Model Validation using Historical SEP Event Analysis of the 3D Physics-Based Forecasting Tool SPARX

Damini Bhagwath,[a,*] Timo Laitinen[a], Silvia Dalla[a] and Mike Marsh[b]
[a] University of Lancashire, UK, [b] Met Office, UK

1 Introduction

- Solar Energetic Particles (SEPs) are high-energy electrons, protons, and ions accelerated to relativistic velocities during solar eruptions. They pose a significant radiation hazard to astronauts, spacecraft, and high-altitude aviation. The prediction of their arrival at Earth is complicated by the turbulent nature of the interplanetary magnetic field which governs SEP transport. To address these challenges, models such as SPARX (Solar Particle Radiation SWx; (1)) offer a physics-based approach to operational space weather forecasting. In this work, we focus on validating the forecasting performance of SPARX using a systematic analysis of historical and recent SEP events.
- We adapt and extend the methodology from Dalla et al. (2018) (2), which used a comprehensive set of X-class flares (1997–2017) to evaluate SEP forecasting performance via standard metrics such as Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI).
- We also present preliminary comparisons between SPARX forecasts and observations (SOHO and GOES) to assess model accuracy in reproducing SEP onset times and flux profiles.

2 Space Weather Forecasting tool SPARX

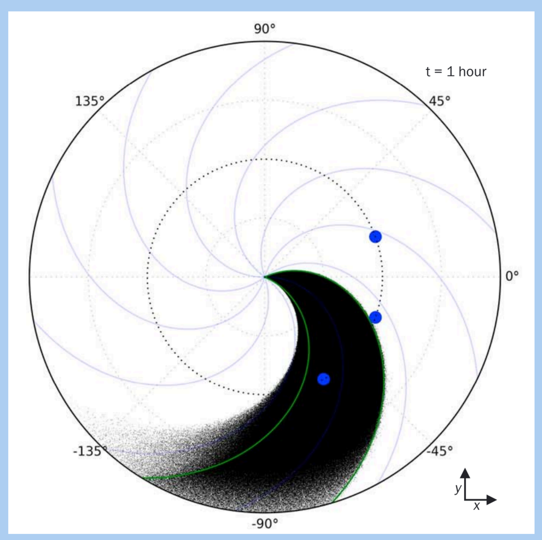

Figure 1: Evolution of the Corotating Solar Energetic Particle Stream due to injection region produced by SPARX (1)

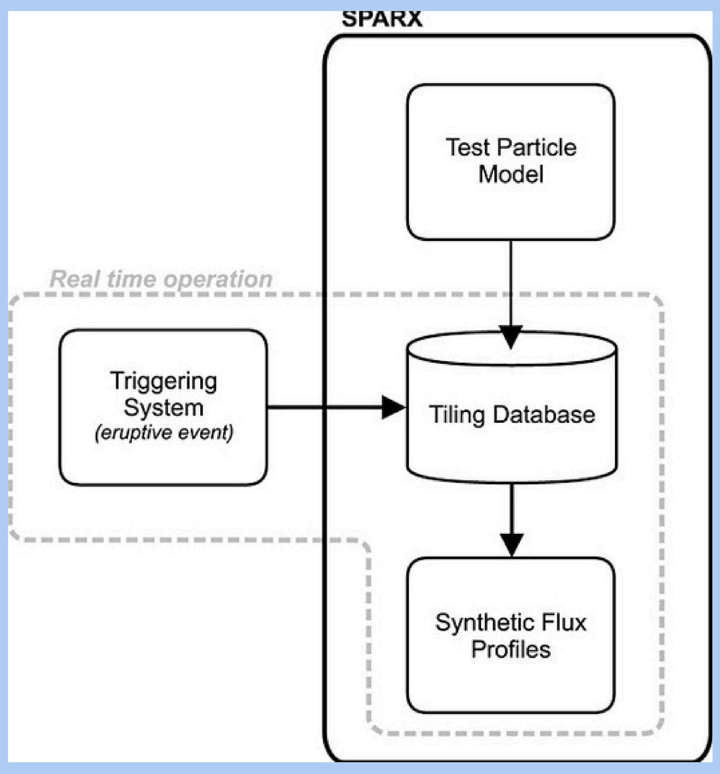

Figure 2: Schematic diagram of SPARX's modelling framework (1)

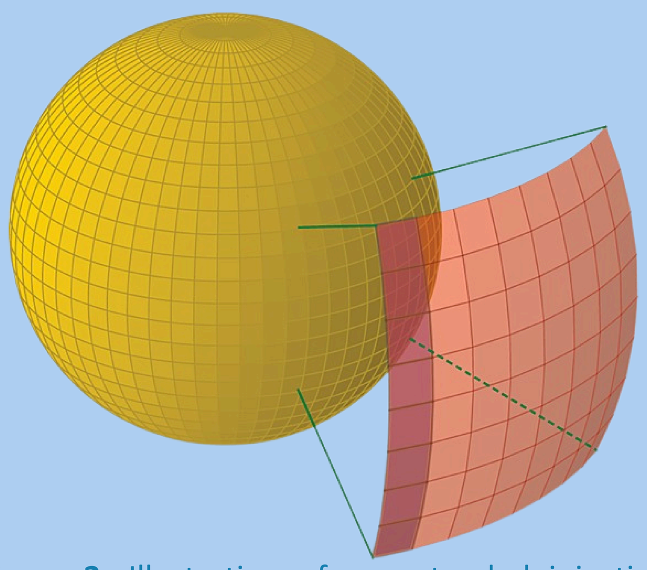

Figure 3: Illustration of an extended injection region at 2 solar radii created in the forecasting system (1)

- Forecasting SEP events accurately is challenging due to the complex particle acceleration and transport processes involved.
- SEPs propagate from the Sun along Parker spiral (Figure 1), and scatter due to turbulence.
- SPARX employs a test particle approach where each simulation follows a large number of particles and these particle trajectories are integrated to produce synthetic flux profiles of SEP events.

1. SEP source characterisation:
Flare intensity = >M1.0, Flare longitude, Flare latitude, Flare peak. The source is a 48° × 48° region

2. Particle Injection:
A large number of test particles injected instantaneously at 2 solar radii from the Sun over energy range 10–400 MeV (Figure 3)

3. Particle Transport:
Simulated protons propagate along the Parker Spiral magnetic field, experiencing drift, scattering (mean free path = 0.3 AU), and deceleration effects

4. Detection and Flux Profiles:
Every time a particle crosses 1 AU, its parameters are recorded to build synthetic SEP flux profiles (>10 MeV, >60 MeV channels) (See Figure 5 (c))

5. Forecast Database:
Flux profiles are gathered into a pre-generated database, SPARX queries the database to rapidly generate forecasts (Figure 2)

- SPARX forecasting system is extensible and can be modified to produce output from different injection spectra, particle species, output flux profile energy ranges and flux profiles at any point in the heliosphere.
- The SPARX model includes some description of perpendicular transport via particle drift.
- In near future, a clear understanding of turbulence parameters will be developed and how they vary across the heliosphere will be incorporated to improve the description of particle propagation.

3 Validation Strategy

- **X-class solar flares list** (1997-2017) obtained from Heliophysics Events Knowledgebase (HEK) (3).
- Flare list served as basis for identifying associated SEP events in SEPEM v3 reference dataset.
- **SEP event identification criteria**: 2.5× increase above quiet-time background, adapted from Swalwell et al.(4).
- Cases where events with pre-flare background already elevated: SEP event was classified if post-flare flux increased by 1.5× the immediate pre-flare level.
- **Ambiguous SEP event classification**: further checked for visibility in > 20 MeV proton threshold.
- Of 174 flares, 160 flares were retained and 8 flares excluded because of SEP flux elevation at the time of flare.
- The **thresholds** taken into account:

1. NOAA threshold of F_{10} =10 pfu
2. F_1= 1 pfu

- Contingency table for SEP event forecasts:

Forecast \ Observed	SEP Occurred	No SEP Occurred
SEP Forecast (Yes)	a (Hits)	b (False Alarms)
No SEP Forecast (No)	c (Misses)	d (Correct Negatives)

- Forecast performance metrics derived from above:

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

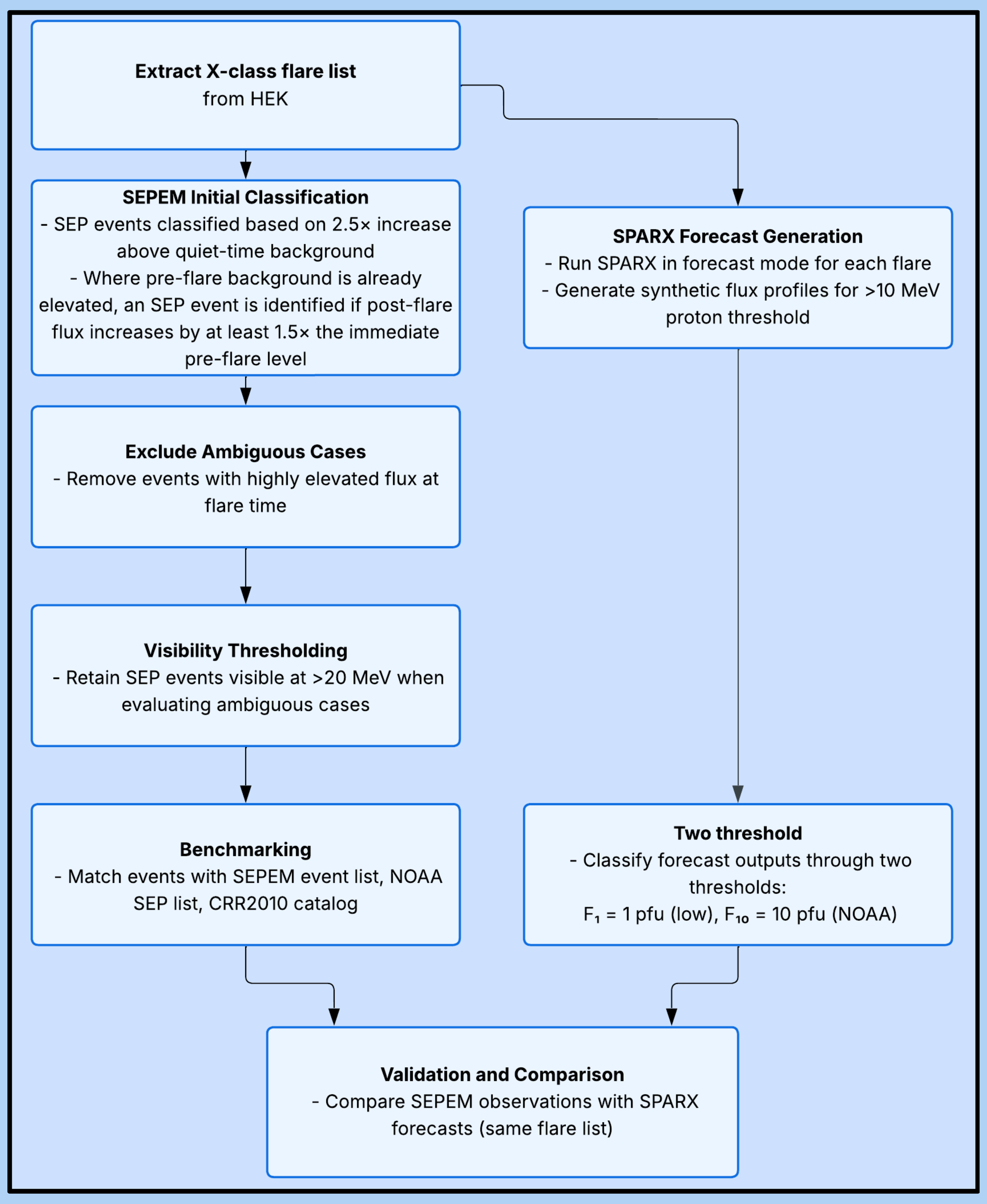

Figure 4: Flowchart summarising the methodology used to validate SPARX forecasts against observed SEP events.

- Extended validation of SPARX was performed using >10 MeV proton flux data accessed via NASA's Integrated Space Weather Analysis (Between 2018-2025) (5) through Heliophysics Application Programming Interface (HAPI).
- Time range was from 1 January 2018 to 3 March 2025. Final X class flare list consisted of 44 flares.

4 Results

- Contingency tables and skill scores constructed for F_{10} and F_1 thresholds using SEPEM and ISWA data (Table 1 and 2).
- **Table 1:**

1. Bias score is > 1 for both F_{10} and F_1 indicating overforecasting, especially for lower thresholds.
2. Heidke Skill Score (HSS) for F_1 was 0.22, indicating that the SPARX model performs better than chance in distinguishing SEP events from non-events.
3. SPARX demonstrates baseline SEP forecast capability over 1997–2017 but suffers from frequent false positives

- **Table 2:**

4. SPARX achieves a TSS of 0.418 and HSS of 0.414 at F_1 during the ISWA validation period. These results indicate that SPARX retains predictive capability when applied to post-2017 operational datasets.

- Figure 5 (a) and (b) showcase SPARX simulation overlaid on SEPEM observation (historical GOES data) for the events on 2001-03-29 and 2012-03-07 respectively.
- Figure 5 (a) evidences SPARX's better performance for western longitude events compared with the case (Figure 5 (b)) where it is an eastern event.

(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

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

Table 1:Contingency tables for SPARX forecasts for F_{10} and F_1 thresholds (top);SPARX skill scores for F_{10} and F_1 thresholds (bottom). Dataset: SEPEM v3 (1997-2017)

(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	24	44		23	21	44

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

Table 2: (As mentioned above) For dataset: GOES operational data from ISWA (2018-2025)

- Further statistics explaining SPARX's forecast accuracy varying with flare longtitude (Figure 5 (c) and (d)) are showcased through the spread observed in flare longitude binning and peak flux comparisons between SEPEM vs SPARX.

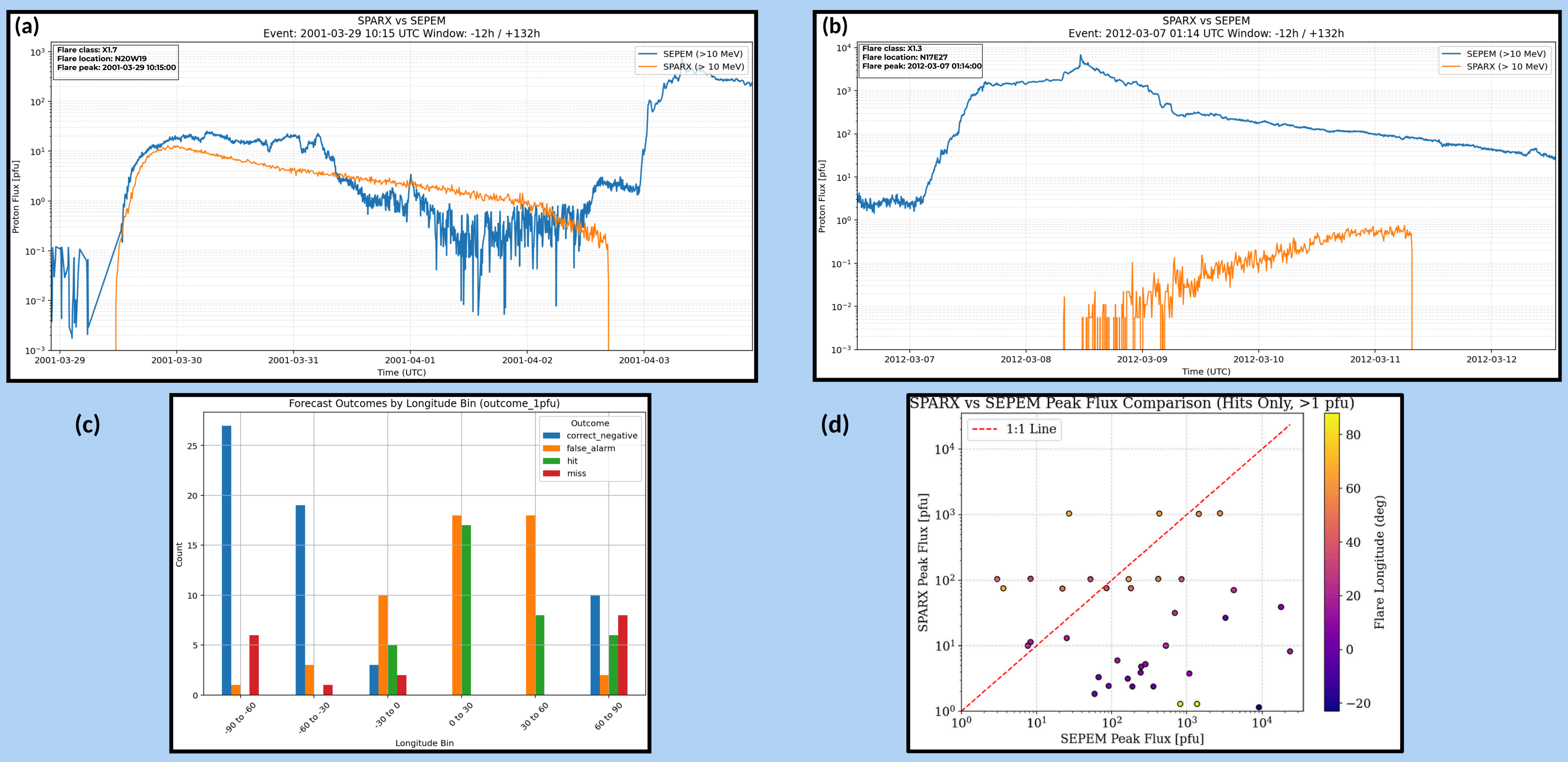

Figure 5: **(a)** SPARX (orange) vs SEPEM (blue) plot for the SEP event on 2001-03-29, flare location = N20W19, flare class = X1.7 **(b)** Same as (a) but for the event 2012-03-07, flare location = N17E27, flare class = X1.3 **(c)** Forecast outcomes by flare longitude bin for SPARX. Bars show number of events in each 30° bin classified as hit, miss etc. **(d)** SPARX peak flux forecast performance vs SEPEM observations

5 Forthcoming Developments

- Constructing ROC curves: Showcasing SPARX's performance skill more accurately.
- Incorporating Uncertainty Quantification: Employing probabilistic metrics within SPARX to provide a more comprehensive evaluation framework.
- Turbulence-Driven Enhancements: Updating SPARX simulations to explicitly model a more detailed description of turbulence.
- Database Expansion: Construction of a new SPARX database consisting of improved turbulence parameters for operational use.

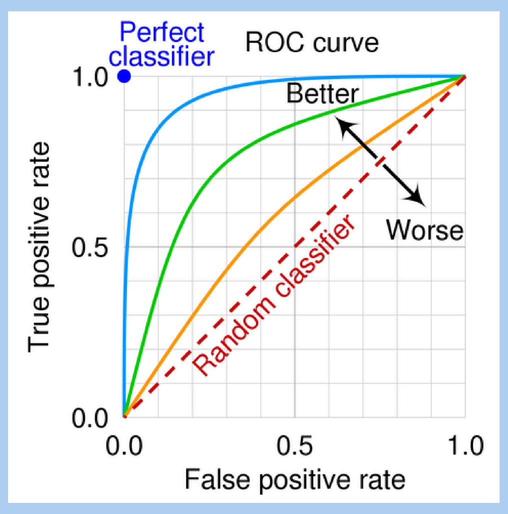

Figure 6: Conceptual illustration (6) of ROC curve showing classifier performance.

Conclusion

Evaluation through contingency-based metrics confirms that SPARX has better performance at lower thresholds, particularly for well-connected events. Notably, SPARX retains predictive capability post-2017, as demonstrated through ISWA-derived SEP events. Our future 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, particularly for eastern SEP events which are currently poorly forecasted.

References

(1): Marsh, M. et al (2015), Space Weather, 10.1002/2014SW001120
(2): Dalla, S. et al. (2018), Space Weather of the Heliosphere: Processes and Forecasts, Proceedings IAU Symposium No. 335 (eds Foullon & Malandraki), p 268, 2018
(3): Hurlburt, N. et al. (2010), Solar Phys 275(1–2), 67–78
(4): Swalwell, B. et al. (2017), Solar Phys 292, 173
(5): Weigel, R. et al. (2022), COSPAR Sci Assem 2022, 3488
(6): Hammouda, M. et al (2024), Nat Hazards 120, 1–23

dsbhagwath@uclan.ac.uk

Jeremiah Horrocks Institute

University of Lancashire