First Results from NASA's SEP Model Validation (SEPVAL) Effort

Kathryn Whitman, Phil Quinn, Ricky Egeland, Luke Stegeman, Clayton Allison on behalf of the ISEP Project
NASA JSC Space Radiation Analysis Group (SRAG)

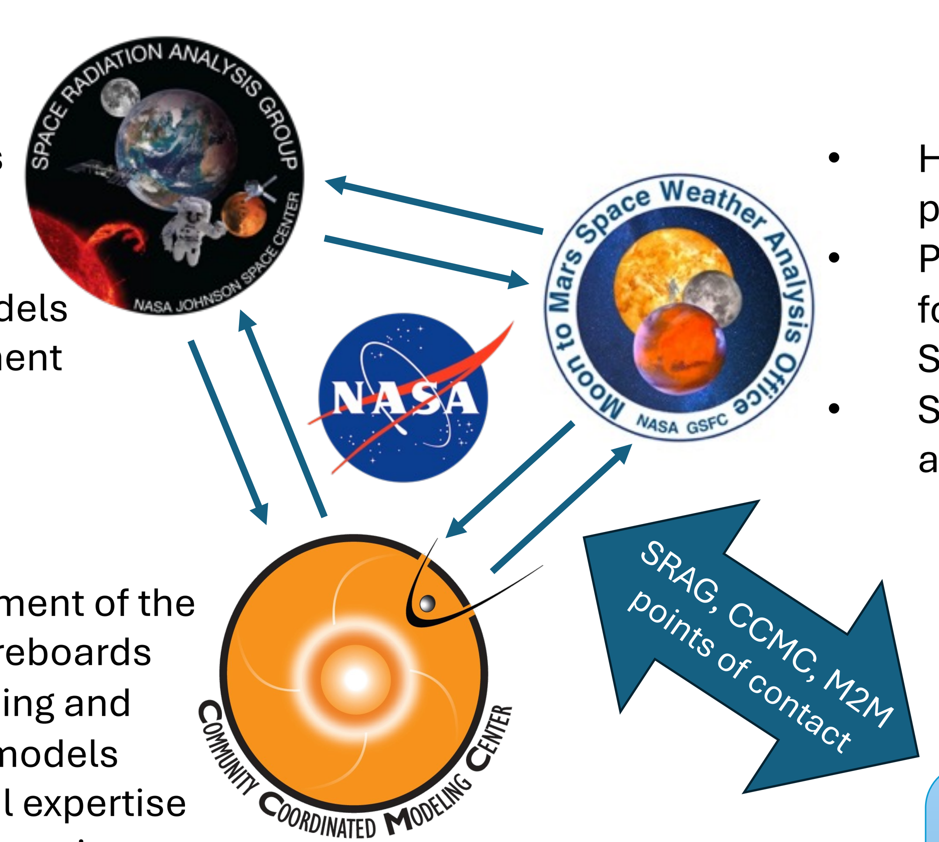
SHINE workshop August 13 – 17, 2024 in Juneau, Alaska

Motivation: The ISEP Project

The Integrated Solar Energetic Proton Event Alert/Warning System (ISEP) project is a collaborative effort to enhance space weather alerting capabilities for SRAG's operations whose mission is to protect astronauts from space radiation.

Ingredients for effective R2O2R2O

- Space Radiation operations
- Expert end users
- Development of SPHINX and validation of models
- Model development



- Human-in-the-loop products
- Provide necessary inputs for models to run on the SEP Scoreboards
- Space weather monitoring and analyses

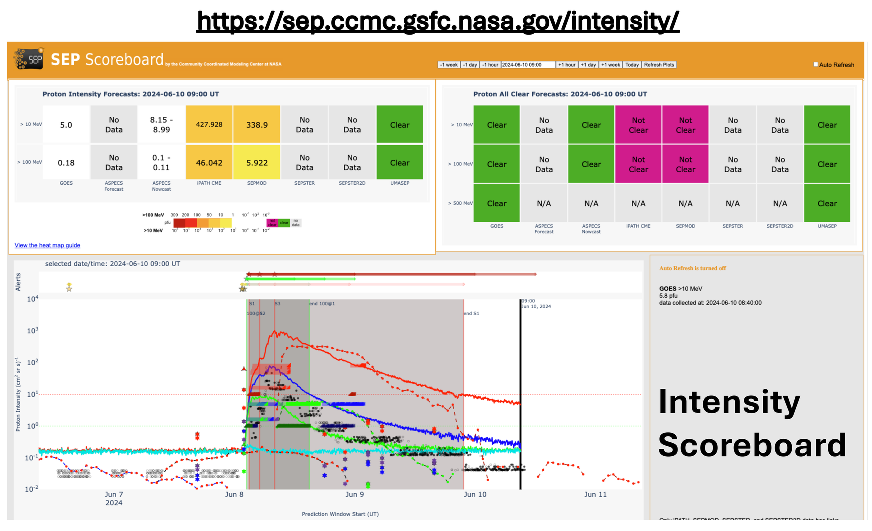
- Development of the SEP Scoreboards
- Onboarding and hosting models
- Technical expertise
- Model expertise

Small, dedicated ISEP grants, technical support and end-user feedback for focused R2O efforts

Model Developers at Research Institutions

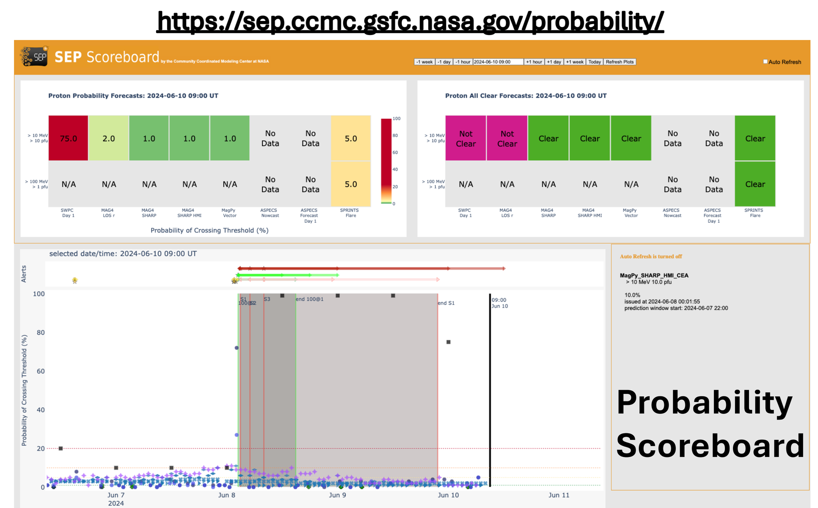
Product: SEP Scoreboards

https://sep.ccmc.gsfc.nasa.gov/intensity/



Intensity Scoreboard

https://sep.ccmc.gsfc.nasa.gov/probability/

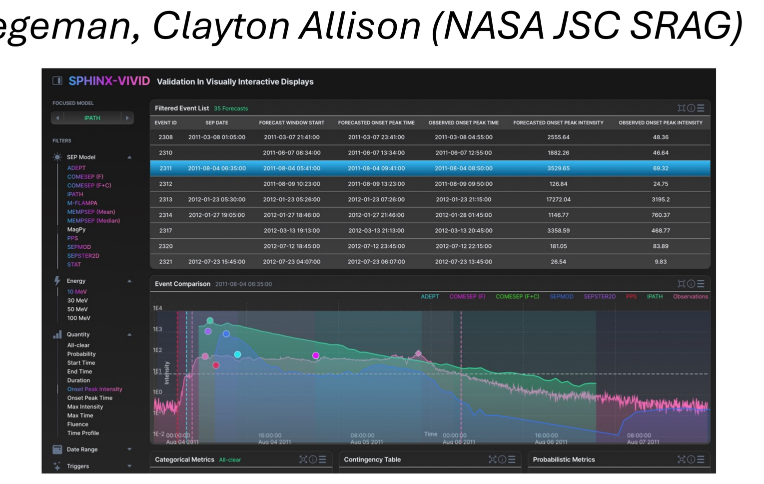


Probability Scoreboard

SPHINX Validation Framework

Solar Particles in the Heliosphere validation INfrastructure for SpWx (SPHINX)
SPHINX Goal: A generalized, automated tool that can validate any kind of forecasted quantity from any type of solar energetic particle (SEP) prediction model.
Developers: *Kathryn Whitman, Ricky Egeland, Luke Stegeman, Clayton Allison (NASA JSC SRAG)*
Github: https://github.com/ktindiana/sphinxval

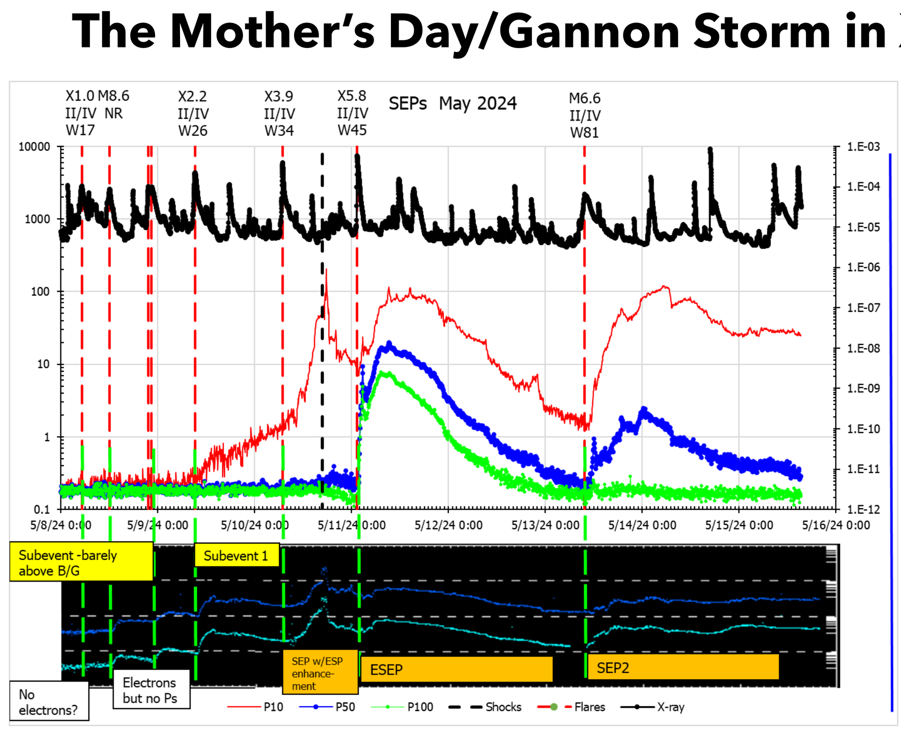
Validation in Visually Interactive Displays (VIVID)
VIVID Goal: Web application for displaying the validation results of SPHINX in a dashboard of interactive plots and tables
Developer: *Phil Quinn (NASA JSC SRAG)*



Availability: Both will be hosted at CCMC in the future.

SEP Scoreboards, Gannon Storms May 2024

The Mother's Day/Gannon Storm in X-Rays and Solar Energetic Particles



Definitions
SPE: >10 MeV exceeds 10 pfu,
ESPE: >100 MeV exceeds 1 pfu

Events
Sub-Event 1
Flare: May 09, 09:13Z, X2.2, W26
Radio: Type II/1004 km/s Type IV

SPE 1
Flare: May 10, 06:54Z, X3.9 W34
Radio: Type II/489 km/s Type IV

Energetic SPE Event 1
Flare: May 11, 01:23Z, X5.8, W45
Radio: Type II/564 km/s Type IV

SPE 2
Flare: May 13, 09:44Z, M6.6, W81
Radio: Type II/683 km/s Type IV

Courtesy Steve Johnson (NASA JSC SRAG)

SRAG Operational Response

G5 (Extreme) Geomagnetic Storm
Most intense since October 2003
Aurora sightings across the globe
Lots of media attention

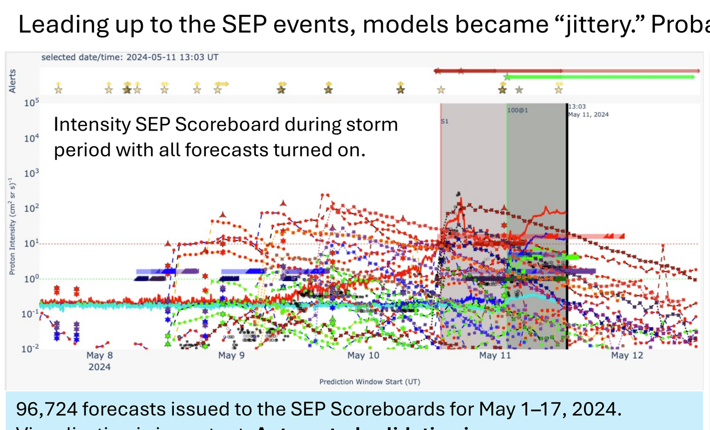
+

Energetic Solar Particle Event (ESPE)
Onset: 02:10 UTC on 11 May
No impact to ISS crew

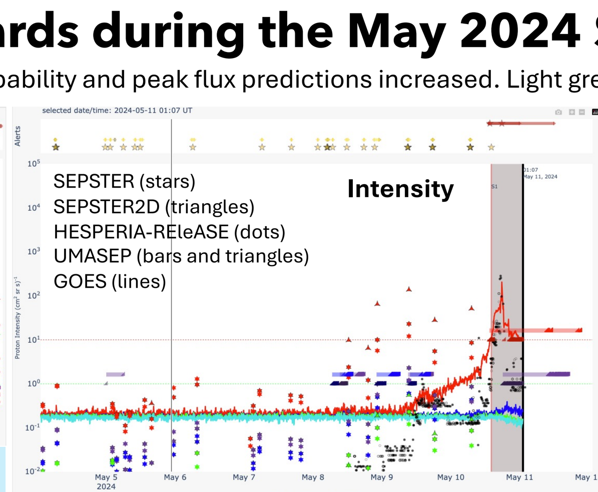
Space weather phenomena triggered **37-hour** continuous SRAG console support

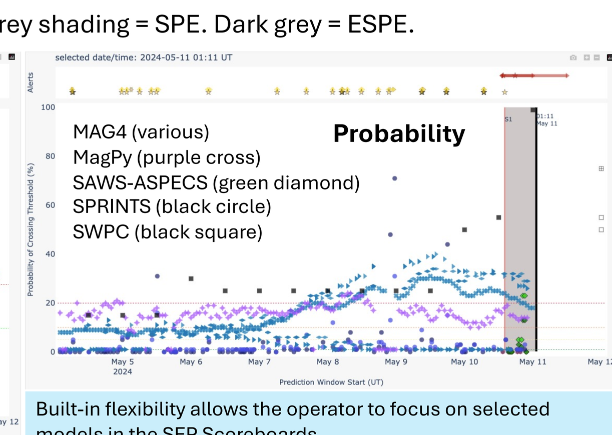
SEP Scoreboards during the May 2024 Storms

Leading up to the SEP events, models became "jittery." Probability and peak flux predictions increased. Light grey shading = SPE. Dark grey = ESPE.



96,724 forecasts issued to the SEP Scoreboards for May 1–17, 2024. Visualization is important. **Automated validation is necessary.**





Built-in flexibility allows the operator to focus on selected models in the SEP Scoreboards.

Automated Validation with SPHINX

SPHINX takes the approach that **model output is considered a forecast if all data input into the model is earlier in time than the observed phenomenon.**
Criteria related to the timing of flares/CMEs and observed proton flux threshold crossings are applied to associate forecasts with observed SEP events.
SPHINX correctly matched forecasts to all SPEs and ESPE.

Model	ESPE (>100 MeV) 2024-05-11 02:10:00
iPATH CME	Miss
SEPSTER	Miss
SEPSTER2D	Miss
SEPMOD	Hit (forecast not produced in RT)
SAWS-ASPECS flare	-25 minutes
SPRINTS Post Eruptive 0-24 hrs	29.55 minutes
UMASEP-100	9.35 minutes

Advanced Warning Time
A subset of SEP Scoreboard models make forecasts for >100 MeV, 1 pfu. 3 models correctly forecasted a hit for the ESPE. AWT is shown in the table. Negative time indicates the forecast was issued after the threshold was already crossed.

SEPVAL 2023 Challenge and Meetings

Following a multi-year validation effort through the SHINE, ISWAT, and ESWW workshops, the SEPVAL challenge and workshops (US and Europe) were focused on validation and SEP forecasting in ops.

SEPVAL 2023 (US)
September 5 - 7, 2023
Southwest Research Institute
San Antonio, TX

SEPVAL 2023 (Europe)
ESWW, Toulouse, France

SEPVAL 2023 organizers:

- Provided a list of challenge time periods and triggers (flares, CMEs) for **33 SEP events and 30 non-events**
- Defined rules of participation to encourage modelers to produce forecasts in a real time-like scenario
- Developed the SPHINX infrastructure and performed the validation
- Made the validation results available to attendees (R2O2R)

SEP model developers:

- Provided forecasts and supplementary information
- Followed the rules of participation
- To provide feedback about the forecast/prediction process and the validation results

SEPVAL Website

Forecasts Received for SEPVAL 2023

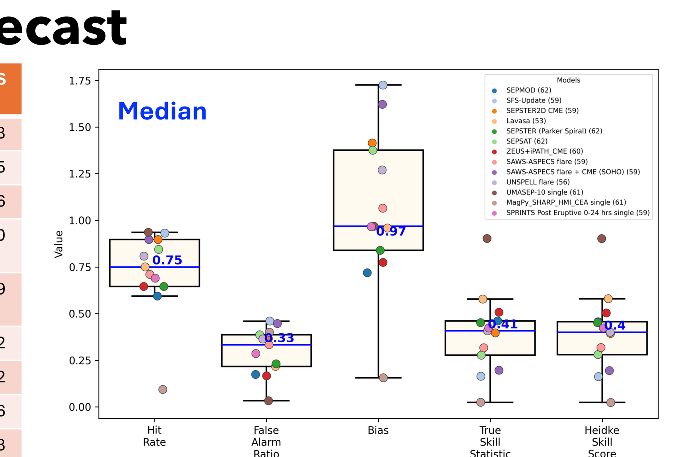
Model	Developer Point of Contact	Affiliation	Method	Energy Channels (MeV)	Forecasted Quantities	# Forecasts Submitted
ADEPT 1hr, 6hr	Stephen White	US Air Force	Empirical	>10	Time Profile	25
COMESEP flare, flare+CME	Mark Dierckxsens	BIRA	Empirical	>10	Probability, Peak	60, 63
cRT+AE10	Ming Zhang	Florida Institute of Technology	Machine Learning	>10	Probability	63
ENLIL+SEPMOD	Janet Luhmann	UC Berkeley	Physics-based	>10, >30, >50, >100	Time Profile	63
Lavasa	Eleni Lavasa	National Observatory of Athens	Machine Learning	>10	All Clear	58
MagPy	David Falconer, Tilaye Tedesse	UA Huntsville, NASA JSC SRAG	Empirical	>10	Probability	2182
MEMPSEP Mean, Median	Subhamoy Chatterjee	Southwest Research Institute	Machine Learning	>10	Probability	60, 60
MFLAMPA	Igor Sokolov	University of Michigan	Physics-based	>10, >30, >50, >100	Time Profile	9
PPS (SFS Update)	Stephen White	US Air Force	Empirical	>10, >100	Peak Flux	61
SEPSAT	Ming Zhang	Florida Institute of Technology	Physics-based	>10, >100	Time Profile	64
SEPSTER	Ian Richardson	University of Maryland	Empirical	>10, >30, >50, >100	Peak Flux	64
SEPSTER2D	Alessandro Bruno	NASA GSFC	Empirical	>10, >30, >50, >100	Peak, Fluence	60
SPREAdFAST	Kamen Kozarev	Bulgarian Academy of Sciences	Physics-based	>10, >30, >50, >100	Time Profile	8
SPRINTS 0-24 hour	Alec Engell	NextGen	Machine Learning	>10, >30, >50, >100	Probability, Peak	15263
STAT	Jon Linker	Predictive Science, LLC	Physics-based	>10, >30, >50, >100	Time Profile	6
UMASEP-10, -100	Marlon Nunez	University of Malaga	Machine Learning, Empirical	>10, >30, >50, >100	Peak, Start	27572, 32240
UNSPELL	Sigiava Alminalragia-Giamini	SPARC	Machine Learning	>5 (validated to >10)	Probability	61
ZEUS+iPATH	Gang Li, Junxiang Hu	UA Hunstville, NASA GSFC	Physics-based	>10, >30, >50, >100	Time Profile	60
SAWS-ASPECS	Athanasios Papaioannou	National Observatory of Athens	Statistical, Empirical, Physics-based	>10, >100	Probability, Peak, Time Profile	57 – 63 for 42 submodels

SEPVAL 2023 First Results

>10 MeV, 10 pfu, Real-Time-like Forecasts

All Clear Binary Forecast

Model	Hits	False Alarms	Correct Negatives	Misses	Total Forecasts	Hit Rate	False Alarm Ratio	Bias	TSS	HSS
Lavasa	18	5	24	6	53	0.75	0.22	0.96	0.58	0.58
MagPy*	3	2	27	29	62	0.12	0.33	0.18	0.05	0.05
PPS SFS-Update	27	23	7	2	59	0.93	0.46	1.7	0.16	0.16
SAWS-ASPECS flare + CME (SOHO)	26	21	9	3	59	0.90	0.7	1.62	0.20	0.20
SAWS-ASPECS flare + CME (SOHO) electrons	24	6	24	4	59	0.90	0.45	1.62	0.20	0.19
SAWS-ASPECS flare	22	11	17	9	59	0.71	0.39	1.06	0.32	0.32
SAWS-ASPECS flare electrons	19	3	25	11	58	0.63	0.14	0.73	0.53	0.52
SEPMOD	19	4	26	13	62	0.73	0.17	0.72	0.46	0.46
SEPSAT	27	17	13	5	62	0.84	0.39	1.38	0.28	0.28
SEPSTER	20	6	25	11	62	0.65	0.23	0.84	0.45	0.45
SEPSTER2D	26	15	15	3	59	0.90	0.37	1.4	0.40	0.39
SPRINTS Post-Eruptive 0-24 hrs	20	8	22	9	59	0.69	0.29	0.97	0.42	0.42
UMASEP-10**	29	1	29	2	61	**0.94**	**0.03**	**0.97**	**0.90**	**0.90**
UNSPELL flare	21	12	18	5	56	0.81	0.36	1.27	0.41	0.40
ZEUS+iPATH	20	4	25	11	60	0.65	0.17	0.77	0.50	0.50



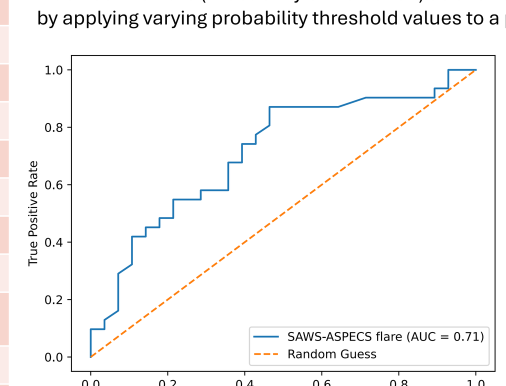
Hit Rate: % SEP events correctly forecast as yes
False Alarm Ratio: Out of all the yes forecasts, what % were false alarms?
Bias: Tendency to forecast false alarms compared to misses
True Skill Statistic: How well can the forecasts separate yes and no events?
Heidke Skill Score: How accurate is the forecast compared to random chance?

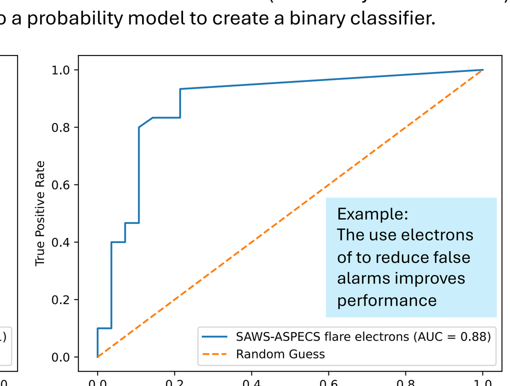
*Unlike all other models listed here, MagPy uses only magnetograms and does not use eruption information as input.
**UMASEP is the only model listed here that uses in situ proton flux as input.

Probability Forecast

Model	Brier Score Best Score = 0	Brier Skill Score Best Score = 1	Area Under the Curve Best Score = 1
COMESEP flare+CME	0.23	0.51	0.72
COMESEP flare only	0.25	0.46	0.77
cRT+AE10	0.17	0.65	**0.87**
MagPy	0.35	0.03	0.55
MEMPSEP Mean	0.20	0.49	0.74
MEMPSEP Median	0.23	0.40	0.72
PPS SFS-Update	**0.16**	**0.67**	0.86
SAWS-ASPECS flare + CME SOHO	0.28	0.40	0.83
SAWS-ASPECS flare + CME SOHO electrons	0.11	0.75	0.92
SAWS-ASPECS flare	0.26	0.48	0.71
SAWS-ASPECS flare electrons	0.23	0.52	0.88
SPRINTS Post-Eruptive 0-24 hrs	0.30	0.34	0.62
UNSPELL flare	0.25	0.42	0.75

Receiver Operating Characteristic:
True Positive Rate (Probability of Detection) as a function of False Positive Rate (Probability of False Alarm) by applying varying probability threshold values to a probability model to create a binary classifier.

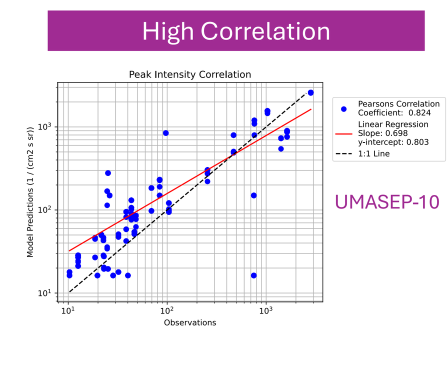




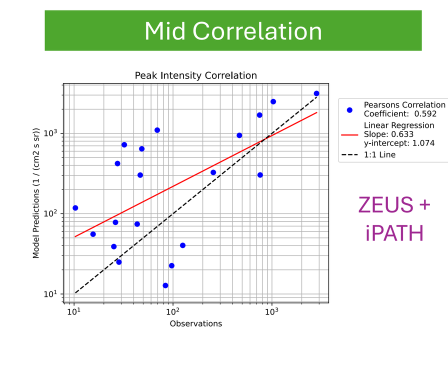
Onset Peak Flux

	Correlation		Bias		Accuracy		
Model	Linear Regression Slope	Pearson Correlation Coefficient (Log)	Mean Log Error (MLE)	Median Log Error (MedLE)	Mean Absolute Log Error (MALE)	Median Absolute Log Error (MedALE)	Median Symmetric Accuracy (MdSA)
COMESEP flare+CME	0.23	0.29	**0.01**	**0.00**	0.57	0.43	1.70
COMESEP flare only	0.25	0.36	-0.15	-0.04	0.51	0.36	1.30
SEPMOD	0.44	0.43	-0.08	0.15	0.62	0.47	1.95
SFS-Update	0.24	0.39	-0.43	-0.36	0.60	0.43	1.67
SEPSTER2D CME	0.78	0.60	0.00	0.09	0.55	0.39	1.47
ADEPT-AFRL 1hr	0.22	0.27	-0.16	0.07	0.48	**0.25**	0.79
ADEPT-AFRL 6hr	0.74	**0.82**	-0.12	0.11	**0.36**	0.28	0.91
SEPSTER (Parker Spiral)	0.80	0.61	-0.48	-0.43	0.68	0.65	3.46
SEPSAT	0.23	0.24	0.02	0.05	0.68	0.63	3.29
ZEUS+iPATH_CME	0.63	0.59	0.35	0.33	0.59	0.48	2.04
SAWS-ASPECS flare + CME (SOHO) 50%	0.33	0.21	0.29	0.77	0.99	1.00	8.84
SAWS-ASPECS flare + CME (SOHO) 90%	0.44	0.21	1.47	2.11	1.91	2.24	172.3
SAWS-ASPECS flare 50%	0.60	0.35	-0.70	-0.81	1.15	1.27	17.51
SAWS-ASPECS flare 90%	**1.03**	0.41	0.38	0.86	1.50	1.62	40.84
UMASEP-10	0.70	**0.82**	0.16	0.18	**0.36**	0.29	**0.93**

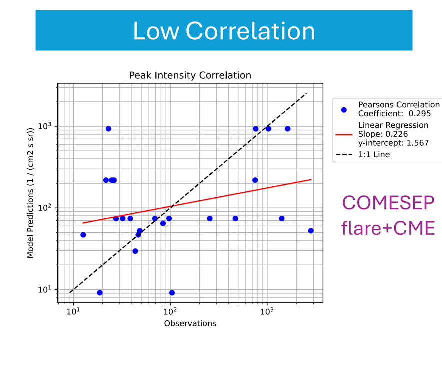
High Correlation



Mid Correlation



Low Correlation



Caveats

SEPVAL results are derived from a wide array of SEP models with a variety of approaches, inputs, and predictions

There are many subtleties involved in making cross-model comparisons which will be addressed in a journal article

Time sampling affects validation outcomes

SEPVAL's challenge event list tells a different kind of story than validation of real time forecasts in the SEP Scoreboards

Future Focus

What do the validation results say about the predictive physical characteristics being captured by the models? What works and what is missing?