Key Points:

- An end-to-end deep neural network model ensemble, trained on remote sensing and in situ data sets, is used to forecast SEP properties
- Model ensemble maximizes the utilization of an imbalanced data set and enhances forecasting confidence with uncertainty estimates
- Solar energetic particle occurrence probabilities are utilized as weights in the loss function to enhance regression performance for events

Correspondence to:

M. A. Dayeh,
maher.aldayeh@swri.org

MEMPSEP-II. Forecasting the Properties of Solar Energetic Particle Events Using a Multivariate Ensemble Approach

Maher A. Dayeh[1,2] , Subhamoy Chatterjee[3] , Andrés Muñoz-Jaramillo[3] , Kimberly Moreland[1,2,4,5], Hazel M. Bain[4,5] , and Samuel T. Hart[1,2]

[1]Southwest Research Institute, San Antonio, TX, USA, [2]The University of Texas at San Antonio, San Antonio, TX, USA, [3]Southwest Research Institute, Boulder, CO, USA, [4]Cooperative Institute for Research in Environmental Sciences, University of Boulder, Boulder, CO, USA, [5]Space Weather Prediction Center, NOAA, Boulder, CO, USA

Abstract Solar Energetic Particles (SEPs) form a critical component of Space Weather. The complex, intertwined dynamics of SEP sources, acceleration, and transport make their forecasting very challenging. Yet, information about SEP arrival and their properties (e.g., peak flux) is crucial for space exploration on many fronts. We have recently introduced a novel probabilistic ensemble model called the Multivariate Ensemble of Models for Probabilistic Forecast of Solar Energetic Particles (MEMPSEP). Its primary aim is to forecast the occurrence and physical properties of SEPs. The occurrence forecasting, thoroughly discussed in a preceding paper (MEMPSEP-I by Chatterjee et al., 2024a, https://doi.org/10.1029/2023sw003568), is complemented by the work presented here, which focuses on forecasting the physical properties of SEPs. The MEMPSEP model relies on an ensemble of Convolutional Neural Networks, which leverage a multi-variate data set comprising full-disc magnetogram sequences and numerous derived and in-situ data from various sources (MEMPSEP-III by Moreland et al., 2024, https://doi.org/10.1029/2023SW003765). Skill scores demonstrate that MEMPSEP exhibits improved predictions on SEP properties for the test set data with SEP occurrence probability above 50%, compared to those with a probability below 50%. Results present a promising approach to address the challenging task of forecasting SEP physical properties, thus improving our forecasting capabilities and advancing our understanding of the dominant parameters and processes that govern SEP production.

Plain Language Summary Solar Energetic Particles (SEPs) are important for understanding space weather. SEPs are hard to predict because they come from various sources and go through complicated processes. However, knowing when and how intense SEPs are in space is crucial for things like space exploration. In this work, we created a new model called MEMPSEP to help predict when and what the properties of these SEPs will be when they are detected. We used data from different sources to train the model on how to make these predictions. The results show that MEMPSEP is better at predicting the properties of SEPs when the chances of them happening are more than 50%. This is a significant step forward in improving our ability to forecast SEPs and understanding the factors that influence them in space.

1. Introduction

Solar energetic particles (SEPs) are an integral component of Space Weather. SEPs in the inner heliosphere comprise different origins, acceleration, and transport mechanisms that alter their properties significantly (see review by Desai and Giacalone (2016)). SEPs are generally classified into two main classes, impulsive and gradual, which correspond to the way SEPs are accelerated. Impulsive SEPs (ISEPs) are generated during explosive flare phases (Reames, 1999), while gradual SEPs (GSEPs) are produced by diffusive acceleration processes at traveling interplanetary (IP) shocks (Cane, 1988), which are associated with the expulsion of coronal mass ejections (CMEs), or develop at corotating IP structures (e.g., Gosling & Pizzo, 1999; Mason et al., 2008). Although both ISEPs and GSEPs exhibit distinct properties when considered individually, the majority of SEP events observed in the heliosphere display a combination of both properties, with large event-to-event variability. This is attributed to the complex environment surrounding the origin (Mason et al., 2004; Desai et al., 2006; Dayeh et al., 2009), acceleration (e.g., Desai et al., 2016), and propagation (e.g., Mason et al., 2012) of SEPs, leading to three unresolved primary concerns, with a mounting body of work continuing to understand these major processes (see review by Zhang et al. (2021) and references therein).

The first issue pertains to the origin of SEPs. Observations of exceptionally rare solar wind elements in SEP events provide compelling evidence that CME-driven IP shocks selectively accelerate material from a

Project administration: Maher A. Dayeh
Resources: Maher A. Dayeh
Software: Maher A. Dayeh, Subhamoy Chatterjee, Kimberly Moreland, Hazel M. Bain, Samuel T. Hart
Supervision: Maher A. Dayeh
Validation: Maher A. Dayeh, Subhamoy Chatterjee, Kimberly Moreland, Hazel M. Bain
Visualization: Maher A. Dayeh, Subhamoy Chatterjee, Andrés Muñoz-Jaramillo
Writing – original draft: Maher A. Dayeh, Subhamoy Chatterjee
Writing – review & editing: Maher A. Dayeh, Subhamoy Chatterjee, Andrés Muñoz-Jaramillo, Kimberly Moreland, Hazel M. Bain, Samuel T. Hart

suprathermal pool consisting of heated solar wind, coronal material, remnants of solar transient events, and more (see Gloeckler, 2003; Desai et al., 2006; Mason et al., 2004; Dayeh et al., 2009, 2017; Alterman et al., 2023; Wijsen et al., 2023). Quantifying the contributions of these different seed populations to SEPs is crucial for understanding the variability observed from one event to another. The second challenge lies in understanding the acceleration of SEPs. Diffusive shock acceleration (DSA) is considered the primary candidate of SEP acceleration at IP shocks, comprising the shock-drift mechanism at quasi-perpendicular shocks (Decker, 1981), and the first-order Fermi mechanism at quasi-parallel shocks (Lee, 1983); however, observations showed that theory can only explain certain behaviors of SEPs and does not provide a universal explanation for a plethora of mixed cases. The third aspect requiring attention is SEP transport. Scattering within the IP medium plays a significant role in shaping SEP observations at 1 au. Studies have revealed evidence of heavy-ion scattering by Alfven waves generated by streaming protons accelerated at shocks, indicating that IP scattering is sometimes dominated by a dynamic wave spectrum rather than a universal background spectrum (e.g., Ng et al., 1999; Tylka et al., 2005). Work by Cohen et al. (2005); Desai et al. (2016); Mason et al. (2006) attributed SEP rigidity-dependent scattering as particles propagate through the corona and IP medium, or to substantial cross-field diffusion near the Sun. However, SEP scattering alone does not fully account for the puzzling variability observed in SEP properties at 1 au.

SEP hazards in the near-Earth environment are of multiple folds. The first and foremost important risk is the radiation hazard for astronauts in Human spaceflight (e.g., Cucinotta, 2022 and references therein), and passengers on high-flying polar flight routes (Beck et al., 2005). SEPs are intense ionizing radiation that could have acute or chronic effects on humans. Acute effects cause radiation sickness that puts astronaut lives at immediate risk. Chronic dose effects include carcinogenesis, neural damage, and degenerative tissue damage diseases affecting major organs. In addition to the human element, high-energy penetrating SEPs are known to cause irreversible malfunctions in spacecraft electronics (e.g., Webb & Allen, 2004). While heavy ions cause most damage to electronics, protons account for most radiation hazard because they are abundant and can penetrate easily through spacecraft shielding. To mitigate SEP radiation hazards, space designers rely on models of the extreme space environment to estimate the radiation protection calculations of instrumentation and crew protection. This prediction is often done for several years ahead of the anticipated missions. While this task is feasible for the more-stable and seasonal energetic sources such as the galactic cosmic rays (GCRs) and anomalous cosmic rays (ACRs), it is not applicable to SEPs, not only because of their transient nature, but also due to the large event-to-event variability in SEP properties, such as their peak intensities, spectral indices, elemental abundances, maximum energy attained by particles, etc. Forecasting SEPs (e.g., Dayeh et al., 2010; Kozarev et al., 2010) and their properties, such as peak flux, event duration, composition, etc., ahead of their arrival is thus critically informative to mitigate SEP radiation hazards. With the emergence of Machine Learning (ML) algorithms in space weather applications, physical processes are often ignored at the expense of trend detection and correlation analysis. While ML shows promising results in predicting space weather events, a complete understanding of the underlying physics remains challenging. Therefore, it is essential to recognize the importance of utilizing both ML and traditional physics-based approaches to gain comprehensive insights into space weather phenomena. Relating SEP characteristics to their precursors has been the subject of substantial research (e.g., Anastasiadis et al., 2017; Belov et al., 2005; Dayeh et al., 2018; Garcia, 2004a, 2004b; Gopalswamy et al., 2004; Kahler and Ling (2018); Lara et al., 2003; Lario et al., 2023, Papaioannou et al. (2018); Richardson et al., 2018). This research revealed correlative relationships that have enabled the development of qualitative predictive models of SEP phenomena. For instance (Dayeh et al., 2018), explored the parameter space of ESPs with their shock properties and found that shock speed is directly related to the associated ESP peak, a result that is also reached by Lario et al. (2023) for a different set of events. Richardson et al., (2018) leveraged CME speeds and magnetic connectivity to estimate the SEP peak proton intensity. Furthermore, Posner, (2007) utilized relativistic electrons to forecast SEP arrival times and profiles. However, to enhance these predictive models, there is a pressing need for a comprehensive multivariate analysis that considers all available relevant parameters and interwoven physical processes, including the conditions of solar eruptions, as also detailed in Bain et al. (2021). This should also advance the capability to have probabilistic forecast models as opposed to simple binary-outcome models. For a detailed review of current models, see, Whitman et al. (2022).

Ensemble models enable proper estimation of prediction uncertainty and ultimately improves correct decisions in critical situations. A new model, the "Multivariate Ensemble of Models for Probabilistic Forecast of Solar Energetic Particles" or MEMPSEP, has recently been developed and extensively validated (Chatterjee et al., 2024a;

Table 1
Energy Ranges for SEP Peak Flux

Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
≥5 MeV	≥10 MeV	≥30 MeV	≥60 MeV	≥100 MeV

hereafter MEMPSEP-I). Their work also investigates the advantage of utilizing an ensemble of models to produce confident and reliable outcome. MEMPSEP enables forecasting the occurrence of an SEP event and in the case of event forecast, it provides the forecast of its physical properties such as peak intensity, duration, and elemental composition, among other user-defined properties. This work complements MEMPSEP-I and describes the regression branch of MEMPSEP which allows SEP properties forecasting. Section 2 briefly describes the data set, Sections 3, 4, and 5 describe the ensemble of models setup, Section 6 presents the results, and Section 7 if for discussion and conclusions.

2. Data Set

The data set used in MEMPSEP development is extensively described in Moreland et al. (2024; hereafter MEMPSEP-III). The data set is comprised of a large set of parameters from remote (e.g., images) and in situ (e.g., time series and pre-eruptive event properties) sources. In this context, ingested data are the "predictors" and the outcome "targets" are the SEP parameters. The sections below provide a brief description of both.

2.1. Predictors

It has been already established in MEMPSEP-I that both in-situ and remote-sensing parameters are crucial in reliably forecasting the occurrence of SEPs. We use a sequence of full-disk line-of-sight magnetograms (SoHO/MDI + SDO/HMI) prior to flare onset as remote-sensing input. For the sequence, we use a sampling interval of 6 hr over a period of 3 days and rebin full-disc magnetograms to a size of 256pixels $\times$ 256pixels creating a data cube of shape $256 \times 256 \times 13$. To homogenize MDI and HMI magnetograms we use a conversion factor of 1.3 (MDI = 1.3*HMI) and clip the field strength within $[-1000\,G, 1000\,G]$. This creates a continuous data set with a longer observational baseline than from a single instrument. We use a set of in-situ properties namely: solar wind (SW) temperature, SW velocity, SW density, interplanetary magnetic field (IMF) field strength (B), IMF Bx, By and Bz components, as well as particle fluxes and ratios (Fe/O, H, O, and Fe) calculated prior to every flare onset. These pre-event quantities describe the suprathermal population prior to the event along with the IP conditions. We emphasize here that the data is pre-event and the forecasting is thus a true forecast as opposed to nowcasting. Additionally, we use Wind/Waves radio-burst time-frequency images (432 [time] $\times$ 80 [frequency]), L1-electrons (7-channels of 8640 time points), and X-ray (2-channels of 1441 time points) time series as in-situ inputs. For details about the data preparation see MEMPSEP-III.

2.2. Targets

We set forecast targets as 7 SEP properties, namely, (i to v) integrated peak flux (in five energy bands determined from GOES; see Table 1), (vi) onset time of SEP w.r.t. onset of flare (named as "FlrtoSEP dt") and (vii) duration of SEPs. It must be noted that the onset time and duration of SEPs named as "evens" are available only when the peak flux (>10 MeV) exceeds a threshold of 5 p.f.u. within 6 hr from the flare onset. This user-defined threshold was chosen to ensure that selected events are gradual SEPs and not impulsive or particle bursts, as well as to optimize the number of positive events while still lying above the instrument background level. For non-events (e.g., flares or CMEs that do not show a rise in particle flux near Earth above the pre-defined p.f.u. threshold), we set those parameters to a constant value of 720 hr (see definition of "events" and "non-events" in MEMPSEP-I paper. Before fitting a model we take the logarithm of targets to minimize the effect of outliers.

3. Model Description

3.1. Architecture

MEMPSEP workhorse is an ensemble of Convolutional Neural Network (CNN; see LeCun et al. (2015)) architecture that ingests multiple inputs such as images, scalar parameters and time series (Figure 1) and forecasts the indicated 7 SEP properties (i.e., Targets). The CNN processes the input image sequences with repeated 2D convolution, batch-normalization, non-linear activation and max-pooling. The flattened layers are concatenated with scalar parameters and features extracted from time series data, and passed through dense (fully connected) layers to produce the SEP properties. Dropout is used before the last layer causing random disconnections
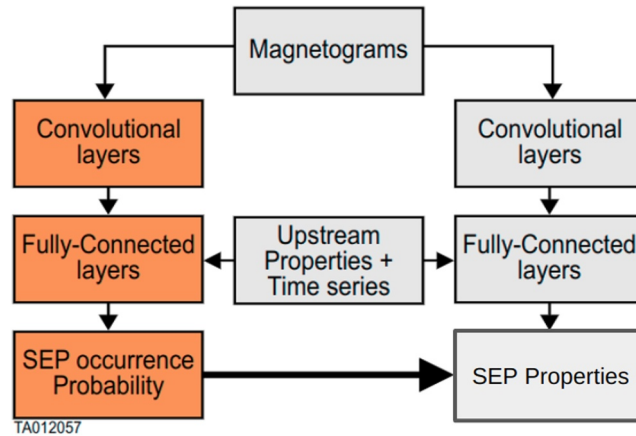
Figure 1. Convolutional neural network model to perform SEP property regression using feedback from the pre-trained classification model. We first train the classification branch (for detailed architecture see MEMPSEP-I) depicted in orange and then freeze weights and biases of that while training the regression branch shown in gray. The outcome of the classification branch (i.e., the SEP occurrence probability) penalizes the loss function of the regression branch depicted by the horizontal arrow at the bottom of the figure.

between nodes and the helping in regularization of the model. It must be noted that the temporal information of the time series data is fused in the initial layers of the CNN and then broken into multiple feature layers that further passed onto subsequent layers. The model architecture is similar to that of the model for forecasting SEP occurrence (described in MEMPSEP-I) except for the lack of any non-linear activation at the end. We avoid any non-linearity after the final layer as activation functions like "ReLU" may cause a dying-ReLU problem and drive the model optimization toward all-zero predictions.

3.2. Optimization

To optimize the model's weights and biases we minimize the mean squared error (MSE) loss described by:

$$\mathrm{MSE} = \frac{1}{\mathrm{NB}}\sum_i\sum_{j=1}^{B}\left(\mathrm{target}_{i,j} - \mathrm{predicted}_{i,j}\right)^2 \tag{1}$$

where $i \in \{\mathrm{peak1, peak2, peak3, peak4, peak5, onset, duration}\}$, N stands for the number of targets (i.e., 7) and j runs over a batch of size B. We set B to 10 for training the model. We use a random training-validation split of 4:1 (1020 (non-events) + 680 (events) for training and 255 (non-events) + 170 (events) for validation). We make sure that there is no temporal leakage between the training and validation set. We stop training the model when the validation accuracy stops improving (Figure 2). We call this model 'non-gated". We also make use of SEP occurrence probability produced by another CNN model depicted in MEMPSEP-I. At this stage, we freeze the weights and biases that were learned during the optimization of the SEP classification branch and just train the regression branch penalizing the loss function of the classification outcome through a custom probability-weighted mean-squared error (PSE) loss described as follows:

$$\mathrm{PSE} = \frac{1}{\mathrm{NB}}\sum_i\sum_{j=1}^{B} p_j\left(\mathrm{target}_{i,j} - \mathrm{predicted}_{i,j}\right)^2 \tag{2}$$

p_j represents the predicted SEP occurrence probability by the frozen classification branch for training instance j. We adapt this approach so that the model is driven to match the target better for the instances that are predicted to have a higher chance of SEP occurrence. We call this model "gated." Figure 2 shows smaller values of PSE as compared to MSE as the probability is always less than 1. Also, the gated model shows faster decay of PSE as compared to MSE in non-gated model for both training and validation. The shaded regions in Figure 2 depict uncertainty from the ensemble models described in detail in the next section.
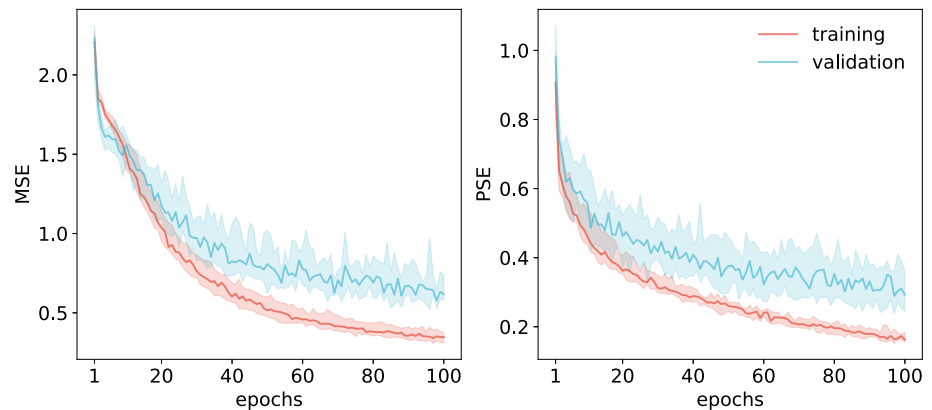
Figure 2. Covergence of loss function with epoch. Left panel depicts mean squared error loss for training and validation set as a function epoch. Right panel shows PSE loss for training and validation set as a function of epoch. The shaded regions depict the uncertainty (10th to 90th percentile produced by model ensemble.).

4. Ensemble of Models

The fact that there is a significant class imbalance between SEP events and non-events is widely recognized as a challenging issue for the effectiveness of Machine Learning-based SEP prediction models (Kasapis et al., 2022). To address this problem, we adopt a strategy where we create several training and validation sets by combining substantially diverse non-events with the same set of events. To achieve this, we begin by assessing the distribution of flare peak strength in events using an equal-frequency histogram with 10 bins. Next, we utilize the edges of these bins to randomly select non-events, aiming for a number that is 1.5 times greater than the number of events in each bin. This selection process is repeated ten times, ensuring minimal overlap between the chosen non-events for each bin. We have approximately 850 event samples (85 per bin), and for each draw, we randomly select around 1275 non-events (approximately 127 per bin). As described in the previous section, 1,275 non-events are divided into 1020 for training and 255 for validation; 850 events are divided into 680 for training and 170 for validation. So, in total, for 10 draws, we end up with close to 10,000 non-event samples. These non-event samples, along with the 850 events, are used to train and validate a model (Figure 2). By doing this, we

create an ensemble of 10 models. This approach is necessary because training a single model would have resulted in a severe class imbalance of 1 event to 15 non-events.

5. Design of Test Set

We design a test set for the forecast of SEP occurrence probability (also used in MEMPSEP-I) that represents all the flare classes and is not affected by the solar cycle phase in event selection (Figure 3). We select 5 events and 10 non-events per flare bins for all the 10 bins and also ensure that a year is dropped from random draw When it appears more than $\frac{N_{\text{events}} * N_{\text{bins}}}{N_{\text{years}}}$ times. N_{years} stands for the number of years shown on the x-axis of Figure 3 (i.e., 16). We also ensure that the test set events are well separated in time from training and validation sets. This test set allows us to provide an unbiased estimate of model-ensemble performance. Also note that the test set is chosen before the selection of the training and validation set. In addition, the model ensemble is never tuned on the test set.

6. Regression Results

For every data point on the test set, we acquire the model-ensemble inference for both 'non-gated" and 'gated" regression. We produce scatter-plots of target versus predicted values of SEP peaks (1–5), FlrtoSEP dt, and SEP duration and color the points according to the median SEP occurrence
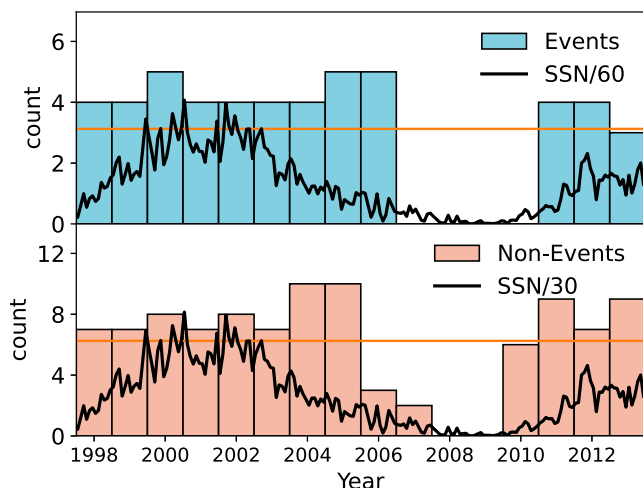
Figure 3. The top panel of the yearly histogram displays events in sky-blue, while the bottom panel represents non-events shown in pink. The horizontal orange line indicates the uniform event count per bin if the events were evenly distributed. The black curve depicts the smoothed monthly sunspot number cycle. From the graph, we can observe that event counts do not follow the modulation of the sunspot number cycle, ensuring an unbiased estimation of model performance on the test set.
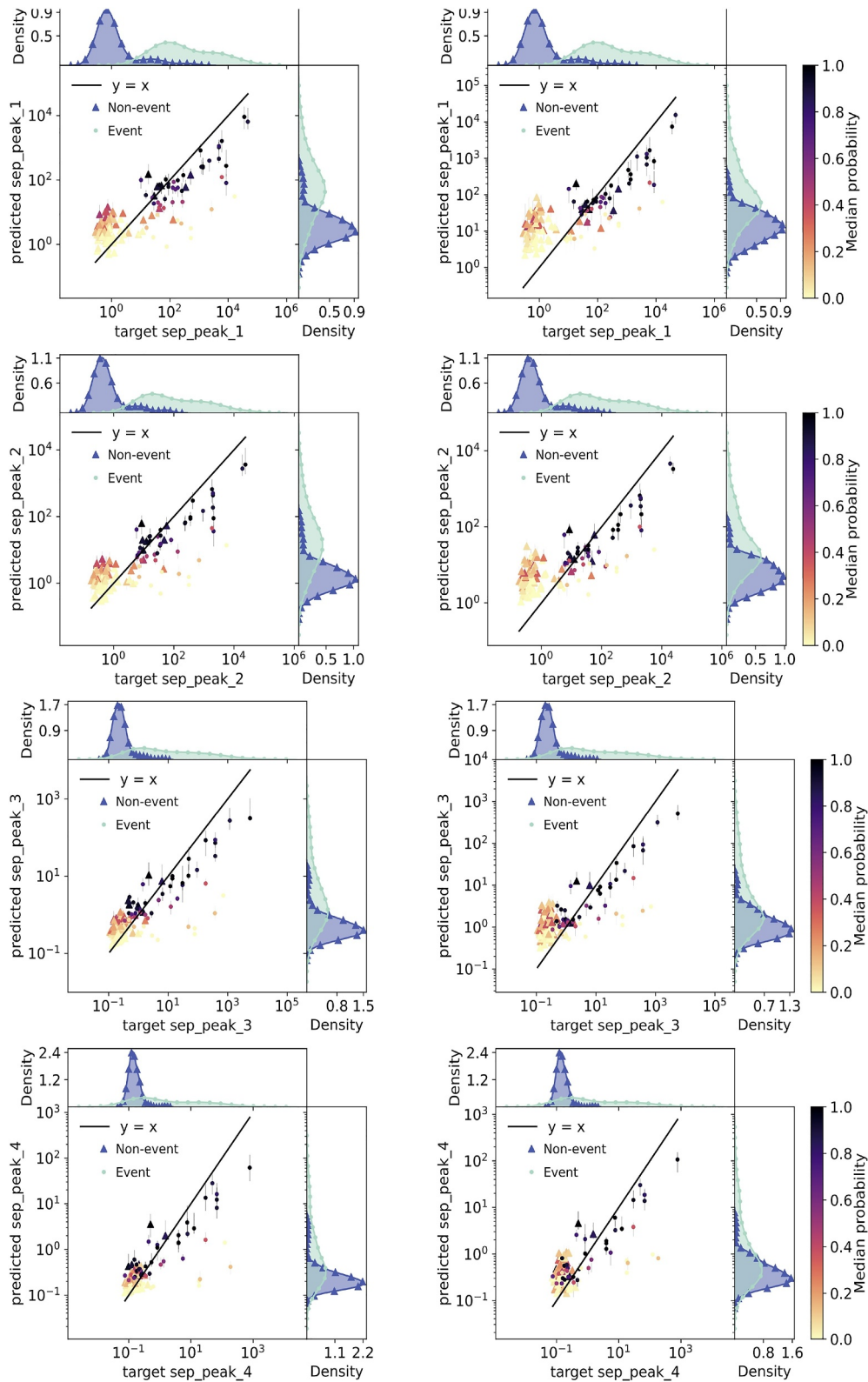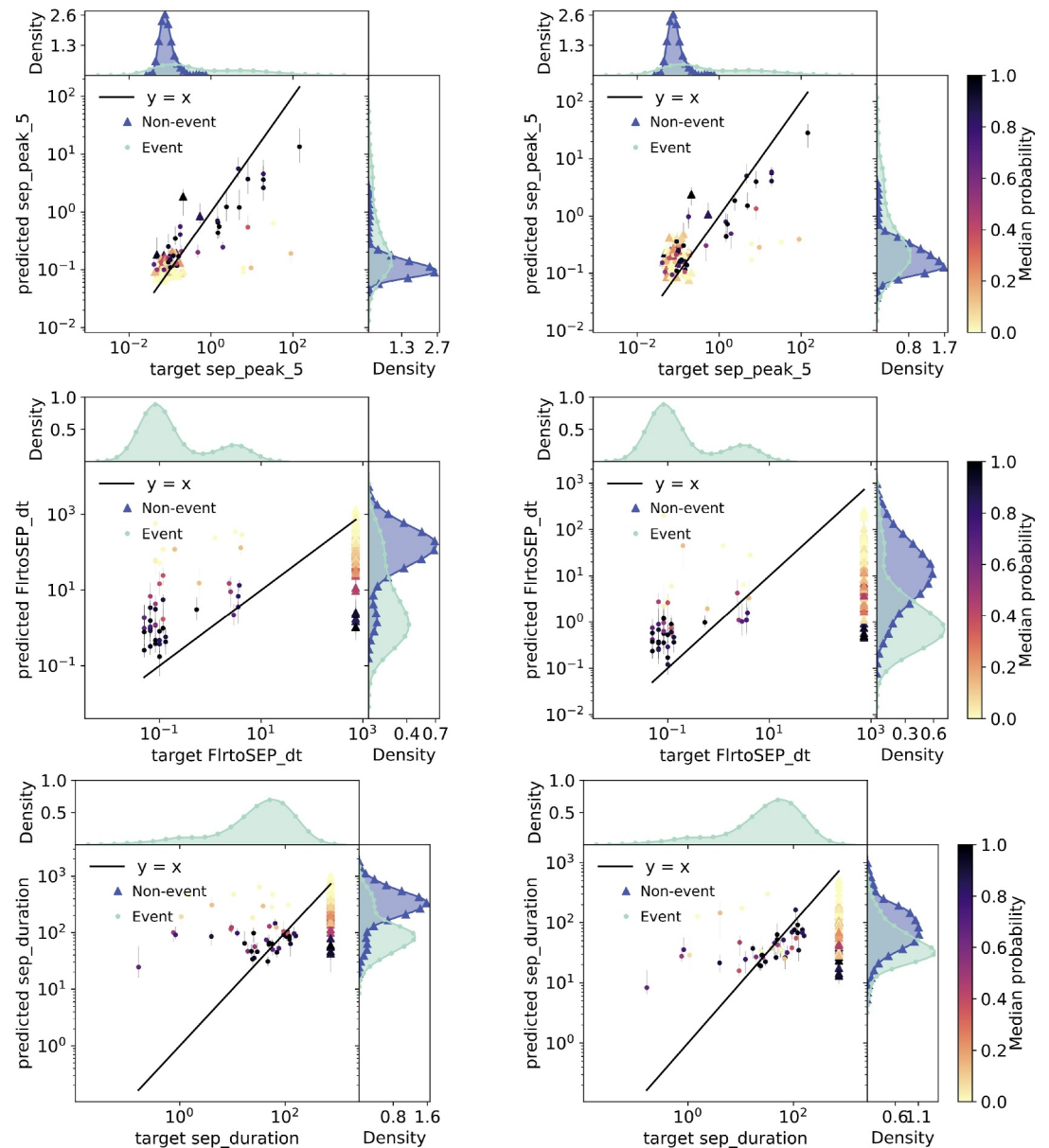
Figure 4.

Figure 5. Same as Figure 4 except that here the Regression results are shown for solar energetic particles (SEP) peak flux in the highest energy range, SEP onset time, and duration.

probability (p) predicted by the model ensemble. We also add different symbols to distinguish between the points representing SEP "events" and "non-events" ground truths (Figures 4 and 5). We use different metrics to quantify the regression performance, namely, Pearson Correlation Coefficient (CC), Root Mean Squared Error (RMSE), and R^2 score for groups of points having $p \geq 0.5$ and $p < 0.5$ R^2 score, also known as the coefficient of

Figure 4. Regression of solar energetic particles (SEP) peak flux in different energy bands. Outcome of non-gates and gated models are depicted through left and right columns respectively. Rows represent different SEP peaks. Each scatter plot depicts the target versus median of predicted values with Model-ensemble on the test set. The points are colored by the predicted SEP probability through the classification branch. Ground-truth of event class is depicted in different symbols (triangular for non-events; circular for events). The error bars represent the 25th to 75th percentile of predictions by the model-ensemble. Marginals represent the probability density of the target and predicted medians for events and non-events.

Table 2
Skill Scores for Predicted Events and Non-Events (Non-Gated)

	Peak 1		Peak 1		Peak 1		Peak 1		Peak 1		Time to SEP		SEP duration	
	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$
CC	0.83	0.48	0.88	0.48	0.92	0.48	0.92	0.51	0.9	0.5	0.31	0.35	0.09	0.22
RMSE	0.68	0.92	0.61	0.84	0.53	0.67	0.48	0.55	0.46	0.51	1.44	1.22	0.82	0.63
R^2 score	0.52	0.2	0.64	0.2	0.76	0.21	0.77	0.22	0.74	0.2	0.04	0.1	−0.02	0.0

Table 3
Skill Scores for Predicted Events and Non-Events (Gated)

	Peak 1		Peak 1		Peak 1		Peak 1		Peak 1		Time to SEP		SEP duration	
	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$	$p \geq 0.5$	$p < 0.5$
CC	0.86	0.36	0.91	0.38	0.94	0.39	0.92	0.42	0.89	0.39	0.3	0.28	0.24	0.2
RMSE	0.6	1.22	0.53	1.1	0.48	0.81	0.46	0.62	0.44	0.57	1.41	1.75	0.81	0.94
R^2 score	0.63	−0.41	0.73	−0.34	0.81	−0.14	0.79	0.02	0.75	0.02	0.07	−0.85	0.01	−1.27

determination and represents the fraction of the variability in the regression target that can be explained by the model. Thus an R^2 score close to 1 is desirable. An R^2 score less than or equal to zero is an indicator that the model is not performing any better than just an average of the regression target.

We observe a general tendency of the points to get closer to the "y = x" line for higher occurrence probabilities. Except for "FlrtoSEP dt" and "SEP duration" we find higher contrast between the metric corresponding to those two groups ($p \geq 0.5$ and $p < 0.5$) for "gated" regression (Table 3) as compared to "non-gated" (Table 2). For both the regression types, we find the CC between the logarithm of the target and the predicted ensemble-median of SEP peaks to be statistically significant and >0.8. Also, RMSE ranges between 0.44 and 0.68 signifying a factor of 3–5 difference between target and predicted values. For "non-gated" and "gated" regression, we find a minimum R^2-score of 0.52 and 0.63 respectively for SEP peaks with $p \geq 0.5$. For "FlrtoSEP dt" and "SEP duration" we generally find worse performance than other regression targets and also observe that the "non-gated" models perform slightly superior to "gated" ones (Table 3).

To understand the ability of the model ensemble to identify "events" and "non-events" as separate populations, we visualize the probability density of the predicted values and check if there is a significant difference between those for "non-events" and "events." We perform an independent one-tailed *t*-test between those and find a significant difference for all the targets (*p*-values<0.05; see Table 4) with "non-gated" and "gated" models. Lastly, to examine the effectiveness of the "gated" model, we plot skill scores as functions of probability threshold for all regression targets (Figures 6 and 7). We find the "gated" model to cause improvements for most of the SEP peaks as compared to "non-gated" for higher values of the threshold with a crossover happening at around a probability of 0.4. For "FlrtoSEP dt" and 'SEP duration" we find that crossover happening for RMSE and R^2 scores with "non-gated" model having slightly superior scores as compared to "gated". Although it should be noted R^2 score for those two targets is close to 0 or negative indicating the "gated" and "non-gated" model performance to be no better that a constant outcome of average target value.

7. Discussion and Conclusions

It can be observed that skills scores for "FlrtoSEP dt" and "SEP duration" are significantly worse than those for other regression targets. False positives depicted by triangles that are dark ($p \geq 0.5$) in the scatterplots play a major role in worsening the Skill scores for 'FlrtoSEP dt" and "SEP duration." Even though the model ensemble is able to predict SEP Peak values for those non-events, there may not be enough information in the inputs to predict such large values of 'FlrtoSEP dt" and 'SEP duration". This could also be the under-represented set of such non-events in the training set that cross the 5 p.f.u. cutoff for SEP flux (>10 MeV) but takes longer than 6 hr. Thus the model ensemble tends to wrongly predict shorter "FlrtoSEP dt" and "SEP duration" for those. The "gated" model producing even smaller values for those targets corresponding to false positives worsens the skill scores further as compared to "non-gated." Even though we evaluate uncertainties in prediction through the model ensemble, it is

Table 4
*p-Values for Independent t-Test Between Predicted Values of Events and Non-Events*

	Peak 1	Peak 1	Peak 1	Peak 1	Peak 1	Time to SEP	SEP duration
Non-gated	0.021	0.023	0.013	0.014	0.008	0.0	0.0
Gated	0.033	0.023	0.019	0.027	0.027	0.001	0.0

unknown when those uncertainties converge. Further investigations are necessary to ensure well-calibrated uncertainties.

Unlike a model like logistic regression or random forest, MEMPSEP integrates multi-modal data in embedding space. Due to the nonlinearities involved in the construction of the model and the lack of physical meaning in the embedding space, it is thus not straight forward to quantify the impact of each observable on the regression outcome. However, as MEMPSEP is driven by SEP occurrence probability, we have evaluated the impact of adding in-situ parameters to remote-sensing inputs on classification accuracy and found significant improvements. This is described in detail in MEMPSEP-I.

In summary, we make use of the class imbalance between SEP events versus non-events and train an ensemble of CNNs to forecast a set of SEP parameters. The skills scores show that the ensemble is able to make tighter (i.e., smaller deviation of the prediction from target) predictions on SEP properties for the test set data points with SEP occurrence probability ≥ 0.5 as compared to those with probability <0.5. We experiment by applying weightage to the loss function with occurrence probability and observe improvement as increased contrast between the metrics for those two probability groups.

With our current data set and approach, the regression results are not as good for "FlrtoSEP dt" and "SEP duration" as for the remaining properties. In some aspects, this is not surprising. SEP profiles get altered en route to 1 au by several convoluted processes, including scatterings, diffusion, turbulence, and other effects. The pre-event predictors are thus part of the puzzle and further physics-based properties relating to transport
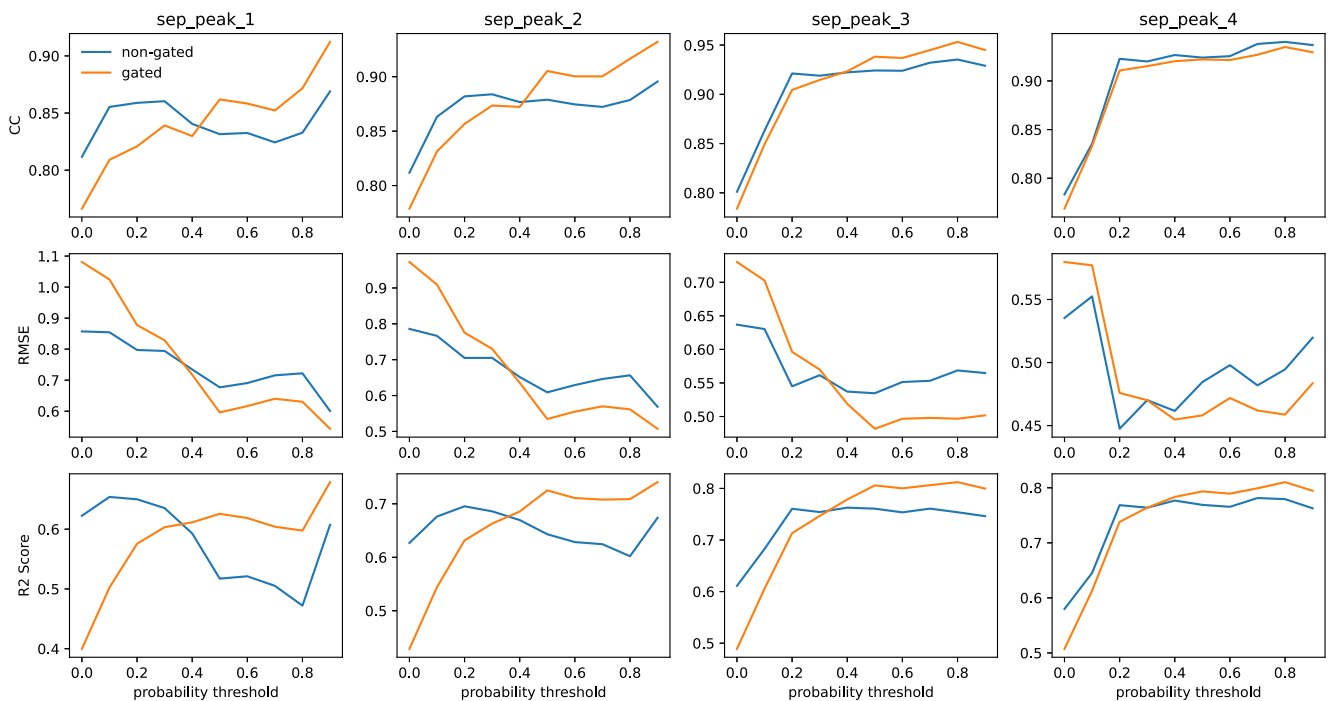
Figure 6. Skill scores versus probability threshold for non-gated and gated regression. The top row depicts a variation of the Pearson correlation coefficient between predicted and target solar energetic particles (SEP) properties as a function of the probability threshold to filter out points with higher predicted SEP occurrence probability. The Middle and bottom rows show variations of Root Mean squared Error (RMSE) and R^2 score respectively as a function of the probability threshold. Columns depict different SEP properties. The outcomes of "Non-gated" and "Gated" regression models are shown with blue and orange curves respectively.
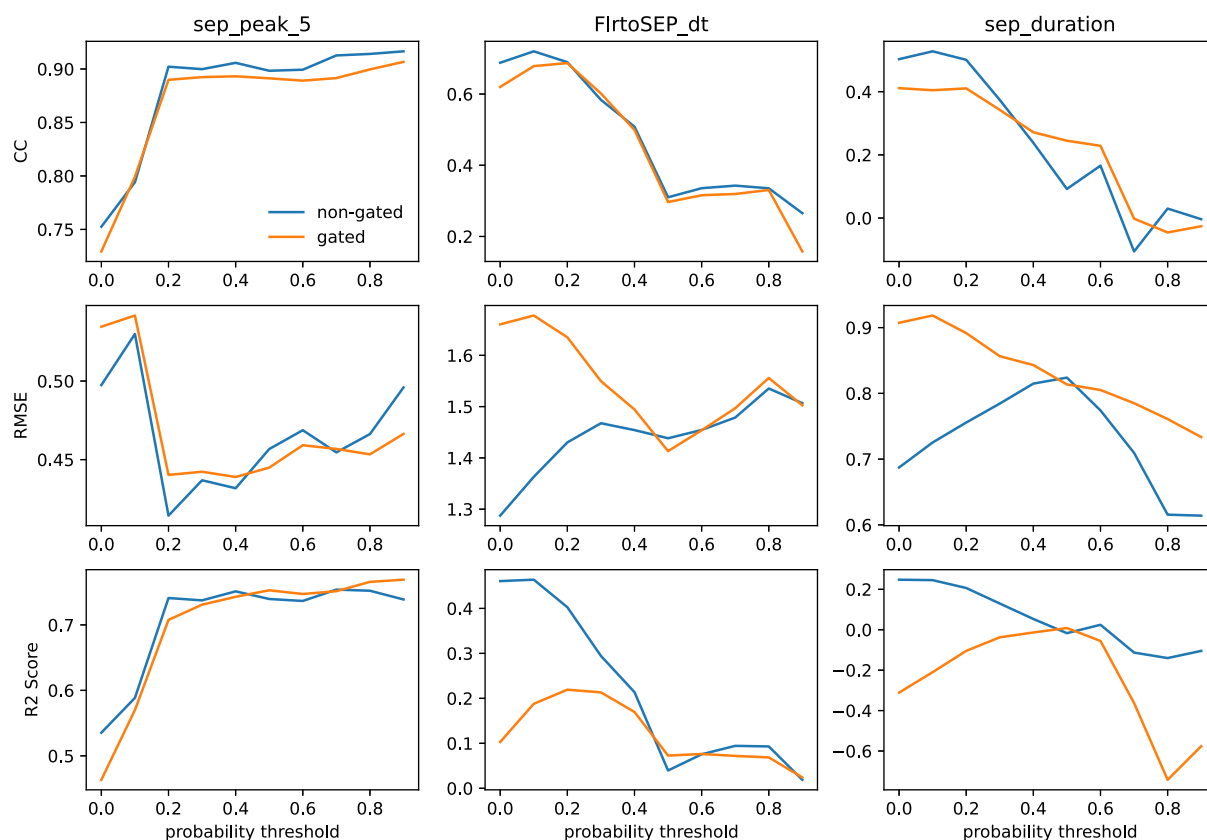
Figure 7. Same as Figure 6 except that here the Regression results are shown for solar energetic particles (SEP) peak flux in highest energy range, SEP onset time, and duration.

in the IP medium would likely improve the prediction of SEP event durations. This problem should be minimized further as we acquire more data across the range for each regression target. Also, having the model-ensemble members trained on a balanced data set, we plan to perform interpretability analysis in the future to decrypt the inner working of the models. Additionally, we would like to make further refinements in the generation of model-ensemble in the future and find the convergence point for the estimated uncertainties.

Data Availability Statement

The data set used in this paper comes from public repositories (e.g., Heliospheric Event Knowledgebase (HEK), JSOC) and are rigorously described in MEMPSEP-III (Moreland et al., 2024). The AI/ML-ready data is publicly available through Zenodo under https://doi.org/10.5281/zenodo.10044864. The codebase developed for this work is also publicly accessible from https://doi.org/10.5281/zenodo.11201195 (Chatterjee et al., 2024b).

References

Alterman, B. L., Desai, M. I., Dayeh, M. A., Mason, G. M., & Ho, G. (2023). Solar cycle variation of 0.3–1.29 mev nucleon 1 heavy ion composition during quiet times near 1 au in solar cycles 23 and 24. *The Astrophysical Journal*, *952*(1), 42. https://doi.org/10.3847/1538-4357/acd24a

Anastasiadis, A., Papaioannou, A., Sandberg, I., Georgoulis, M., Tziotziou, K., Kouloumvakos, A., & Jiggens, P. (2017). Predicting flares and solar energetic particle events:The forspef tool. *Solar Physics*, *292*(9), 134. https://doi.org/10.1007/s11207-017-1163-7

Bain, H. M., Steenburgh, R. A., Onsager, T. G., & Stitely, E. M. (2021). A summary of national oceanic and atmospheric administration space weather prediction center proton event forecast performance and skill. *Space Weather*, *19*(7), e2020SW002670. https://doi.org/10.1029/2020SW002670

Beck, P., Latocha, M., Rollet, S., & Stehno, G. (2005). Tepc reference measurements at aircraft altitudes during a solar storm. *Advances in Space Research*, *36*(9), 1627–1633. https://doi.org/10.1016/j.asr.2005.05.035

Belov, A., Garcia, H., Kurt, V., Mavromichalaki, H., & Gerontidou, M. (2005). Proton enhancements and their relation to the x-ray flares during the three last solar cycles. *Solar Physics*, *229*(1), 135–159. https://doi.org/10.1007/s11207-005-4721-3

Cane, H. V. (1988). The large-scale structure of flare-associated interplanetary shocks. *Journal of Geophysical Research*, *93*(A1), 1–6. https://doi.org/10.1029/JA093iA01p00001

Chatterjee, S., Dayeh, M., Munoz-Jaramillo, A., Bain, H., Moreland, K., & Hart, S. (2024a). MEMPSEP-I. Forecasting the probability of solar energetic particle event occurrence using a multivariate ensemble of convolutional neural networks. *Space Weather*. https://doi.org/10.1029/2023SW003568

Chatterjee, S., Dayeh, M., Munoz-Jaramillo, A., Moreland, K., Bain, H., & Hart, S. (2024b). MEMPSEP codebase (v1.0.1). *Zenodo*. https://doi.org/10.5281/zenodo.11201195

Cohen, C. M. S., Stone, E. C., Mewaldt, R. A., Leske, R. A., Cummings, A. C., Mason, G. M., & von RosenvingeWiedenbeck, T. T. M. E. (2005). Heavy ion abundances and spectra from the large solar energetic particle events of october-november 2003: Heavy ion abundances and spectra. *Journal of Geophysical Research*, *110*(A9). https://doi.org/10.1029/2005JA011004

Cucinotta, F. A. (2022). Flying without a net: Space radiation cancer risk predictions without a gamma-ray basis. *International Journal of Molecular Sciences*, *23*(8), 4324. https://doi.org/10.3390/ijms23084324

Dayeh, M. A., Desai, M. I., Dwyer, J. R., Rassoul, H. K., Mason, G. M., & Mazur, J. E. (2009). Composition and spectral properties of the 1 au quiet-time suprathermal ion population during solar cycle 23. *The Astrophysical Journal*, *693*(2), 1588–1600. https://doi.org/10.1088/0004-637x/693/2/1588

Dayeh, M. A., Desai, M. I., Ebert, R. W., Elliott, H., Farahat, A., Kozarev, K., & Li, G. (2018). What causes the variability in the properties of energetic storm particle (ESP) events? *Journal of Physics: Conference Series*, *1100*(1), 012008. https://doi.org/10.1088/1742-6596/1100/1/012008

Dayeh, M. A., Desai, M. I., Kozarev, K., Schwadron, N. A., Townsend, L. W., PourArsalan, M., et al. (2010). Modeling proton intensity gradients and radiation dose equivalents in the inner heliosphere using emmrem: May 2003 solar events. *Space Weather*, *8*(11). https://doi.org/10.1029/2009sw000566

Dayeh, M. A., Desai, M. I., Mason, G. M., Ebert, R. W., & Farahat, A. (2017). Origin and properties of quiet-time 0.11–1.28 mev nucleon1 heavy-ion population near 1 au. *The Astrophysical Journal*, *835*(2), 155. https://doi.org/10.3847/1538-4357/835/2/155

Decker, R. B. (1981). The modulation of low-energy proton distributions by propagating interplanetary shock waves: A numerical simulation. *Journal of Geophysical Research*, *86*(A6), 4537–4554. https://doi.org/10.1029/JA086iA06p04537

Desai, M., & Giacalone, J. (2016). Large gradual solar energetic particle events. *Living Reviews in Solar Physics*, *13*(1), 3. https://doi.org/10.1007/s41116-016-0002-5

Desai, M., Mason, G., Mazur, J., & Dwyer, J. (2006). Solar cycle variations in the composition of the suprathermal heavy-ion population near 1 au. *The Astrophysical Journal Letters*, *645*(1), L81–L84. https://doi.org/10.1086/505935

Desai, M. I., Mason, G. M., Dayeh, M. A., Ebert, R. W., McComas, D. J., Li, G., et al. (2016). Charge-to-mass dependence of heavy ion spectral breaks in large gradual solar energetic particle events. *Journal of Physics: Conference Series*, *767*(1), 012004. https://doi.org/10.1088/1742-6596/767/1/012004

Garcia, H. A. (2004a). Forecasting methods for occurrence and magnitude of proton storms with solar hard x rays: Forecasting proton events with hard x rays. *Space Weather*, *2*(6). https://doi.org/10.1029/2003SW000035

Garcia, H. A. (2004b). Forecasting methods for occurrence and magnitude of proton storms with solar soft x rays: Forecasting proton events. *Space Weather*, *2*(2). https://doi.org/10.1029/2003SW000001

Gloeckler, G. (2003). Ubiquitous suprathermal tails on the solar wind and pickup ion distributions. *AIP Conference Proceedings*, *679*(1), 583–588. https://doi.org/10.1063/1.1618663

Gopalswamy, N., Yashiro, S., Krucker, S., Stenborg, G., & Howard, R. A. (2004). Intensity variation of large solar energetic particle events associated with coronal mass ejections. *Journal of Geophysical Research*, *109*(A12). https://doi.org/10.1029/2004JA010602

Gosling, J., & Pizzo, V. (1999). Formation and evolution of corotating interaction regions and their three dimensional structure. *Space Science Reviews*, *89*(1–2), 21–52. https://doi.org/10.1023/A:1005291711900

Kahler, S. W., & Ling, A. G. (2018). Forecasting solar energetic particle (sep) events with flare x-ray peak ratios. *Journal of Space Weather and Space Climate*, *8*, A47. https://doi.org/10.1051/swsc/2018033

Kasapis, S., Zhao, L., Chen, Y., Wang, X., Bobra, M., & Gombosi, T. (2022). Interpretable machine learning to forecast sep events for solar cycle 23. *Space Weather*, *20*(2), e2021SW002842. https://doi.org/10.1029/2021SW002842

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). https://doi.org/10.1029/2009sw000550

Lara, A., Gopalswamy, N., Nunes, S., Mun˜oz, G., & Yashiro, S. (2003). A statistical study of CMES associated with metric type ii bursts: CMES and metric type ii bursts. *Geophysical Research Letters*, *30*(12). https://doi.org/10.1029/2002GL016481

Lario, D., Richardson, I. G., Aran, A., & Wijsen, N. (2023). High-energy (¿40 mev) proton intensity enhancements associated with the passage of interplanetary shocks at 1 au. *The Astrophysical Journal*, *950*(2), 89. https://doi.org/10.3847/1538-4357/acc9c5

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, M. A. (1983). Coupled hydromagnetic wave excitation and ion acceleration at interplanetary traveling shocks. *Journal of Geophysical Research*, *88*(A8), 6109–6119. https://doi.org/10.1029/JA088iA08p06109

Mason, G. M., Desai, M. I., Cohen, C. M. S., Mewaldt, R. A., Stone, E. C., & Dwyer, J. R. (2006). The role of interplanetary scattering in western hemisphere large solar energetic particle events. *The Astrophysical Journal Letters*, *647*(1), L65–L68. https://doi.org/10.1086/507469

Mason, G. M., Leske, R. A., Desai, M. I., Dwyer, J. R., Mazur, J. E., Mewaldt, R. A., et al. (2008). Abundances and energy spectra of corotating interaction region heavy ions. *AIP Conference Proceedings*, *1039*(1), 101–110. https://doi.org/10.1063/1.2982430

Mason, G. M., Li, G., Cohen, C. M. S., Desai, M. I., Haggerty, D. K., Leske, R. A., & Zank, G. P. (2012). Interplanetary propagation of solar energetic particle heavy ions observed at 1 au and the role of energy scaling. *The Astrophysical Journal*, *761*(2), 104. https://doi.org/10.1088/0004-637x/761/2/104

Mason, G. M., Mazur, J. E., Dwyer, J. R., Jokipii, J. R., Gold, R. E., & Krimigis, S. M. (2004). Abundances of heavy and ultraheavy ions in 3he-rich solar flares. *The Astrophysical Journal*, *606*(1), 555–564. https://doi.org/10.1086/382864

Moreland, K., Dayeh, M., Bain, H. M., Chatterjee, S., Munoz-Jaramillo, A., & Hart, S. (2024). MEMPSEP-III. A machine learning oriented multivariate data set for forecasting solar energetic particle event occurrence and subsequent properties. *Space Weather*. https://doi.org/10.1029/2023SW003765

Ng, C. K., Reames, D. V., & Tylka, A. J. (1999). Effect of proton-amplified waves on the evolution of solar energetic particle composition in gradual events. *Geophysical Research Letters*, *26*(14), 2145–2148. https://doi.org/10.1029/1999GL900459

Papaioannou, A., Anastasiadis, A., Kouloumvakos, A., Paassilta, M., Vainio, R., Valtonen, E., et al. (2018). Nowcasting solar energetic particle events using principal component analysis. *Solar Physics*, *293*(7), 100. https://doi.org/10.1007/s11207-018-1320-7

Posner, A. (2007). Up to 1-hour forecasting of radiation hazards from solar energetic ion events with relativistic electrons. *Space Weather*, *5*(5). https://doi.org/10.1029/2006SW000268

Reames, D. V. (1999). Particle acceleration at the sun and in the heliosphere. *Space Science Reviews*, *90*(3–4), 413–491. https://doi.org/10.1023/A:1005105831781

Richardson, I. G., Mays, M. L., & Thompson, B. J. (2018). Prediction of solar energetic particle event peak proton intensity using a simple algorithm based on cme speed and direction and observations of associated solar phenomena. *Space Weather*, *16*(11), 1862–1881. https://doi.org/10.1029/2018SW002032

Tylka, A. J., Cohen, C. M. S., Dietrich, W. F., Lee, M. A., Maclennan, C. G., Mewaldt, R. A., et al. (2005). Shock geometry, seed populations, and the origin of variable elemental composition at high energies in large gradual solar particle events. *The Astrophysical Journal*, *625*(1), 474–495. https://doi.org/10.1086/429384

Webb, D. F., & Allen, J. H. (2004). Spacecraft and ground anomalies related to the october-november 2003 solar activity. *Space Weather*, *2*(3). https://doi.org/10.1029/2004SW000075

Whitman, K., Egeland, R., Richardson, I. G., Allison, C., Quinn, P., Barzilla, J., et al. (2022). Review of solar energetic particle models. *Advances in Space Research*, *72*(12), 5161–5242. https://doi.org/10.1016/j.asr.2022.08.006

Wijsen, N., Li, G., Ding, Z., Lario, D., Poedts, S., Filwett, R. J., et al. (2023). On the seed population of solar energetic particles in the inner heliosphere. *Journal of Geophysical Research: Space Physics*, *128*(3). https://doi.org/10.1029/2022ja031203

Zhang, J., Temmer, M., Gopalswamy, N., Malandraki, O., Nitta, N. V., Patsourakos, S., et al. (2021). Earth-affecting solar transients: A review of progresses in solar cycle 24. *Progress in Earth and Planetary Science*, *8*(1), 56. https://doi.org/10.1186/s40645-021-00426-7