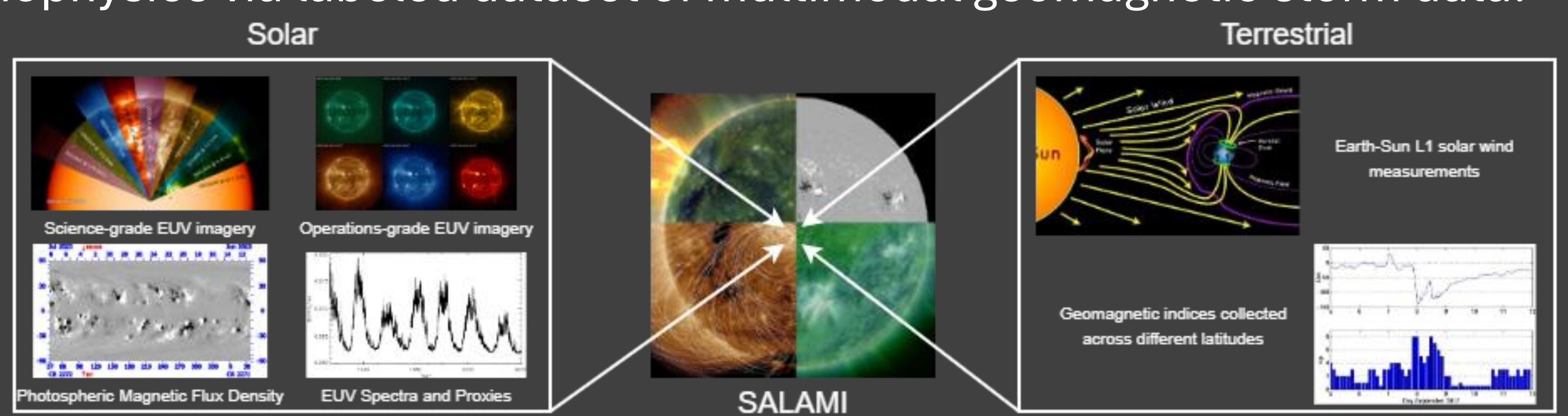
Storm Archive for Learning and Anticipation with Machine Intelligence (SALAMI)

Daniel Brandt[1], Meryl Spencer[1], Nathan Lemus[1], Tristan Singleton[1], Vukosav Subotic[1]

[1]Michigan Tech Research Institute, Michigan Technological University

The Need: An AI/ML-ready Dataset for Geomagnetic Storms

Geomagnetic storms are the mainstays of space weather prediction. In an age with a growing body of space weather-related observations, both scientific and commercial, and at a time when AI-powered data assimilation techniques are rapidly advancing, we are now better positioned to derive predictive insights from combined datasets than ever before. SALAMI provides the foundation for obtaining that insight in heliophysics via labeled dataset of multimodal geomagnetic storm data.



SALAMI expands upon SDOML, which focused exclusively on SDO data [1].

The Data: Times and Places across Interplanetary Spaces

The data in SALAMI are all time series data of a sort: either they are univariate time series data or time series of images.

Imagery: AIA 1600 Å, 304 Å, 211 Å, 193 Å, 171 Å, 94 Å; SUVI 304 Å, 171 Å, 131 Å, 94 Å; Synoptic Magnetograms

Univariate Observables: We use hourly NASA OMNI solar/geomagnetic data. To remove data gaps in the time series, we employ *multivariate Singular Spectrum Analysis* [2]. In the case of gaps in the solar wind alpha-proton ratio, we use a parametric reconstruction.

Gap-Filling:

$$\tilde{\mathbf{X}} = \|\tilde{x}_{i,j}\| = \sum_{k=1}^{l} P_{i_k} P_{i_k}^{\mathrm{T}} \mathbf{X}$$

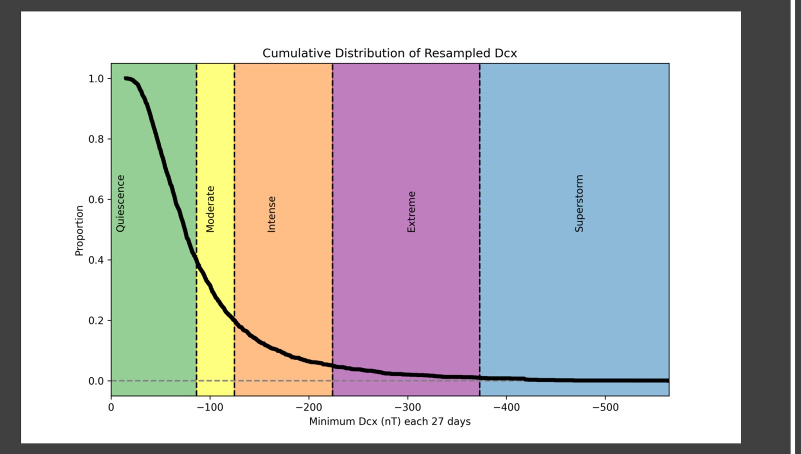
Singular Spectrum Analysis

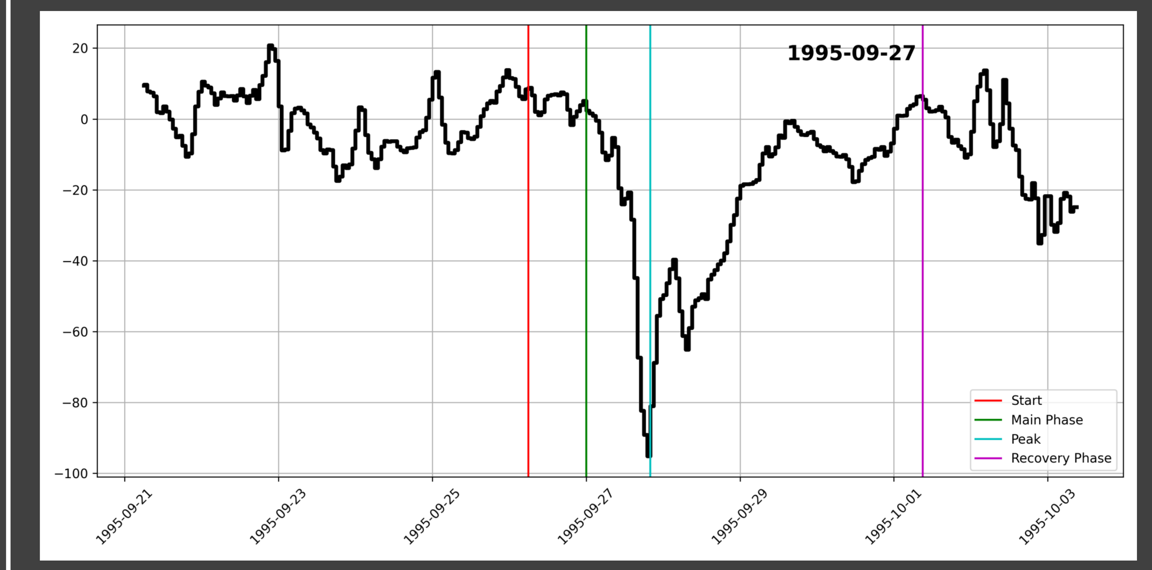
$$\frac{n_a}{n_p} = \frac{1}{4}\left(\frac{p \times 10^6}{1.67 n_p v_p^2} - 1\right)$$

Parametric Reconstruction

The Method: Storm Identification and Labeling

Storms are labeled by intensity based on where their peak Dcx index [3] values fall in an empirical distribution function [4].





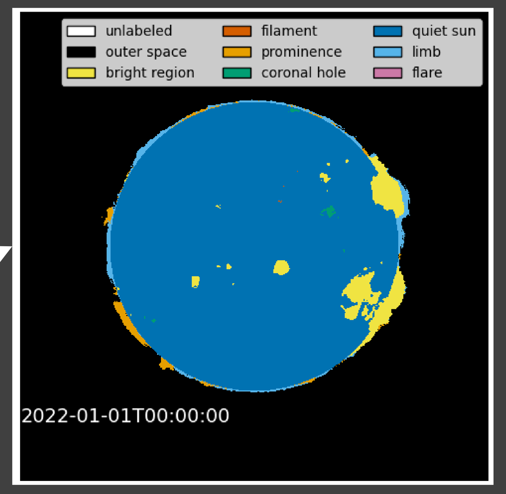
Storms phases are automatically determined based on the heuristics devised by [5].

Storms are labeled by intensity and class. Storms with multiple peaks are treated separately. Time periods of geomagnetic quiet equivalent in total duration to the storm times are also included. Solar imagery and OMNI time series data are obtained in each of these intervals.

The Future: Applications and Next Steps

Applications:

- Development of novel space weather forecasting algorithms
- Precursor identification for geomagnetic disturbances
  - *SALAMI provides scripts for obtaining fine resolution Level 1 AIA imagery and FISM2 flare products during time periods of interest. SALAMI also provides cross-referencing of storm periods with solar flares from the CCMC DONKI Database* [6].
- Space weather classification algorithms
- Model calibration



Next Steps:

- Integration of SUVI thematic maps (high priority)
- Implementation of MSSA with optimized window sizes and reconstruction modes
- Inclusion of quantified uncertainties on reconstructed data
- Functionality for seamless integration of new/updated data (i.e., SWFO L1, ESA VIGIL)
- Integration of tools for data augmentation

Acknowledgements

This work was supported by NASA Grant #80NSSC20K1581. We are additionally grateful to Dr. Brian Thelen (MTRI), and Dr. Barbara Thompson (NASA GSFC) and the rest of the SDO Science Team for providing valuable feedback that improved this program. We thank the AIA, SUVI, GONG, and OMNI project scientists whose tireless work made this project possible.

References

[1] Galvez, et al. 2019; doi:10.3847/1538-4365
[2] Kondrashov, et al. 2014; doi:10.1002/2014GL059741
[3] Mursula, et al. 2008; doi:10.5194/astra-4-41-2008
[4] Collado-Villaverde, et al. 2024; doi:10.1007/s11069-023-06241-1
[5] Katus, et al. 2012; doi:10.1029/2012JA017915
[6] Inceoglu, et al. 2018; doi:10.3847/1538-4357/aac81e

Google Scholar

Contact: daabrand@mtu.edu

This work is in preparation for submission to the journal Space Weather

2026 Space Weather Workshop: April 27 – May 1