A machine-learning oriented remote and in-situ database for forecasting SEP occurrence and properties

K. Moreland[1,2], M.A. Dayeh[2,1], S. Chatterjee[2], A. Munoz-Jamarillo[2], H.M. Bain[3,4], S. Hart[1,2]

[1]University of Texas at San Antonio, [2]Southwest Research Institute, [3]CU Boulder CIRES, [4]NOAA SWPC

Correspondence: kim.moreland@contractor.swri.org

Motivation

We present a new parameter-rich dataset that is tailored for the forecasting of solar energetic particle (SEP) events. The dataset comprises numerous parameters from in situ and remote observatories. It contains over 18,000 flare events and their associated remote images, along with their measured X-ray, radio, proton, electron, upstream interplanetary (IP) plasma, and magnetic field properties. When available, associated SEP, coronal mass ejection, and shock properties are provided, in addition to numerous physics-based derived parameters. In situ data comes from multiple instruments onboard GOES, ACE, and *Wind*. Remote data comes from instruments on board SDO and SOHO and include full-disc magnetograms, EUV, and coronagraph images. Selection criteria for flare event classification and methods for calculating important SEP properties will be explained.

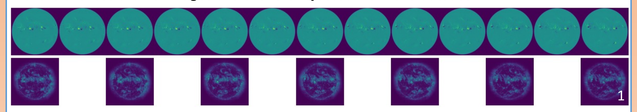
Special consideration is given to data that is currently available in operational real-time or will be available in real-time on upcoming missions. The dataset has already been used in the development of a newly emerging model that forecasts the occurrence and subsequent properties of SEPs at 1 au.

Pre-Flare & Imaging Data

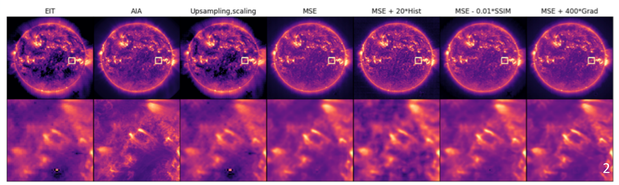
Mission	Instrument	Measurement	Energy Range	Resolution
SOHO	LASCO C2/C3	Coronagraph		1 hour
	MDI	Solar Magnetic Field		6 hour
	EIT	EUV	171Å	12 hour
SDO	HMI	Solar Magnetic Field		6 hour
	AIA	EUV	171Å	12 hour
Wind	WAVES	Solar Radio Emissions	10 kHz – 10 MHz	0.1 second
GOES	EPS/EPEAD	H+, e-	0.6 – 500 MeV	5 minute
ACE	EPAM	H+	0.05 – 5 MeV	12 second
	ULEIS	H+, Fe, O	0.04 – 10 MeV	1 hour
	SIS	Fe, O	3.0 – 170 MeV	1hour
	SWEPAM	Solar Wind Plasma		64 second
	MAG	Local Magnetic Field		16 second

Pre-flare data is the solar wind data 6 hours before the flare onset obtained from the ACE spacecraft at 1 AU. We average valid data from the SWEPAM, ULEIS, and SIS instruments during this time to incorporate 15 values into the dataset. This data provides a snapshot of what is occurring in the solar wind prior to a flare onset.

For each identified event, we stack the previous 3 days of imaging. For EUV we have 7 images (1 image every 12 hours), 13 magnetogram images (1 image every 6 hours) and 73 LASCO C2/C3 images (1 image every hour). Figure 1 shows an example of the magnetogram and EUV images that are saved for a single event. Interpolation is used for missing images; if there are less than 5 images the event is rejected from the dataset.

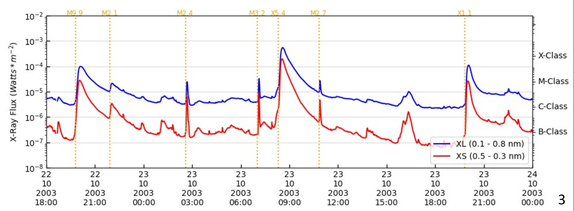


Images from different instruments are homogenized in terms of contrast and resolution using a deep learning (DL) framework. This allows for improvements in terms of contrast and detail. Figure 2 shows the deep learning results used to homogenize the EIT and AIA images by pixel scale, field of view, and point of view.

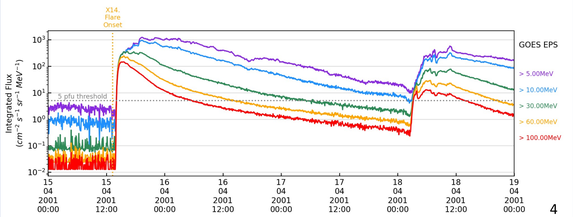


Data Processing

1 SunPy connector to query HEK for GOES Flare Event List (Figure 3).



2 Using GOES X-ray and integrated proton flux search for flux to rise about threshold within 6 hours of flare onset (Figure 4).

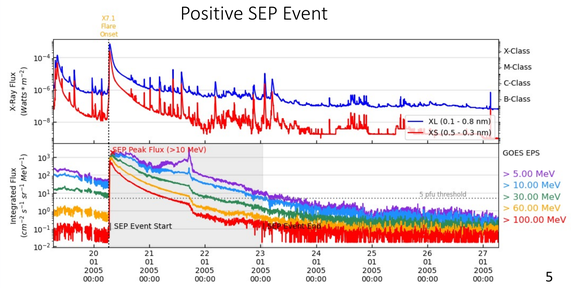


3 Mark events as positive (Figure 5) or negative (Figure 6) for SEPs.

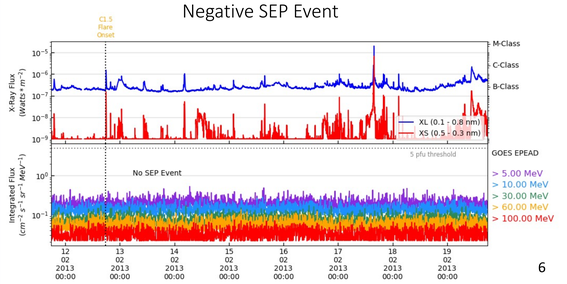
Event Type Determination

4 Calculate SEP related properties (peak flux, fluence, event duration) and save to dataset.

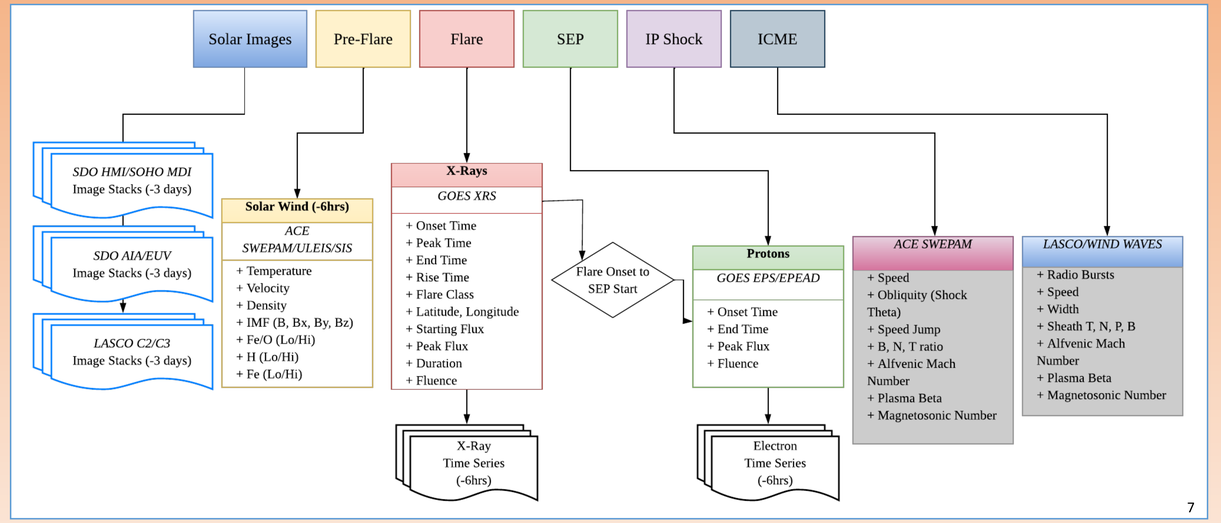
Positive SEP Event



Negative SEP Event



Complete Dataset



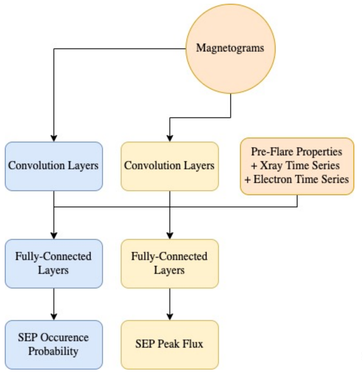
Resulting Database

Flare Class	Positive	Negative	Total Events
X	55	90	135
M	167	1607	1774
C	706	16105	16811

From 1998-2013 we have over 18,000 events, and we associate all of the data shown in the complete dataset flow chart (figure 7) with each event when the data is available. Currently, we are working on adding more data related to IP Shocks and CME (data properties shown in gray boxes) for each event if they are present and the data is available. *Subsequent iterations of the dataset will include data from 2013 onwards.*

Applications

Currently, this dataset is being used as a feature set for a Deep Neural Network (DNN) model to predict whether a SEP event will occur or not and, if one does occur, predict various SEP properties such as peak flux and fluence. First, the solar images undergo convolution before the addition of the pre-flare properties, x-ray, and electron time series are added. These layers are then fully connected to predict the SEP occurrence probability. A diagram of the model is shown in figure 8. You can learn more about the work related to and involving this dataset here the Helio ML conference by visiting Subhamoy Chatterjees' poster "Forecasting the Occurrence Probability and Properties of Solar Energetic Particle Events using a Multivariate Ensemble of Convolutional Neural Networks".



1. Dataset will be available on a dedicated website for community use
2. Additional data relevant products will be added and validated in the coming months

We acknowledge partial support for this work from LWS 80NSSC19K0079 and O2R grants 80NSSC20K0290 and 80HQTR20C0017.