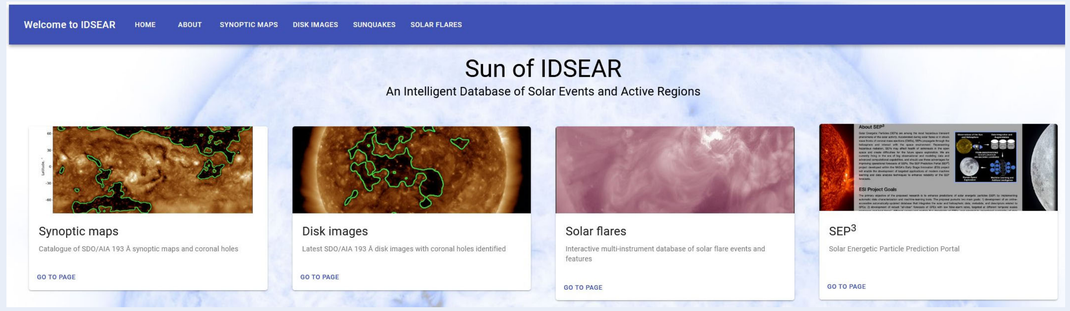
Development of Solar Flare and Energetic Particle Prediction Portal (SEP^3)

Alexander Kosovichev[1], Viacheslav Sadykov[2], Vincent Oria[1], Irina Kitiashvili[3], Patrick O'Keefe[1], Aatiya Ali[2], Chun-Jie Chong[1], Paul Kosovich[1], Samuel Granovsky[1]

[1]New Jersey Institute of Technology, [2]Georgia State University, [3]NASA Ames Research Center

Solar activity is a primary factor determining the state of the Earth's space environment, geomagnetic and ionospheric disturbances, and radiation hazards. In the current state of knowledge, machine learning (ML) methods provide essential tools for processing data, investigating relationships among various physical properties and characteristics, uncovering hidden connections, and predicting hazardous solar events. The primary difficulty in developing and applying modern machine-learning tools in heliophysics is that the essential data are scattered among over a hundred data repositories developed by instrument teams of space missions and ground-based observatories. In addition, statistical and ML methods require long time series of homogeneous measurements. To facilitate ML-ready data preparation and access, we have developed an interactive database of solar flares integrating the most essential datasets (https://solarflare.njit.edu/). The database performs an initial data processing and is automatically updated. In addition, we are developing the Solar Energetic Particle Prediction Portal (SEP3, https://sun.njit.edu/SEP3), which hosts web applications that allow users to retrieve the database records. The Portal has a search page for browsing the events from the most widely used catalogs and a dedicated space to share the most recent achievements of the team. The interactive widget can display soft X-ray and proton flux time series from GOES satellites and the flare records. The data portal has been used to evaluate the forecasts of solar proton events and investigate machine-learning approaches to SEP prediction.



Project Goals and Objectives

Our primary objective is to develop "all-clear" forecasts of Solar Proton Events (SPEs) with low false-alarm rates using a state-of-the-art machine-learning approach.
Specifically, we plan to:

1. develop an online-accessible database that integrates the solar and heliospheric data, metadata, and descriptors related to SPEs;
2. develop robust "all-clear" forecasts of SPEs with low false-alarm rates, targeted at different temporal scales (cadences and lead times), different energy and particle flux thresholds of SPEs, and adapted to the operational availability of data sources and gaps in the data.

Solar Energetic Particle Prediction Portal (SEP^3)

To support the project data needs, we are currently developing an online-accessible database of SPE-related data, metadata, and data products (SEP^3 project).

- The database is available online for the broader research community from the NJIT web server. Core features of the portal include various SPE-related sources collected in one place:
  - properties of active regions (PIL, SMARP, and SHARP Space Weather Active Region Patches, Solar Region Summary records);
  - GOES proton and SXR fluxes;
  - NOAA records of the radio bursts, flares, and SPEs;
  - SOHO/EPHIN energetic particle data;
  - OULU neutron monitor data;
  - CACTUS and SOHO/LASCO CME catalog records
- MySQL database schema developed and optimized to efficiently handle the data queries necessary for the project
- Intuitive Web application with API-based online access to database entries and data products (under development).

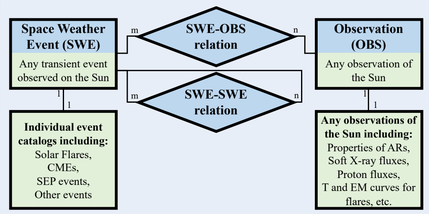
https://sun.njit.edu/

Acknowledgements

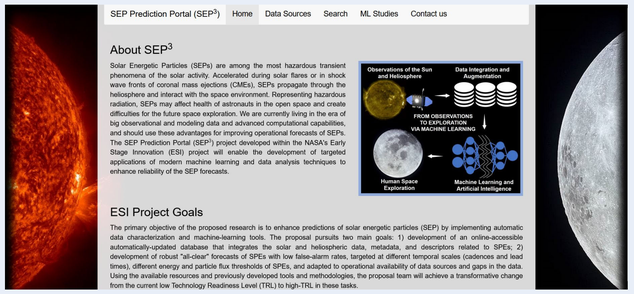
We thank SDO/HMI and GOES satellite teams for providing the high-quality scientific data and related data products. We thank Space Weather Prediction Center at National Oceanic and Atmospheric Administration (SWPC NOAA) for providing operational reports and predictions of space weather events. The Bootstrap API is distributed under MIT license. This work is supported by NASA ESI grant 80NSSC20K0302 and NSF grants 1916509 and 1639683.

Data sources and the SEP^3 database structure

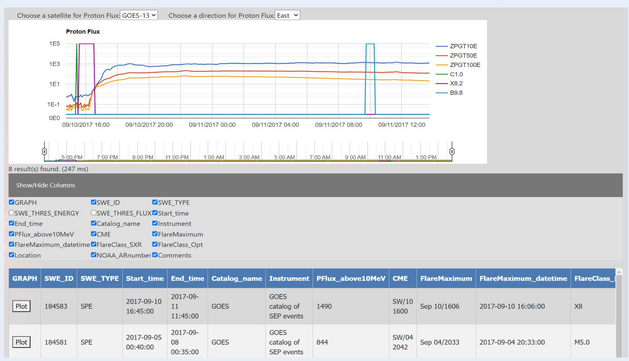
Solar Energetic Particle DB Data Sources	
Polarity Inversion Line (PIL) properties of Active Regions	GOES proton flux measurements and event catalogs
Space Weather HMI Active Region Patches (SHARPs)	GOES soft X-ray measurements
Flare lists from various flare catalogs (GOES, RHESSI, HEK)	GOES Temperatures and Emission measures
CME lists (LASCO/SOHO and CACTus catalogs)	Records of radio bursts (type II and IV)
Measurements of ground-based neutron monitors	Derived lists of SPE events and SPE-quiet periods
Historic classifications of sunspots & ARs (McIntosh and magnetic classes)	Other resources



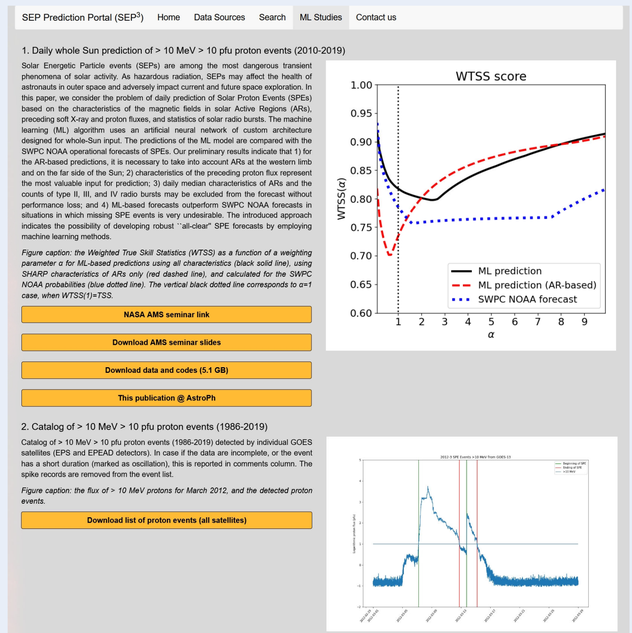
Solar Energetic Particle Prediction Portal (SEP^3) title page



Example of the interactive SEP search output



Datasets and Statistical and Machine-Learning Analysis



SMARP and SHARP: Active Region Data From SOHO and SDO

SMARP:
- 1996.04.23 - 2010.10.27
- 96-minute cadence
- 0.12 degree/pixel resolution
- only include line-of-sight maps

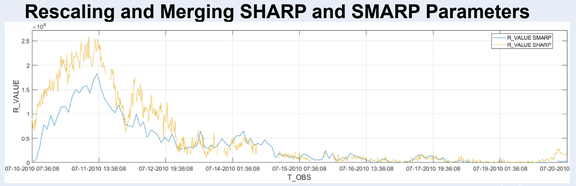
SHARP:
- 2010.05.01 - present
- 12-minute cadence
- 0.03 degree/pixel resolution
- include maps of the entire vector

- Two data products derived from photospheric heliomagnetic field data taken by SOHO MDI (SMARP) and SDO HMI (SHARP).
- Provide coverage from April, 1996 to present (two full solar cycles).
- Both contain AR images and their summary physical parameters:

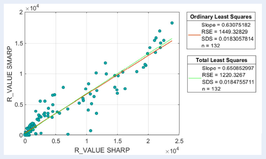
Parameter	Description	Units
USFLUXL	Total line-of-sight unsigned magnetic flux	Maxwells
MEANGBL	Mean value of the line-of-sight-field gradient	Gauss/Mm
R_VALUE	Unsigned flux R near polarity inversion lines	Maxwells
CMASKL	CEA pixels in active region	N/A
MEANGBZ	Mean value of the vertical field gradient	Gauss/Mm
USFLUXZ	Vertical component of the total unsigned flux	Maxwells

CHALLENGES: Data gaps; time series can be difficult to synchronize; unit conversions; SMARP vertical component computations.

Rescaling and Merging SHARP and SMARP Parameters



- Overlap in observations from May 1, 2010 to October 28, 2010.
- 10 most active regions were selected inside the overlap period.
- For each parameter, the slope and intercept were determined, and the weighted averages were computed.
- The example plotted here is R_VALUE for NOAA AR 11087.



A comparison between two linear regression techniques:

1. Ordinary Least Squares - much simpler to apply, and to compute the error.
2. Total Least Squares – more appropriate to use when there is uncertainty in both X and Y.

Total Least Squares was the method used in this study.

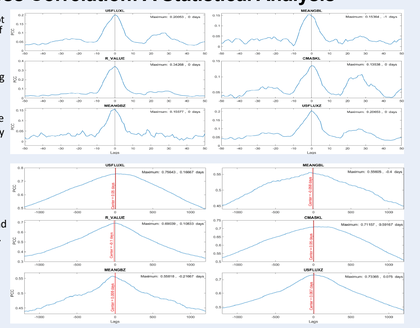
Time-Lagged Cross Correlation: A Statistical Analysis

Challenge: The combined SHARP-SMARP is not one continuous time series, but a collection of time series (one for each active region).

Solution#1:
- Create continuous time series by picking records corresponding to maximum R_VALUE, regardless of coordinates.
- We can then perform TLCC between the daily-averaged parameters and the daily flare index.

Solution#2:
- Create continuous time series by integrating the parameters over the entire solar disk.
- In this example, we use only SHARP, and perform TLCC with the GOES X-ray flux.

Conclusion: Some parameters (i.e. R_VALUE, MEANGBL, MEANGBZ) may serve as better precursors than others.



Rolling-Window Time-Lagged Cross-Correlation

- Applying the rolling window technique to TLCC allows us to see how the leader-follower dynamic varies with time.
- Here, the results of RWTLCC between the six SHARP-SMARP parameters and the daily flare index have been averaged and plotted on the right; the daily flare index is on the left.
- The color bar represents the PCC correlation score.
- Inconsistent pattern: in some cases, the SHARP-SMARP parameters lead the flare index, but in other cases the opposite is true.
- Conclusion: Overall, areas of highest correlation appear to correspond to higher flare activity, regardless of the direction of offset. Lack of flare data during solar minima still presents a challenge.