol. 117, p. 30055–30062). National Acad Sciences.

FDLTeam2017. (2017). *Machine learning approach to Kp prediction*. AGU Fall Meeting.

Forbush, S. E. (1946). Three unusual cosmic-ray increases possibly due to charged particles from the Sun. *Physics Reviews*, *70*(9–10), 771–772. https://doi.org/10.1103/PhysRev.70.771

Gruet, M. A., Chandorkar, M., Sicard, A., & Camporeale, E. (2018). Multiple-hour-ahead forecast of the Dst Index using a combination of long short-term memory neural network and Gaussian process. *Space Weather*, *16*(11), 1882–1896. https://doi.org/10.1029/2018SW001898

Hill, M. E., Schwadron, N. A., Hamilton, D. C., Difabio, R. D., & Squier, R. K. (2009). Interplanetary suprathermal He+ and He++ observations during quiet periods from 1.5 to 9 AU and implications for particle acceleration. *The Astrophysical Journal Letters*, *699*(1), L26–L30. https://doi.org/10.1088/0004637X/699/1/L26

Himes, M. D., Harrington, J., Cobb, A. D., Baydin, A. G., Soboczenski, F., O'Beirne, M. D., et al. (2022). Accurate machine-learning atmospheric retrieval via a neural-network surrogate model for radiative transfer. *The Planetary Science Journal*, *3*(4), 236–250. https://doi.org/10.3847/PSJ/abe3fd

Hundhausen, A. J. (1972). *Coronal expansion and solar wind* (Vol. 5). Springer-Verlag.

Kaiser, M. L., Kucera, T. A., Davila, J. M., Cyr, O. C. S., Guhathakurta, M., & Christian, E. (2008). The STEREO mission: An introduction. *Space Science Reviews*, *136*(1–4), 5–16. https://doi.org/10.1007/s11214-007-9277-0

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

Kóta, J., Manchester, W. B., Jokipi, J. R., de Zeeuw, D. L., & Gombosi, T. I. (2005). Simulation of SEP acceleration and transport at CME-driven shocks. In G. Li, G. P. Zank, & C. T. Russell (Eds.), *The physics of collisionless shocks: 4th annual IGPP international astrophysics conference* (Vol. 781, pp. 201–206). https://doi.org/10.1063/1.2032697

Kozarev, K., Schwadron, N. A., Dayeh, M. A., Townsend, L. W., Desai, M. I., & PourArsalan, M. (2010). Modeling the 2003 Halloween events with EMMREM: Energetic particles, radial gradients, and coupling to MHD. *Space Weather*, *8*(11), S00E08. https://doi.org/10.1029/2009SW000550

Kozarev, K. A., Evans, R. M., Schwadron, N. A., Dayeh, M. A., Opher, M., Korreck, K. E., & van der Holst, B. (2010). Global numerical modeling of energetic proton acceleration in a coronal mass ejection traveling through the solar corona, and coupling to MHD. *ApJ*, *778*(1), 43. https://doi.org/10.1088/0004-637X/778/1/43

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, *30*.

Lundstedt, H. (1996). Solar origin of geomagnetic stroms and predictions. *Journal of Atmospheric and Terrestrial Physics*, *58*(7), 821–830. https://doi.org/10.1016/0021-9169(95)00105-0

Lundstedt, H. (2005). Progress in space weather predictions and applications. *Advances in Space Research*, *36*(12), 2516–2523. https://doi.org/10.1016/j.asr.2003.09.072

Lundstedt, H. (2006). Solar activity modelled and forecasted: A new approach. *Advances in Space Research*, *38*(5), 862–867. https://doi.org/10.1016/j.asr.2006.03.041

Lundstedt, H., Gleisner, H., & Wintoft, P. (2002). Operational forecasts of the geomagnetic Dst index. *Geophysical Research Letters*, *29*(24), 2181. https://doi.org/10.1029/2002GL016151

Martucci, M., Laurenza, M., Benella, S., Berrilli, F., Moro, D. D., Giovannelli, L., et al. (2023). First ground-level enhancement of solar cycle 25 as seen by the high-energy particle detector (HEPD-01) on board the CSES-01 satellite. *Space Weather*, *21*(1), e2022SW003191. https://doi.org/10.1029/2022SW003191

McComas, D. J., Alexander, N., Angold, N., Bale, S., Beebe, C., Birdwell, B., et al. (2016). Integrated science investigation of the Sun (ISIS): Design of the energetic particle investigation. *Space Science Reviews*, *204*(1–4), 187–256. https://doi.org/10.1007/s11214-014-0059-1

McComas, D. J., Christian, E. R., Cohen, C. M. S., Cummings, A. C., Davis, A. J., Desai, M. I., et al. (2019). Probing the energetic particle environment near the Sun. *Nature*, *576*(7785), 223–227. https://doi.org/10.1038/s41586-019-1811-1

Munk, A., Zwartsenberg, B., Ścibior, A., Baydin, A. G., Stewart, A., Fernlund, G., et al. (2022). Probabilistic surrogate networks for simulators with unbounded randomness. In *38th conference on uncertainty in artificial intelligence (UAI)*.

Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. (2018). Deep Flare Net (DeFN) model for solar flare prediction. *The Astrophysical Journal*, *858*(2), 113–121. https://doi.org/10.3847/1538-4357/aab9a7

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*.

Poduval, B., Baydin, A. G., & Schwadron, N. (2023a). Surrogate model: Data and software. Retrieved from https://zenodo.org/records/10109868

Poduval, B., Baydin, A. G., & Schwadron, N. (2023b). *Surrogate model: EPREM simulations of SEP events*. Zenodo. Retrieved from https://zenodo.org/records/10038847

Poduval, B., McPherron, R. L., Walker, R., Himes, M. D., Pitman, K. M., Azari, A. R., et al. (2023). AI-Ready data in space science and solar physics: Problems, mitigation and action plan. *Frontiers in Astronomy and Space Sciences*, *10*. https://doi.org/10.3389/fspas.2023.1203598

Poduval, B., McPherron, R. L., Walker, R., Himes, M. D., Pitman, K. M., Azari, A. R., et al. (2022). AI-ready data in solar physics and space science: Concerns, mitigation and recommendations. White Paper Submitted to the Heliophysics Decadal Survey White Paper Concepts. Retrieved from http://surveygizmoresponseuploads.s3.amazonaws.com/fileuploads/623127/6920789/107-1870187ec154eee48664bed-68513f0cb_PoduvalBala.pdf

Posner, A., Heber, B., Rother, O., & Gueteersloh, S. (2009). A new trend in forecasting solar radiation hazards. *Space Weather*, *7*(5). https://doi.org/10.1029/2009SW000476

Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., & Tucker, P. (2005). Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, *41*(1), 1–28. https://doi.org/10.1016/j.paerosci.2005.02.001

Reames, D. V. (2013). The two sources of solar energetic particles. *Space Science Reviews*, *175*(1–4), 53–92. https://doi.org/10.1007/s11214-013-9958-9

Reames, D. V. (2021). *Solar energetic particles*. Springer.

Reames, D. V. (2022). A perspective on solar energetic particles. *Frontiers in Astronomy and Space Sciences*, *9*. https://doi.org/10.3389/fspas.2022.890864

Richardson, I. G. (2018). Solar wind stream interaction regions throughout the heliosphere. *Living Reviews in Solar Physics*, *15*, 1. https://doi.org/10.1007/s41116-017-0011-z

Rodríguez-Pacheco, J., Wimmer-Schweingruber, R. F., Mason, G. M., Ho, G. C., Sanchez-Prieto, S., Prieto, M., et al. (2020). The energetic particle detector energetic particle instrument suite for the solar orbiter mission. *Astronomy & Astrophysics*, *642*(A7).

Schwadron, N. A., Bale, S., Bonnell, J., Case, A., Christian, E. R., Cohen, C. M. S., et al. (2020). Seed population pre-conditioning and acceleration observed by Parker Solar Probe. *The Astrophysical Journal - Supplement Series*, *246*(2), 33. https://doi.org/10.3847/1538-4365/ab5527

Schwadron, N. A., Cooper, J. F., Desai, M., Downs, C., Gorby, M., Jordan, A. P., et al. (2017). Particle radiation sources, propagation and interactions in deep space, at earth, the moon, mars, and beyond: Examples of radiation interactions and effects. *Space Science Reviews*, *212*(3–4), 1069–1106. https://doi.org/10.1007/s11214-017-0381-5

Schwadron, N. A., Goodrich, C., Spence, H., Townsend, L., Cucinotta, F., Kim, M.-H., et al. (2007). *Earth-Moon-Mars radiation environment module (EMMREM)*. IEEEAC. paper 1001. https://doi.org/10.1109/AERO.2007.352781

Schwadron, N. A., Gorby, M., Török, T., Downs, C., Linker, J. A., Lionello, R., et al. (2014). Synthesis of 3-D coronal-solar wind energetic particle acceleration modules. *Space Weather*, *12*(6), 323–328. https://doi.org/10.1002/2014SW001086

Schwadron, N. A., Lee, M. A., Gorby, M., Lugaz, N., Spence, H. E., Desai, M., et al. (2015). Particle acceleration at low coronal compression regions and shocks. *The Astrophysical Journal*, *810*(2), 97. https://doi.org/10.1088/0004-637X/810/2/97

Schwadron, N. A., Rahmanifard, F., Wilson, J., Jordan, A. P., Spence, H. E., Joyce, C. J., et al. (2018). Update on the worsening particle radiation environment observed by CRaTER and implications for future human deep-space exploration. *Space Weather*, *16*(3), 289–303. https://doi.org/10.1002/2017SW001803

Schwadron, N. A., Townsend, L., Kozarev, K., Dayeh, M. A., Cucinotta, F., Desai, M., et al. (2010). Earth-Moon-Mars radiation environment module framework. *Space Weather*, *8*(1). https://doi.org/10.1029/2009SW000523

Shirobokov, S., Belavin, V., Kagan, M., Ustyuzhanin, A., & Baydin, A. G. (2020). Black-box optimization with local generative surrogates. In *Advances in neural information processing systems* (Vol. 34).

Sobester, A., Forrester, A., & Keane, A. (2008). *Engineering design via surrogate modelling: A practical guide*. John Wiley & Sons.

Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., & Snow, F. (1998). The advanced composition explorer. *Space Science Reviews*, *86*(1/4), 1–22. https://doi.org/10.1023/A:1005082526237

Tan, Y., Hu, Q., Wang, Z., & Zhong, Q. (2018). Geomagnetic index Kp forecasting with LSTM. *Space Weather*, *16*(4), 406–416. https://doi.org/10.1002/2017SW001764

Temmer, M. (2021). Space weather: The solar perspective. *Living Reviews in Solar Physics*, *18*(1), 4. https://doi.org/10.1007/s41116-021-00030-3

Vidotto, A. A. (2021). The evolution of the solar wind. *Living Reviews in Solar Physics*, *18*(1), 3. https://doi.org/10.1007/s41116-021-00029-w

Winter, L. M., & Ledbetter, K. (2015). Type II and Type III radio bursts and their correlation with solar energetic proton events. *The Astrophysical Journal*, *809*(1), 105. https://doi.org/10.1088/0004-637x/809/1/105