

Validation of the 3D Physics-Based SEP Forecasting Tool SPARX Using Historical SEP Events

Damini Bhagwath,^{a,*} Timo Laitinen,^a Silvia Dalla^a and Mike Marsh^b

^aUniversity of Central Lancashire, UK

^bMet Office, UK

E-mail: DSBhagwath@uclan.ac.uk, TLMLaitinen@uclan.ac.uk,
SDalla@uclan.ac.uk, mike.marsh@metoffice.gov.uk

The transport of Solar Energetic Particles (SEPs) through the interplanetary space remains a challenging aspect in space physics. The effects of the interplanetary magnetic field and solar wind turbulence on the SEP arrival to Earth has motivated the adoption of physics-based approaches in operational space weather forecasting. 3D physics-based models offer a framework for incorporating the complexity of SEP transport mechanisms. The forecasting tool, SPARX (Marsh et al., 2015), employs a relativistic full-orbit test particle approach to simulate SEP propagation to predict their arrival and help assess the potential hazard they may cause. Detection of a solar flare with a magnitude $> M1.0$ triggers the SPARX system to produce synthetic particle flux profiles from a database of SEP simulations to forecast observations that would be made at 1AU by a spacecraft.

To enhance SPARX's modeling capabilities, we focus here on validating the model through a systematic analysis of recent SEP events. The performance evaluation methodology implemented in Dalla et al. (2018) used a comprehensive list of X-class solar flares between 1997 and 2017 where for every flare, SPARX was operated in forecast mode. The outputs were analysed for SEP event prediction and performance metrics such as Probability of Detection (POD), False Alarm Ratio (FAR) and Critical Success Index (CSI) were calculated. This evaluation methodology will be adapted for events post 2017 to incorporate a larger dataset and thereby enabling a statistically robust validation of the tool. We chose a proton energy threshold of > 20 MeV for maintaining consistency for model evaluation. We present preliminary analysis of SPARX-generated synthetic SEP flux profiles against multi-spacecraft observations from SOLO, SOHO, and STEREO-A, assessing its current efficiency in reproducing key event characteristics such as onset times and intensity profiles. We discuss how these findings guide next-stage improvements in SPARX. This work aims to contribute to bridging the gap between 3D physics-based SEP modeling and real-time forecasting, ultimately advancing our capability to assess and mitigate SEP-driven space weather hazards. Future enhancement consists of incorporating cross-field transport of SEPs and evaluating its impact on SPARX's forecasting accuracy.

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland



*Speaker

1. Introduction

Solar Energetic Particles (SEPs) are high-energy particles (electrons, protons and ions) that are accelerated to relativistic energies during solar eruptions. The sudden and unpredictable onset of these high-energy solar energetic particle events poses an increased radiation risk to humans in spaceflights, spacecraft instrumentation, aircrafts and satellite electronics [1].

Complex SEP acceleration and propagation mechanisms have prompted space weather forecasters to adopt a multitude of SEP prediction models such as probabilistic models, empirical models, AI-based models and physics-based models [2]. The Solar Particle Radiation SWx (SPARX) [3] utilises a 3D-physics-based SEP simulation approach within a forecasting system. When a flare $>M1.0$ is detected, SPARX uses simulations of an extended injection region that represents a coronal mass ejection-driven shock. It uses a pre-generated database of run outputs with varying proton injection locations, comprising 30 model runs for each IMF polarity. SPARX currently implements a monopolar IMF for each forecast where a run describes a distinct injection tile at a specific central latitude, with particles tracked for 100 hours. 100,000 protons get injected in each tile with energies from 10–400 MeV. The underlying numerical model used to produce these simulations is the relativistic full-orbit test particle code developed by Dalla et al. [4], which was initially used for particle acceleration studies during magnetic reconnection but later adapted to study heliospheric propagation of SEPs [5]. The combined particle outputs from the selected tiles are then used to construct synthetic SEP flux profiles at 1 AU.

A forecasting tool's ability to accurately predict an SEP event's onset time, duration and peak flux is crucial for determining potential for harmful radiation exposure. SPARX generates flux profiles of SEP intensities versus time for protons >10 MeV and 60 MeV. In this paper, we focus on model validation of the current version of SPARX through a systematic analysis of historic and recent SEP events. We also present preliminary analysis of SPARX-generated synthetic SEP flux profiles against observations from SOHO where its current efficiency in reproducing key event characteristics such as onset times and intensity profiles is assessed.

2. Model evaluation

2.1 Observations and analysis

SPARX's performance has been previously evaluated by Dalla et al., [4] whose methodology is primarily adapted here and further extended for an in depth performance evaluation of the current SPARX model. The SEPv3 reference dataset [6, 7] is the primary dataset for SEPs and event lists such as NOAA SEP event list [8] and CRR2010 [9] were referenced to support benchmarking and statistical validation. The SEPv3 dataset offers a continuous, high-quality uniform dataset that is well suited for long-term statistical analysis and skill score evaluation. A list of X-class solar flares between 1 January 1997 and 31 December 2017 was compiled. The X-class flares were obtained from the Heliophysics Events Knowledgebase (HEK), which provides an organized system for the collection and dissemination of heliophysical data [10]. This HEK-derived flare list comprises 174 X-class flares and serves as the basis for identifying associated SEP events in the SEPv3 reference dataset. To aid in the classification of ambiguous SEP events in the observational datasets such as SEPv3 with significant noise due to low counting statistics, we adopted an added

requirement that the SEP event be visible at > 20 MeV proton threshold. This choice allowed for greater consistency across different spacecraft datasets, SEP event lists and it enabled more robust discrimination in marginal or ambiguous cases. Criteria for identifying SEP events was adapted from Swalwell et al.[11] where the SEP events classified are based on a $2.5\times$ increase above the quiet-time background along with a qualitative assessment of elevated pre-event conditions.

In the adapted framework, the methodology is extended by introducing a quantitative criterion: for cases where the pre-flare proton flux background is already elevated, an SEP event is identified if the post-flare flux increases by at least $1.5\times$ the immediate pre-flare level. This was done to remove any possible ambiguity arising from another solar flare being responsible for ongoing activity at the time of the flare being inspected. To complement the SEPTEM-based model evaluation, extended validation of SPARX was then performed using >10 MeV proton flux data accessed via NASA's Integrated Space Weather Analysis (ISWA) system [13], through the Heliophysics Application Programming Interface (HAPI) for the time period 2018 to 2025.

Of 174 flares, 160 flares were retained and 8 flares excluded owing to SEP flux elevation at the time of the flare. SPARX forecasts were evaluated at the >10 MeV integral channel. For the duration of a forecasted event, if the peak proton flux >10 MeV exceeded the two chosen thresholds, a decision of a positive forecast was made. The thresholds taken into account were, the NOAA threshold of $F_{10} = 10$ pfu and another lower threshold of $F_1 = 1$ pfu. The peak flux profiles forecast from SPARX were then verified and benchmarked against SEP event observations from SEPTEM v3 dataset for the same list of X class flares. The > 10 MeV proton integral channel was used for generating the observed SEP event list.

Historical SEP event observations were also analysed using the Serpentine data loader [12], as benchmarks for evaluation of simulations, including both historical and recent SEP events. Figures 1 and 2 illustrate an ambiguous case flagged during compilation of the SEP event list. This was a complex period, as multiple flares occurred close in time and produced overlapping SEP events. Figure 1 shows the SPARX forecast flux profile overlaid on ISWA GOES observations for the event on 5 September 2024.

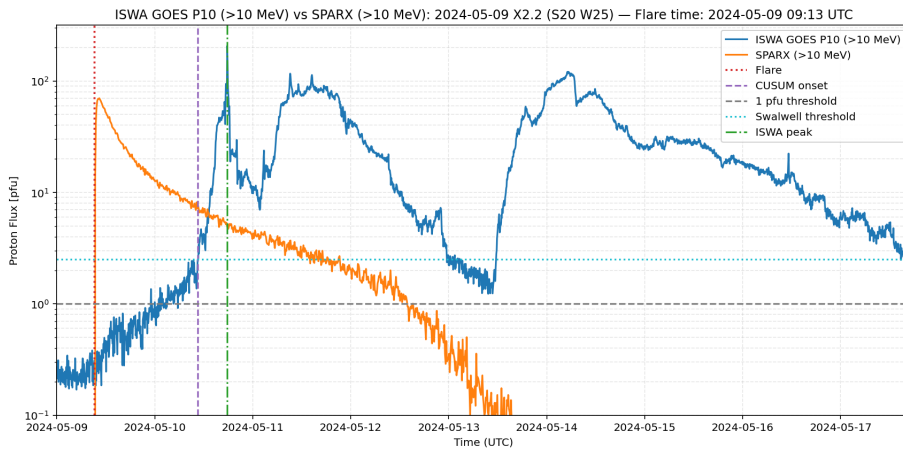


Figure 1: SPARX (orange) flux profile overlaid on ISWA GOES (blue) plot for the SEP event on 2024-05-09., flare location = S20W25, flare class = X2.2

Figure 2 consists of the plot for SOHO/ERNE 25–32 MeV proton flux threshold which was

used to confirm the SEP event occurrence. Analyzing historical SEP events using data from SOHO provides a multi-perspective observational basis for benchmarking. The analysis allowed for a detailed understanding of various characteristics of an SEP event ranging from rise phase to connectivity to the solar source. An extension of the SEP event list with SOHO, SOLO and STEREO observations is planned for the near future to increase the statistics of SPARX validation.

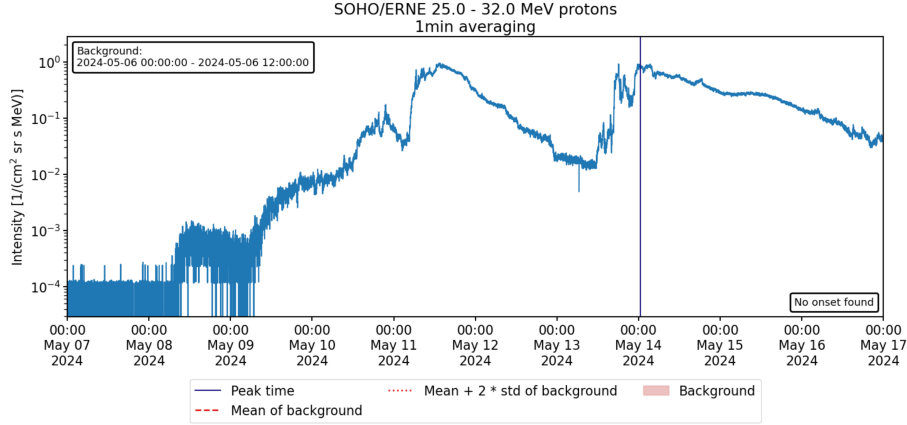


Figure 2: SOHO/ERNE 25–32 MeV proton flux time profile for 3 sep consecutive SEP events (2024-05-11 to 2024-05-16).

3. Results

Contingency table analysis was used to quantify the predictive performance of SPARX's forecasts, as detailed in Table 1. Forecast results are classified as: hits (a) when an SEP event was both forecast and observed; misses (c) when an SEP event was observed but was not forecast; false alarms (b) when an SEP forecast was issued but no event was observed; and correct negatives (d) when no SEP event was forecast and neither was it observed.

| Forecast \ Observed | SEP Occurred | No SEP Occurred |
|----------------------|-------------------|------------------------------|
| SEP Forecast (Yes) | <i>a</i> (Hits) | <i>b</i> (False Alarms) |
| No SEP Forecast (No) | <i>c</i> (Misses) | <i>d</i> (Correct Negatives) |

Table 1: Contingency table for SEP event forecasts

Skill score analysis was then performed (Table 2) for quantification of the performance of the current model. The bias score (BIAS) quantifies under forecasting if the score is <1 and over forecasting if the score is >1 . The Probability of Detection (POD) represents the ratio of SEP events that were forecast correctly to the total number of observed events. It serves as an indicator of the model's sensitivity, with a value of 1 denoting complete success in identifying all observed SEP events. POD does not account for false alarms and therefore False Alarm Ratio (FAR), Probability of False Detection (POFD) and Critical Success Index (CSI) are necessary to be calculated. POFD which is also known as the false alarm rate measures the fraction of actual "no" events that were forecast as "yes". FAR indicates the proportion of positive forecasts that were incorrect. For both

POFD and FAR, the perfect score is denoted by 0. CSI measures the agreement between predicted and observed "yes" events, accounting for both misses and false alarms. A CSI of 1 indicates perfect correspondence.

| Metric | Equation | Perfect Score |
|---------------------------------------|-----------------------|---------------|
| Bias | $\frac{a + b}{a + c}$ | 1 |
| POD (Probability of Detection) | $\frac{a}{a + c}$ | 1 |
| FAR (False Alarm Ratio) | $\frac{b}{a + b}$ | 0 |
| POFD (Probability of False Detection) | $\frac{b}{b + d}$ | 0 |
| CSI (Critical Success Index) | $\frac{a}{a + b + c}$ | 1 |

Table 2: Forecast performance metrics derived from the contingency table, with corresponding perfect score values.

These skill scores provide a first-level assessment of contingency based metrics. Additional quantitative metrics such as the Heidke Skill Score (HSS) and True Skill Score (TSS) were further derived. The Hanssen-Kuipers Discriminant also known as True Skill Statistic (TSS), is a measure of how efficiently the model discriminates between the observed "yes" and "no" events. It ranges from -1 to 1 where 1 is the perfect score and 0 is no skill. HSS is used to measure the accuracy of correct forecasts relative to random chance. The perfect score is 1 while the range is from -1 to 1. The contingency tables corresponding to both the NOAA threshold F_{10} and the lower threshold F_1 are presented in Tables 3a and 3b, respectively. The SEPv3 reference dataset covering the period 1997 to 2017 was used in this analysis.

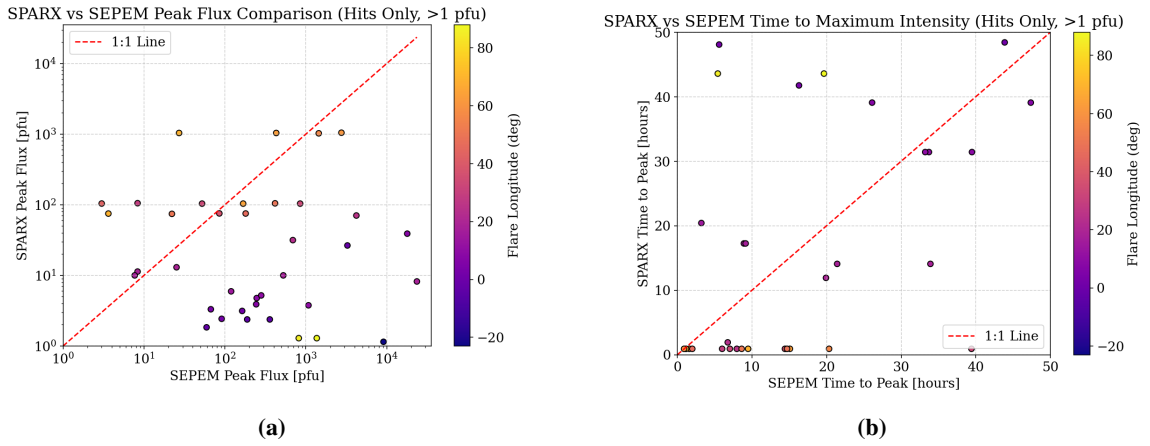
The associated skill scores, derived using the equations outlined in Table 2, are subsequently reported in Table 4. The bias score remains > 1 for both F_{10} and F_1 and indicates overforecasting, especially for lower thresholds. This is confirmed by the presence of some skill at F_1 . The Heidke Skill Score (HSS) for the F_1 threshold was found to be 0.22, indicating that the SPARX model performs better than chance in distinguishing SEP events from non-events. For events where SPARX successfully predicted the occurrence of an SEP event, a comparison was made between the forecast and observed proton flux profiles. Figure 3(a) exhibits forecast accuracy variance with source location. Figure 3(b) shows the time to peak fluxes for SEPv3 vs SPARX. The peak flux is forecast better for well-connected longitudes.

Table 3: Contingency tables for SPARX forecasts for F_{10} and F_1 thresholds, evaluated against the SEPTEM event list (1997–2017).

| (a) F_{10} Threshold | | | (b) F_1 Threshold | | |
|------------------------|------------------------|-----|---------------------|------------------------|-----|
| hits = 16 | false alarms = 31 | 47 | hits = 37 | false alarms = 46 | 83 |
| misses = 30 | correct negatives = 83 | 113 | misses = 17 | correct negatives = 60 | 77 |
| 46 | 114 | 160 | 54 | 106 | 160 |

Table 4: SPARX skill scores for F_{10} and F_1 thresholds.

| Threshold | BIAS | POD | FAR | POFD | CSI | TSS | HSS |
|-----------|------|------|------|------|------|------|------|
| F_{10} | 1.02 | 0.35 | 0.66 | 0.27 | 0.21 | 0.08 | 0.08 |
| F_1 | 1.54 | 0.69 | 0.55 | 0.43 | 0.37 | 0.25 | 0.22 |

**Figure 3:** (a) SPARX peak flux forecast performance (1 pfu threshold) plotted against SEPTEM observations, with source longitudes indicated by color. This visualisation highlights the dependence of forecast accuracy on longitudinal connectivity. (b) Time to peak intensity by SPARX vs SEPTEM time to peak intensity for >10 MeV protons.

The classification methodology used in the SEPTEM validation in the previous section based on flux thresholds was applied here to ISWA-derived data. The time range was from 1 January 2018 to 3 March 2025. The final X class flare list consisted of 44 flares. From the SPARX forecast decisions and ISWA-derived event classifications, a contingency table was constructed (Table 5 and 6). The use of ISWA data reinforces the trend observed with SEPTEM validation: model performance improves with relaxed thresholds, but at the cost of higher false alarm rates. Further integration of uncertainty quantification methods and improved turbulence parameterisation could enhance performance at operationally relevant thresholds.

Lastly, despite variations in ISWA's calibration procedures, SPARX achieves a TSS of 0.418 and HSS of 0.414 at the F_1 threshold during the ISWA validation period. These results indicate that SPARX retains predictive capability when applied to post-2017 operational datasets.

Table 5: Contingency tables for SPARX forecasts for F_{10} and F_1 thresholds, evaluated against the ISWA event list (2018–2025).

| (a) F_{10} Threshold | | | (b) F_1 Threshold | | |
|------------------------|------------------------|----|---------------------|------------------------|----|
| hits = 5 | false alarms = 6 | 11 | hits = 14 | false alarms = 4 | 18 |
| misses = 15 | correct negatives = 18 | 33 | misses = 9 | correct negatives = 17 | 26 |
| | | 20 | | | 23 |
| | | 24 | | | 21 |
| | | 44 | | | 44 |

Table 6: SPARX skill scores for F_{10} and F_1 thresholds.

| Threshold | BIAS | POD | FAR | POFD | CSI | TSS | HSS |
|-----------|------|-------|-------|------|-------|-------|-------|
| F_{10} | 0.55 | 0.25 | 0.545 | 0.25 | 0.192 | 0.00 | 0.00 |
| F_1 | 0.78 | 0.609 | 0.222 | 0.19 | 0.519 | 0.418 | 0.414 |

4. Conclusion

Evaluation through contingency-based metrics confirms that SPARX performs better at lower thresholds. Notably, SPARX retains predictive capability after 2017, as demonstrated through ISWA-derived SEP events. Further analysis will be carried out by integrating probabilistic metric analysis alongside current deterministic analysis which would provide a more comprehensive evaluation framework for both the current version of SPARX and the updated version. SPARX exhibits measurable skill at F_1 threshold and marginal skill at the operational NOAA threshold (F_{10}). Bias score indicates over-forecasting in the SEP-EM derived SEP event sample and slight under-forecasting in the ISWA sample.

Event analysis agrees with the longitudinal dependence stated in [4]. For western, well-connected sources SPARX tends to underestimate time-to-peak, whereas eastern events show closer agreement in timing but under prediction of peak fluxes. These behaviors will be further investigated by enhancing turbulence parameterization within SPARX.

Acknowledgments

DB acknowledges the support from the University of Lancashire’s Doctoral Training Centre. TL and SD acknowledge support from the UK Science and Technology Facilities Council (STFC) through grant ST/Y002725/1. We acknowledge the Community Coordinated Modeling Center (CCMC) at Goddard Space Flight Center for the use of the Integrated Space Weather Analysis (ISWA) system, <https://iswa.gsfc.nasa.gov>.

References

- [1] Committee on the Evaluation of Radiation Shielding for Space Exploration, *Managing Space Radiation Risk in the New Era of Space Exploration*, The National Academies Press, Washington, D.C., 2008

- [2] Whitman, K., Egeland, R., Richardson, I.G., Allison, C., Quinn, P., Barzilla, J., Kitiashvili, I., Sadykov, V., Bain, H.M., Dierckxsens, M., *et al.*, *Review of Solar Energetic Particle Prediction Models*, *Adv. Space Res.* **72** (2023) 5161–5242, doi:10.1016/j.asr.2022.08.006
- [3] M. S. Marsh, S. Dalla, M. Dierckxsens, T. Laitinen, and N. Crosby, *SPARX: A modeling system for Solar Energetic Particle Radiation Space Weather forecasting*, *Space Weather* **13** (2015) 386
- [4] S. Dalla, B. Swalwell, M. Battarbee, M. Marsh, T. Laitinen, and S. Proctor, *Application of Test Particle Simulations to Solar Energetic Particle Forecasting*, in *Space Weather of the Heliosphere: Processes and Forecasts*, pp. 268–271, 2018
- [5] Marsh, M. S., Dalla, S., Kelly, J., and Laitinen, T., *Drift-induced Perpendicular Transport of Solar Energetic Particles*, *Astrophys. J.* **774** (2013) 4, doi:10.1088/0004-637X/774/1/4
- [6] Jiggins, P., Raukunen, O., Sandberg, I., Mutch, S., Aminimalragia-Giamini, S., Vainio, R., Heynderickx, D., Papadimitriou, C., Rodriguez, J., and Aran, A., *The SEP-EM Reference Data Set – Version 3*, in *44th COSPAR Scientific Assembly*, 16–24 July 2022, p. 1163
- [7] Crosby, N., Heynderickx, D., Jiggins, P., Aran, A., Sanahuja, B., Truscott, P., Lei, F., Jacobs, C., Poedts, S., Gabriel, S., Sandberg, I., Glover, A., and Hilgers, A., *SEP-EM: A tool for statistical modeling the solar energetic particle environment*, *Space Weather* **13** (2015), doi:10.1002/2013SW001008
- [8] https://cdaw.gsfc.nasa.gov/CME_list/sepe/
- [9] Cane, H.V., Richardson, I.G., and von Roseninge, T.T., *A study of solar energetic particle events of 1997–2006: Their composition and associations*, *J. Geophys. Res.* **115** (2010) A08101, doi:10.1029/2009JA014848
- [10] Hurlburt, N., Cheung, M., Schrijver, C., Chang, L., Freeland, S., Green, S., Heck, C., Jaffey, A., Kobashi, A., Schiff, D., Serafin, J., Seguin, R., Slater, G., Somani, A., and Timmons, R., *Heliophysics Event Knowledgebase for the Solar Dynamics Observatory (SDO) and Beyond*, *Solar Phys.* **275** (2012) 67–78, doi:10.1007/s11207-010-9624-2
- [11] Swalwell, B., Dalla, S., and Walsh, R.W., *Solar Energetic Particle Forecasting Algorithms and Associated False Alarms*, *Solar Phys.* **292** (2017) 173, doi:10.1007/s11207-017-1196-y
- [12] Palmroos, C., Gieseler, J., Dresing, N., Morosan, D.E., Asvestari, E., Yli-Laurila, A., Price, D.J., Valkila, S., and Vainio, R., *Solar Energetic Particle Time Series Analysis with Python*, *Front. Astron. Space Sci.* **9** (2022) 395, doi:10.3389/fspas.2022.1073578
- [13] Weigel, R.S., Vandegriff, J., Faden, J., King, T., Roberts, D.A., Harris, B., *et al.*, *HAPI: An API standard for accessing Heliophysics time series data*, *J. Geophys. Res. Space Phys.* **126** (2021) e2021JA029534, doi:10.1029/2021JA029534