https://tinyurl.com/ictp-cheung

Data-Driven Models for Space Weather Prediction

Mark Cheung[1,2]
1. Lockheed Martin Solar & Astrophysics Laboratory, Palo Alto, California, USA
2. Stanford University, California, USA

ISWI Workshop
20-24th May 2019, ICTP, Trieste, Italy

Data-Driven Models

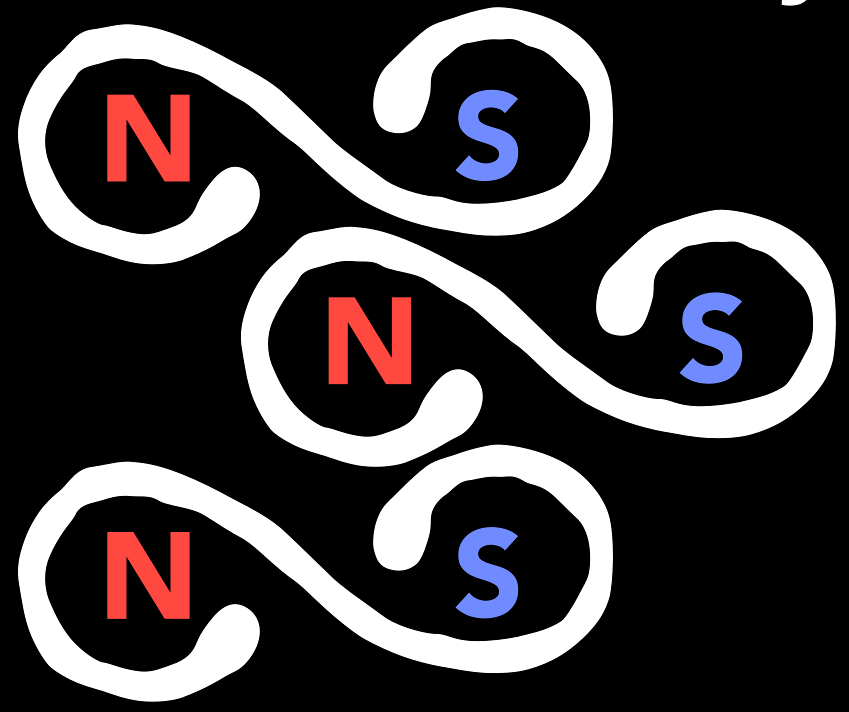
Physics-based Models:

- Data-inspired Models: Simplified simulations to mimic observed scenarios
- Data-constrained Models: Time-independent models satisfying observations at an instant in time. Includes models that may start with a data-constrained initial condition but driven by idealized boundary conditions.
- Data-Driven Models: Time-dependent models evolved in response to evolving boundary conditions

Empirical Data-Driven Models:

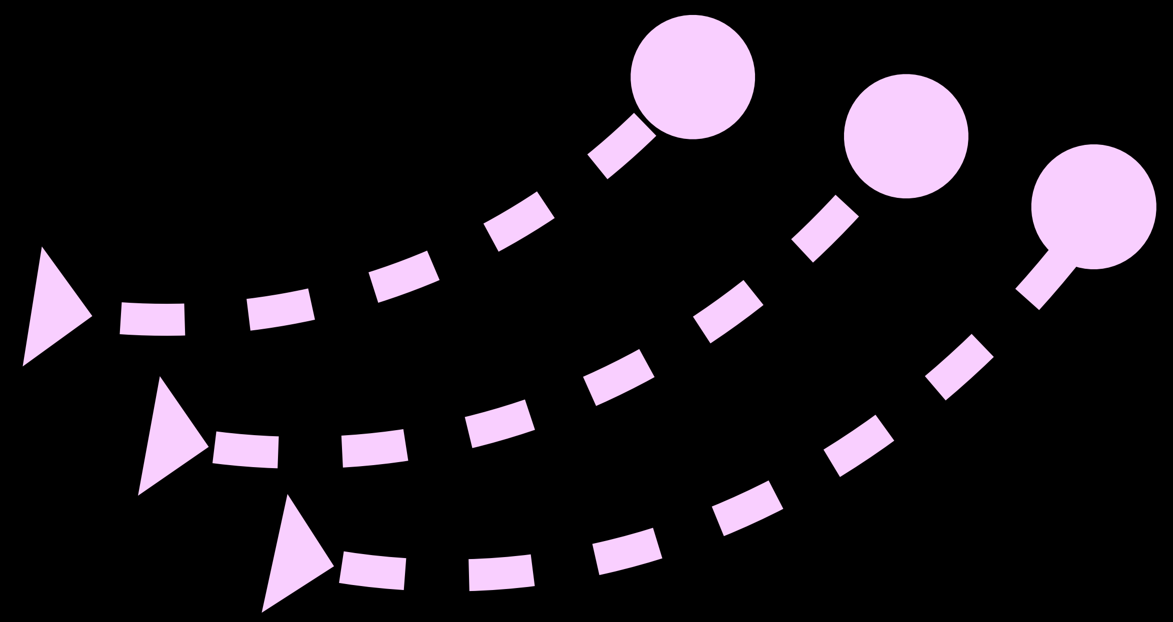
- Physics-rules not prescribed. Try to discover relations in the data.

Magnetized Wind and Ejections

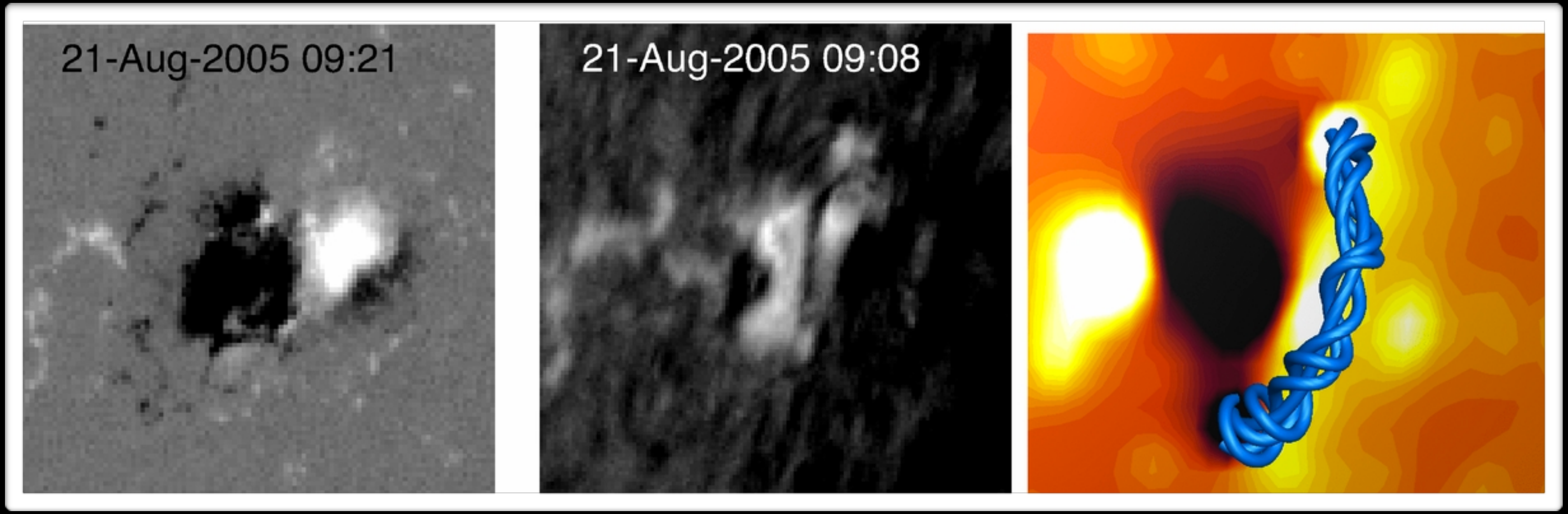


Light (all forms)

Particle storms

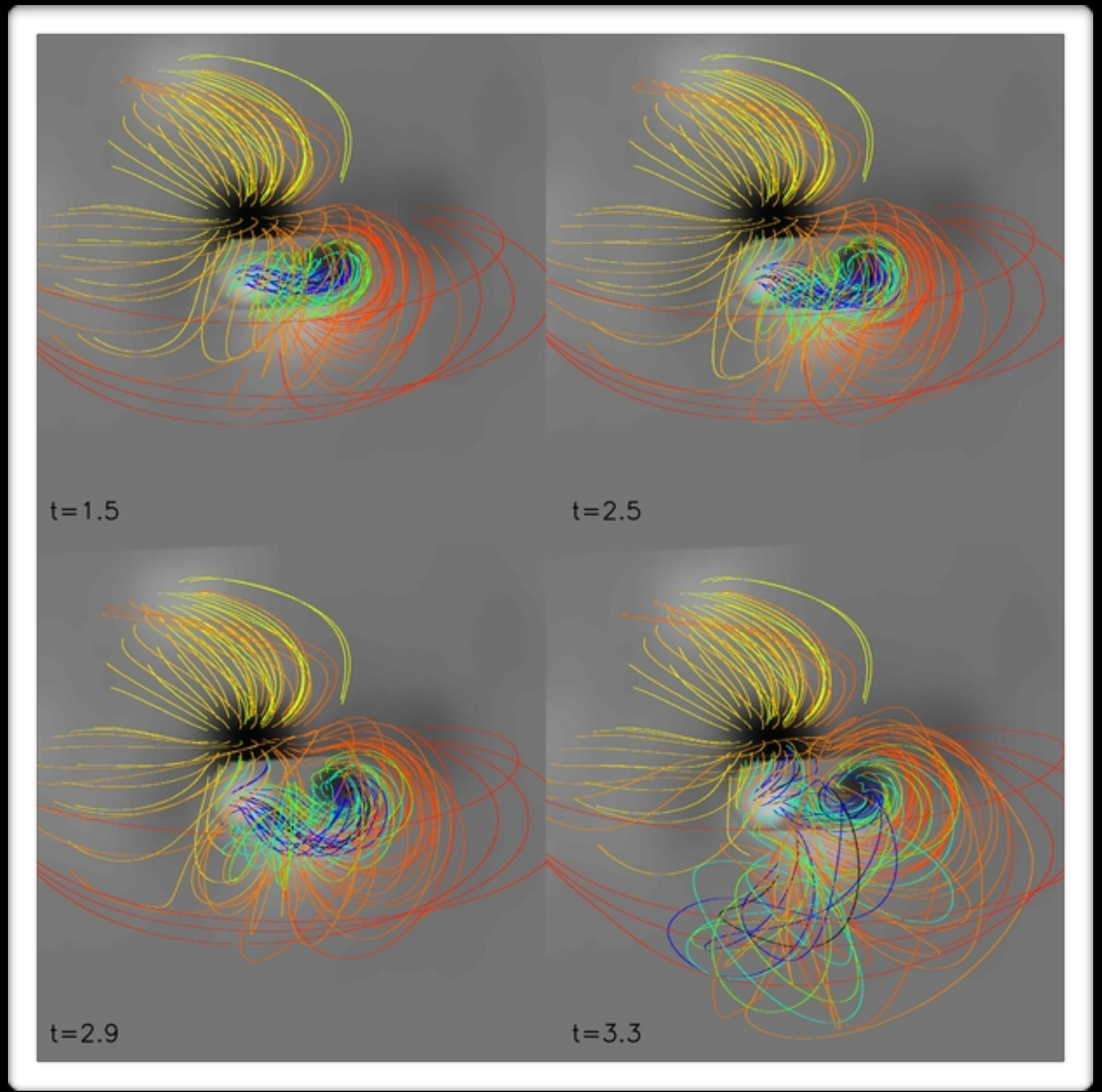


Examples of Data-inspired Models



Left: Lugaz et al. (2011, ApJ)

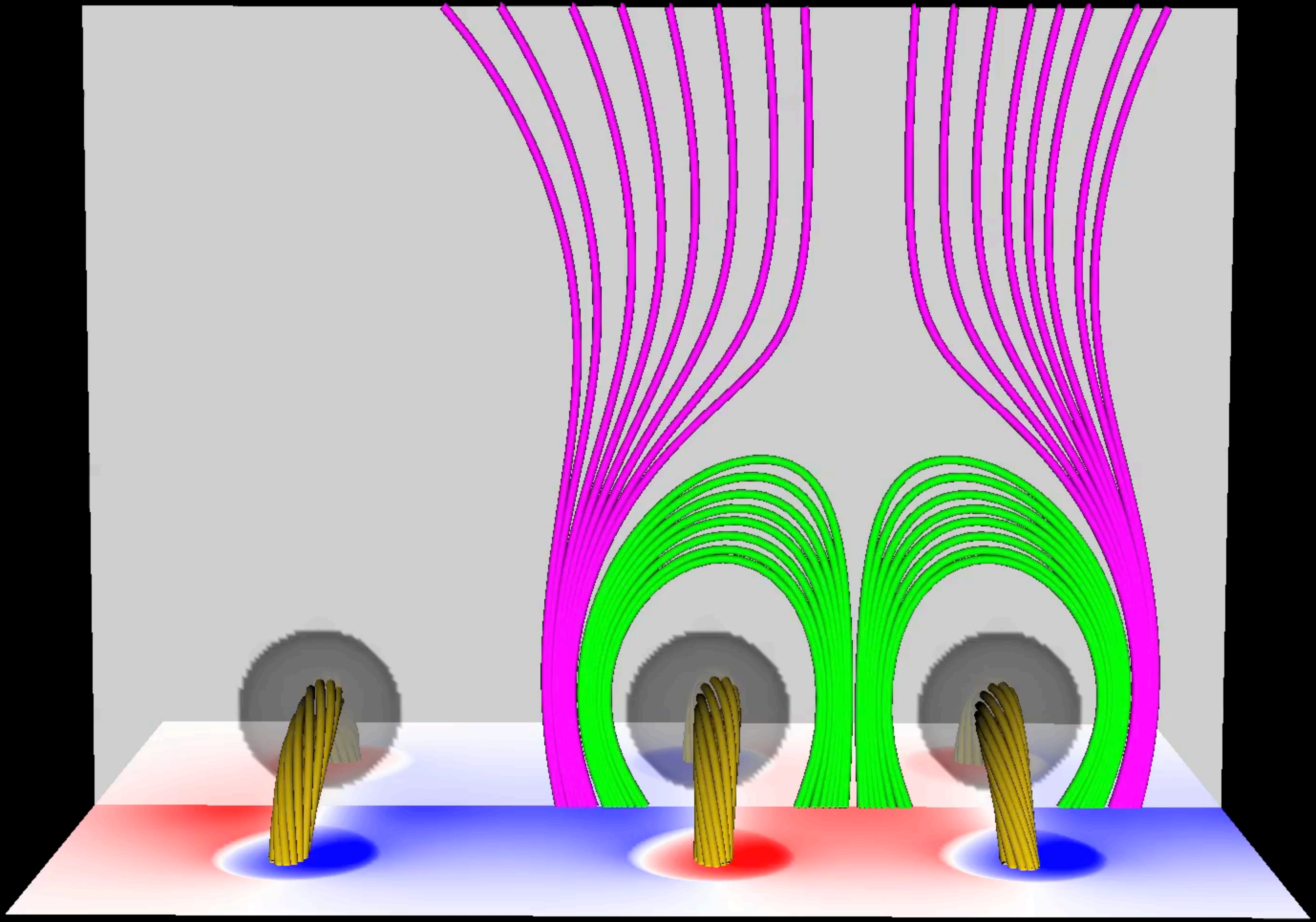
- Idealized flux rope inserted into background field extrapolated from a synoptic magnetogram.
- MHD evolution of the non-force-free initial condition leads to a CME



Right: Fan (2011, ApJ)

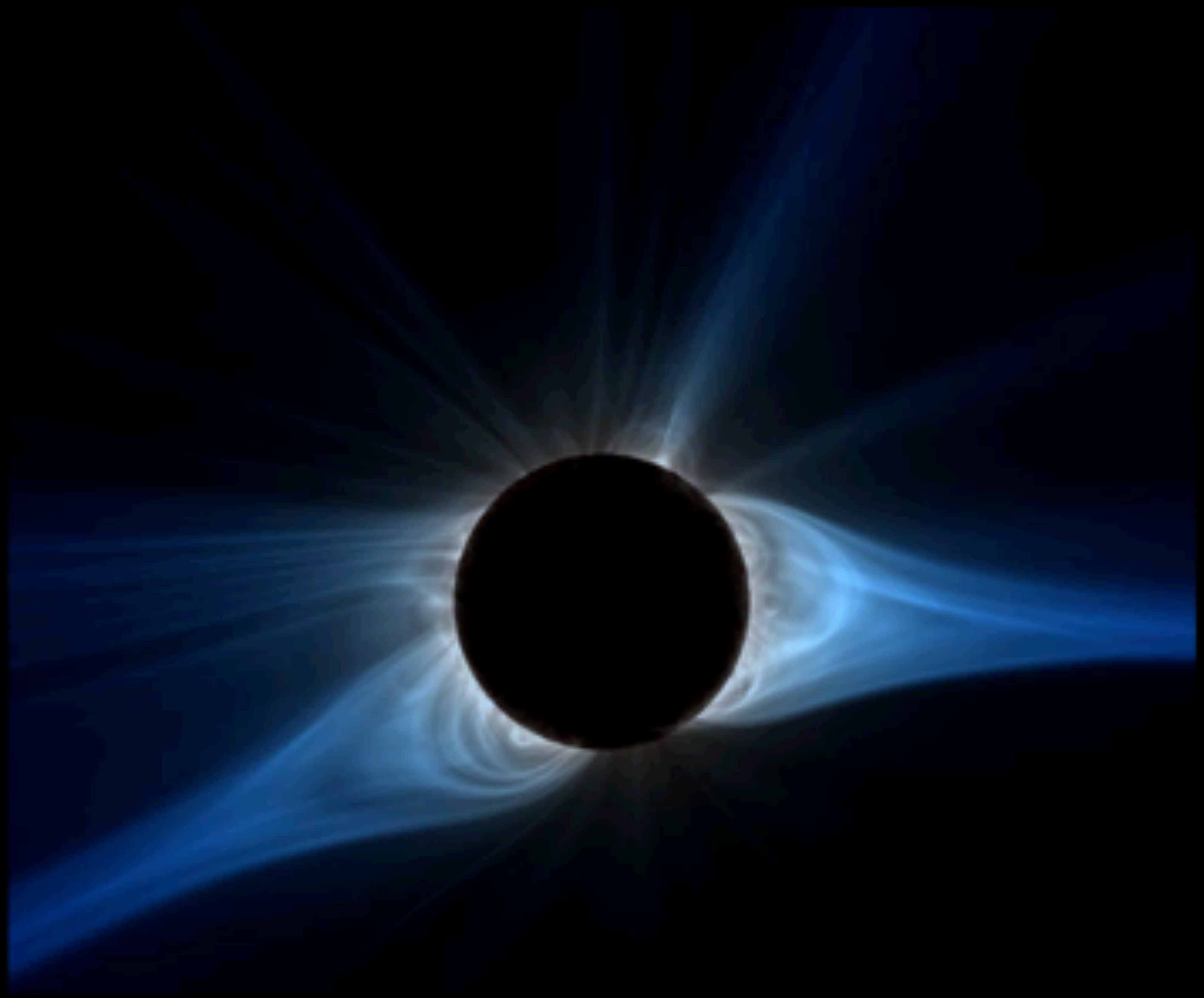
- Smoothed MDI magnetogram of AR 10930 so that B=3 kG -> 200 G
- A twisted flux rope was emerged into the pre-existing sunspot. The interaction between the two magnetic systems leads to an eruption

Examples of Data-inspired Models



Torok et al. (2011, ApJL): MHD model of sympathetic eruptions inspired by Aug 1st 2010 events.

Data-Constrained Model: Aug 21st Eclipse Predictions



Prediction made August 14, 2017
Based on SDO/HMI and SDO/AIA data
Using software developed by Predictive Science, Inc.
http://www.predsci.com/corona/aug2017eclipse/home.php

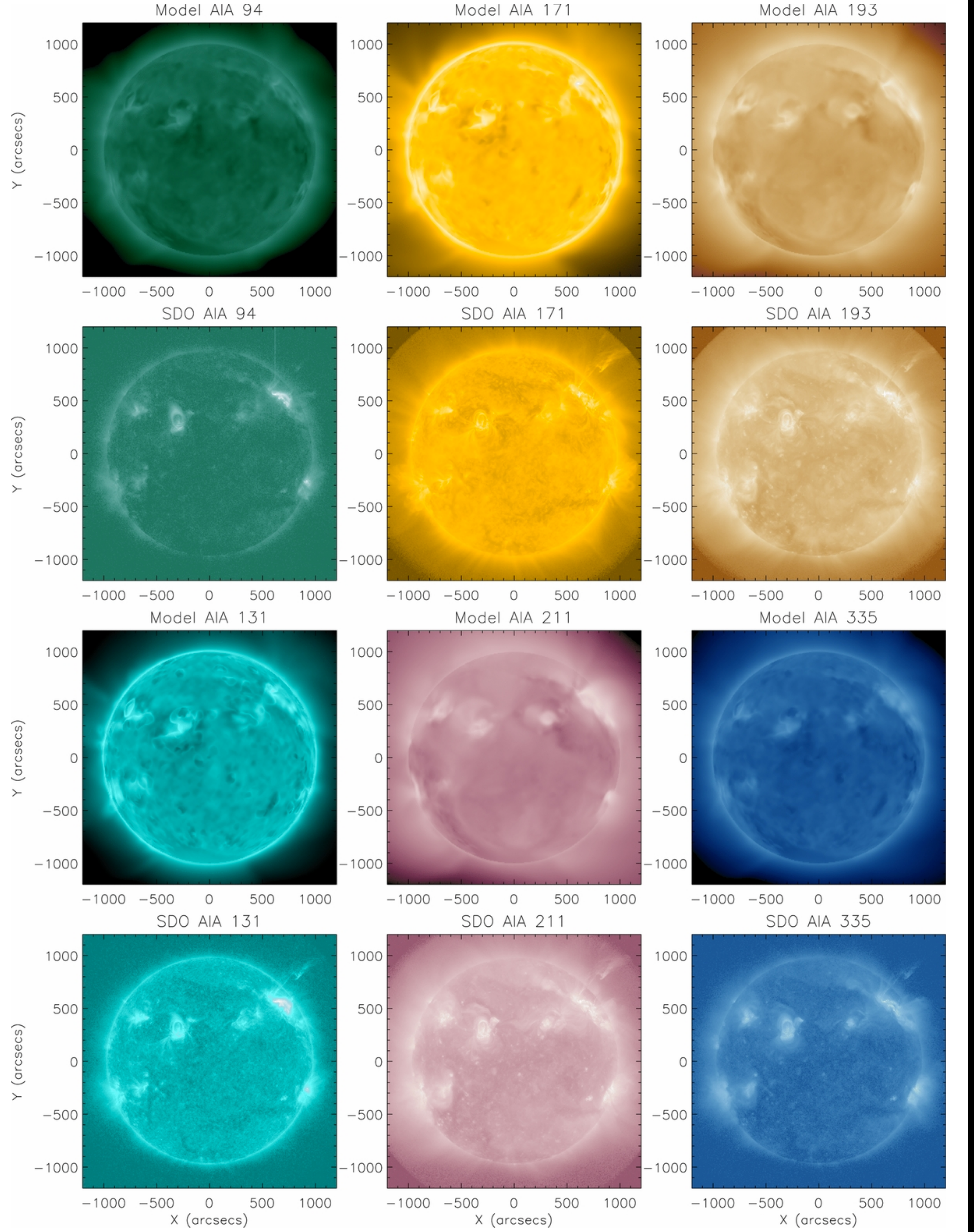
rsackett00@yahoo.com
Cape Girardeau, MO
August 21, 2017 1:21 pm CDT
High dynamic range composite processed to bring out coronal streamers and earthshine on moon. Sky & Telescope online gallery

Courtesy: Z. Mikic

Data-Constrained Models

Alfvén Wave Solar Model (AWSoM) van der Holst+ (2014, ApJ)

- Fully-compressible MHD equations + Alfvén wave propagation and dissipation.
- Used AIA (and STEREO) EUV images to validate the Alfvén wave heating model (as opposed to an analytical spatially-dependent heating model).
- See Alvarado-Gómez et al. (2016, 2018) for application to stellar winds of exoplanet host stars.




Data-inspired Model

SDO's main goal is to understand, driving toward a predictive capability, those solar variations that influence life on Earth.

SDO images the sun's surface, atmosphere and interior.
The mission generates about 3 terabytes worth of science data.

SDO in a Nutshell

- 3 instruments monitoring the Sun all the time since May 2010.
  - Atmospheric Imaging Assembly (AIA): visible, UV, and EUV full disk images of the photosphere, chromosphere, transition region and corona at 4096x4096 pixels.
  - Helioseismic & Magnetic Imager (HMI): visible light full disk dopplergrams and magnetograms at 4096x4096 pixels.
  - EUV Variability Experiment (EVE): disk-integrated EUV irradiance spectra at 1 Å resolution.
- About 12 PBs of data to date.
- SDO science data has been part of over 3000 refereed publications (18 in Science, 17 in Nature, 46 PhD dissertations).
- Easy data access: First authors are spread out over 33 countries with co-authors from at least another 18 (source: NASA SDO project scientist Dean Pesnell).


HMI/Continuum
HMI/Doppler
HMI/Magnetogram
AIA/171A
EVE
2011-09-25T07:59:49
SDO/EVE Data
WHI 2008
Visible
Flux (W/m^2/nm)
EVE/SAM
9.4
13.1
17.1
19.3
21.1
30.4
33.5
160.0
170.0
450.0
617.3
Wavelength (nanometers)
AIA/193A
AIA/304A
AIA/335A
AIA/1600A
2011 Sep 25 08:00:36

DATA

IMPACT

social

scientific

OPEN DATA POLICY

1. No proprietary data withholding period.
2. Anyone with internet access can download full resolution, quick-look images within minutes of their capture. Fully calibrated science data available within days.
3. Mirror data archives located around the world, including at Harvard-Smithsonian Center for Astrophysics, MPI for Solar System Research (Göttingen), University of Lancashire (UK) and Korea.

http://www.lmsal.com/hek

COLLABORATIVE METADATA ENVIRONMENT

1. Researchers / computer algorithms find features and events (e.g. sunspots, flares) and submit them to the Heliophysics Events Knowledgebase (HEK).
2. HEK is like a table of contents for solar data.
3. HEK tells the user which data sets (from different observatories) are available, which events are nearby. This accelerates their workflow and widens their discovery space.

OPEN DATA

APPLICATION
PROGRAMMING
INTERFACES

COLLABORATIVE
METADATA
ENVIRONMENT

Joint Science Operations Center (JSOC)

http://jsoc.stanford.edu

Search | Match ALL words

Data Access | Visual Catalog | Docs | News & Events

- HMI Data Products
- AIA Data Products
- MDI Data Products
- SHA Data Products
- IRIS Data Products
- SID Data Products

** Useful Links **

- SDO Data Use Policy
- HMI Coverage Tables
- HMI Support Information
- AIA Coverage Tables & Release Notes
- JSOC Processing Status
- JSOC System Status
- HMI Event Tables

Welcome to the Joint Science Operations Center (JSOC) Science Data Processing (SDP) home. Data products from the Solar Dynamics Observatory, as well as certain other missions and instruments, are available from the JSOC database. The following instruments and projects have data archived here:

Helioseismic and Magnetic Imager (HMI): is one of three instruments aboard the Solar Dynamics Observatory(SDO) designed to study oscillations and the magnetic field at the solar surface. HMI observes the full solar disk at 6173 Å with a resolution of 1 arc second and is a successor to the Michelson Doppler Imager(MDI) on the Solar and Heliospheric Observatory(SOHO).

Atmospheric Imaging Assembly (AIA): is another instrument board the Solar Dynamics Observatory(SDO) designed to study the solar corona, taking simultaneous full disc images in multiple wavelengths of the corona and transitional region (up to half a solar radius above the solar limb), with 1.5 arc sec resolution and 12 second temporal cadence or better. The primary goal of the AIA Science Investigation is to significantly improve our understanding of the physics behind the activity displayed by the Sun's atmosphere, which drives space weather in the heliosphere and in planetary environments.

Michelson Doppler Imager (MDI): is the predecessor to the current HMI and was launched aboard the Solar and Heliospheric Observatory (SOHO). It is a project of the Stanford-Lockheed Institute for Space Research and part of an international collaboration to study the interior structure and dynamics of the Sun. All the data observed by MDI is now archived in the JSOC.

Stanford Helioseismology Archive (SHA): is a compilation of helioseismology data from various missions including Global Oscillations Network Group (GONG), Mount Wilson, Magneto-Optic Two-Height Instrument (MOTH), Taiwan Oscillations Network (TON) and others to facilitate research.

Interface Region Imaging Spectrograph (IRIS): is a multi-channel imaging spectrograph with a 20 cm UV telescope which will obtain UV spectra and images with high resolution in space (0.33-0.4 arc sec) and time (1s) focused on the chromosphere and transition region of the Sun. The primary goal of the IRIS explorer is to understand how the solar atmosphere is energized.

Sudden Ionosphere Disturbance(SID) Monitors program is an educational project to build and distribute inexpensive ionospheric monitors to students around the world.These monitors detect solar flares and other ionospheric disturbances. JSOC is the central data repository where students can exchange and compare data.

Contacts | JSOC Home | Exportdata | Lookdata | SDO-NASA | Stanford Solar Home | Stanford Solar-Center

SDO Privacy Notice

drms documentation

Release:	0.5.6
Date:	February 19, 2019
Github:	https://github.com/sunpy/drms
PyPI:	https://pypi.python.org/pypi/drms

Python module for accessing HMI, AIA and MDI data.

- Introduction
  - Requirements
  - Installation
  - Acknowledgements
- Tutorial
  - Basic usage
  - Data export requests
  - Example scripts
- API Reference
  - Classes
  - Constants and utility functions
  - Exceptions

LMSAL >> Sungate >> Heliophysics Events Knowledgebase http://www.lmsal.com/hek/api.html

HEK home | Recently reported events | Search Events | Search Data | Request AIA Data | API | Contact Us

A Heliophysics Events Knowledgebase to facilitate scientific discovery

List of Supported Feature/Event types and associated attributes

The full list of Event/Feature types and associated attributes can be found here.

Web API

Web developers who wish to create third-party web applications interacting with the Heliophysics Events Registry should consult the HER Web API wiki, which provides examples on how to query HER, how to submit events to HER as well as other functions.

Sunpy API

Sunpy has a HEK module for using HEK's web API.

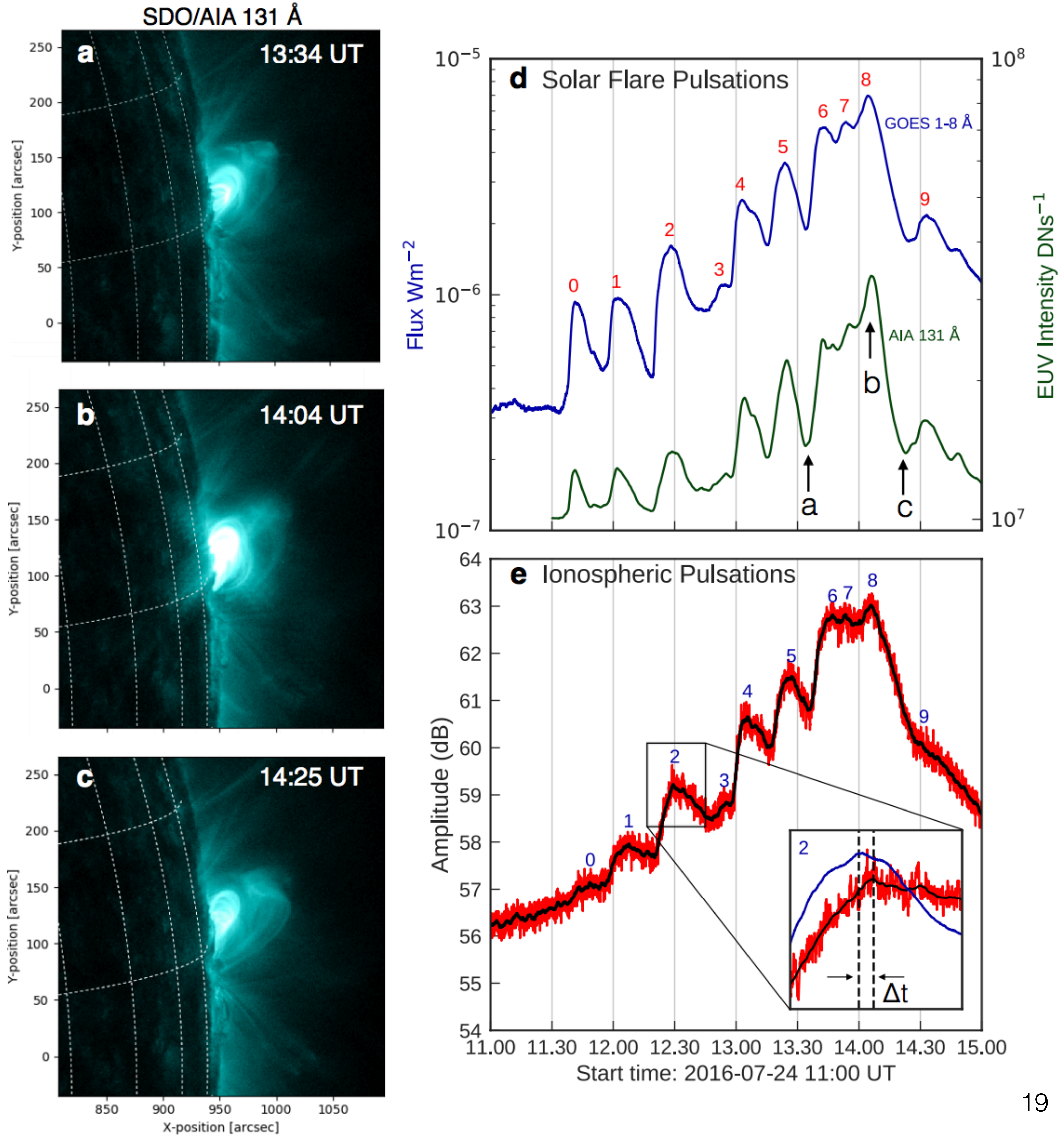
SolarSoft IDL

We are developing a number of software packages to help researchers use and contribute to the HEK project:

- Ontology package: SolarSoft API for reporting events and features to the Heliophysics Events Registry (HER), as well as for querying HER.
- Panorama: an OpenGL based browser for viewing solar data

- Hayes, Gallagher , McCauley, Dennis, Ireland & Inglis, *“Pulsations in the Earth’s Lower Ionosphere Synchronized with Solar Flare Emission”*, JGR, 2017.

- *“To examine the lower ionosphere response to X-ray QPP, VLF radio signals at 24 kHz emitted by the communications transmitter in Maine, U.S. (station ID: NAA; 44.6∘N, 67.2∘W) were monitored at the Rosse Solar–Terrestrial Observatory in Birr, Ireland (53.1∘N, 7.9∘W) using Stanford University Sudden Ionospheric Disturbance (SID) monitor (Scherrer et al., 2008).”*



http://newserver.stil.bas.bg/ISWI/Projects/Instrument_Area.html

Data-Driven Space Weather Models

Physics-based Models:

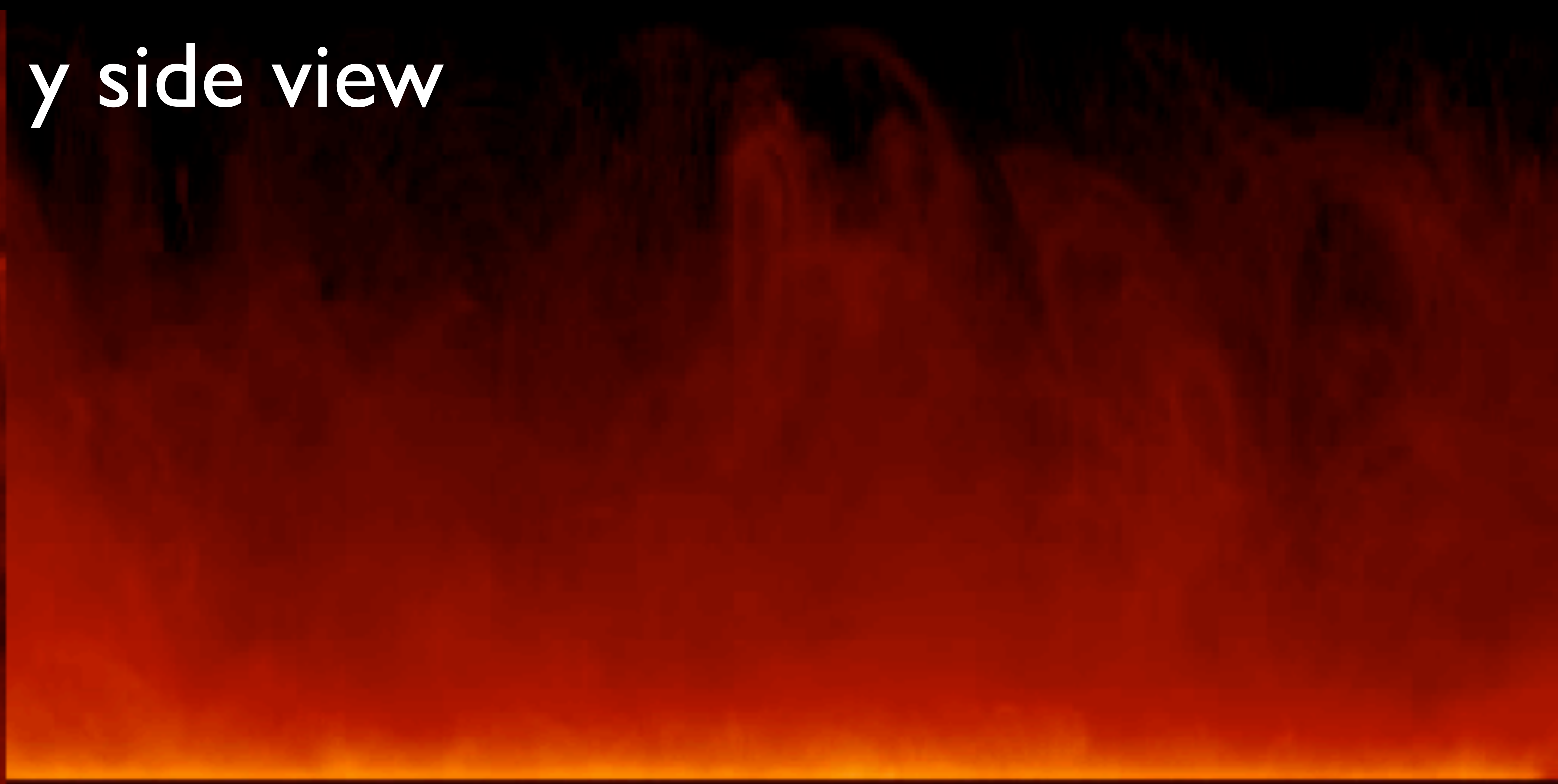
- Data-inspired Models: Simplified simulations to mimic observed scenarios
- Data-constrained Models: Time-independent models satisfying observations at an instant in time. Includes models that may start with a data-constrained initial condition but driven by idealized boundary conditions.
- <u>Data-Driven Models: Time-dependent models evolved in response to evolving boundary conditions</u>

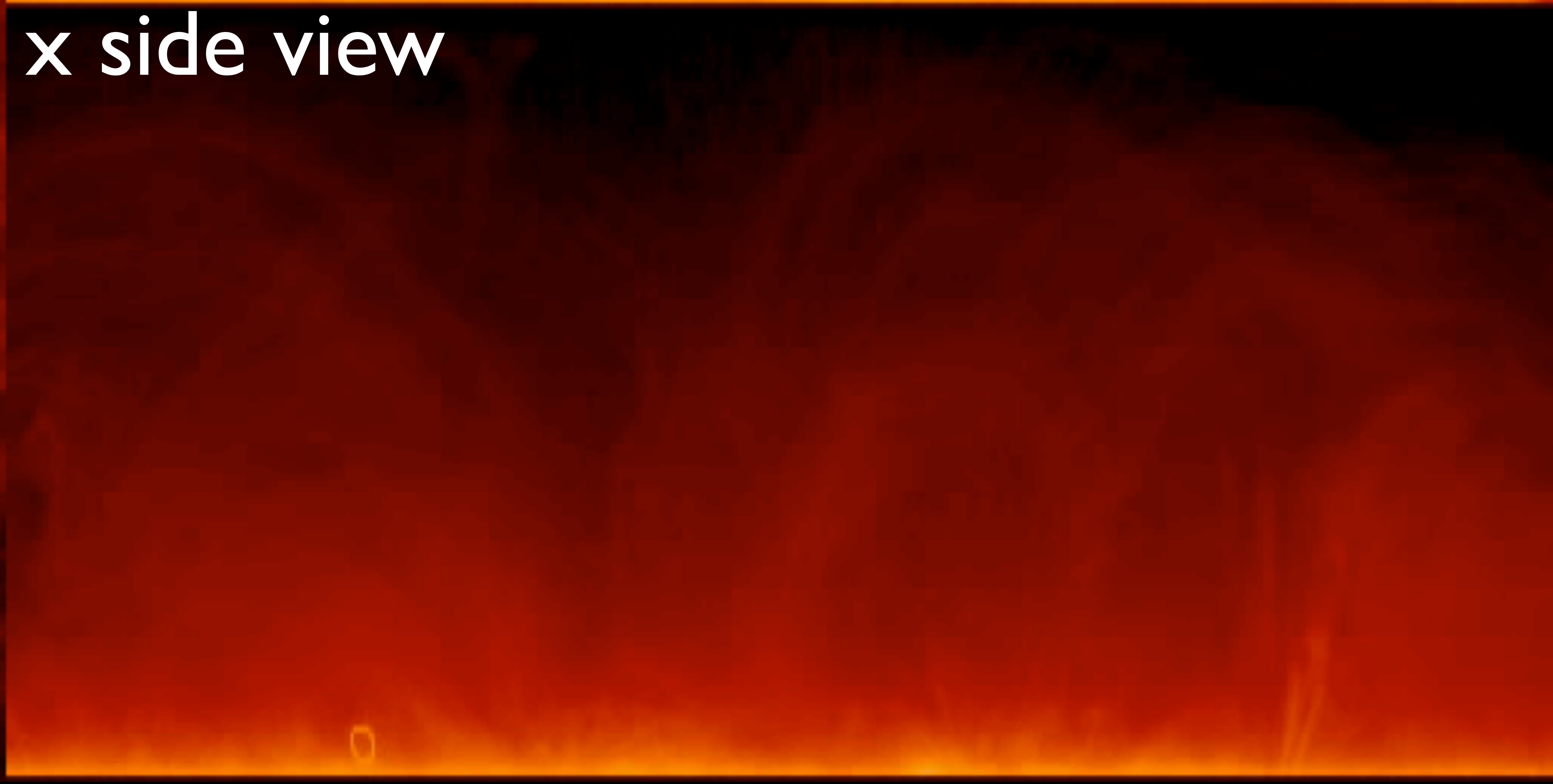
Empirical Data-Driven Models:

- Physics-rules not prescribed. Try to discover relations in the data.


2011/02/12 00:00:00
HMI vector magnetogram sequence of NOAA AR 11158
Credit: Keiji Hayashi (HMI)

Visualization of Field Lines





Orange ~ $\int_{los} \langle j^2 \rangle dl$, where $\langle j^2 \rangle$ is fieldline-averaged j^2. Positive polarity B_{r}. Negative polarity B_r

Data-Driven Space Weather Models

Physics-based Models:

- Data-inspired Models: Simplified simulations to mimic observed scenarios
- Data-constrained Models: Time-independent models satisfying observations at an instant in time. Includes models that may start with a data-constrained initial condition but driven by idealized boundary conditions.
- Data-Driven Models: Time-dependent models evolved in response to evolving boundary conditions

Empirical Data-Driven Models:

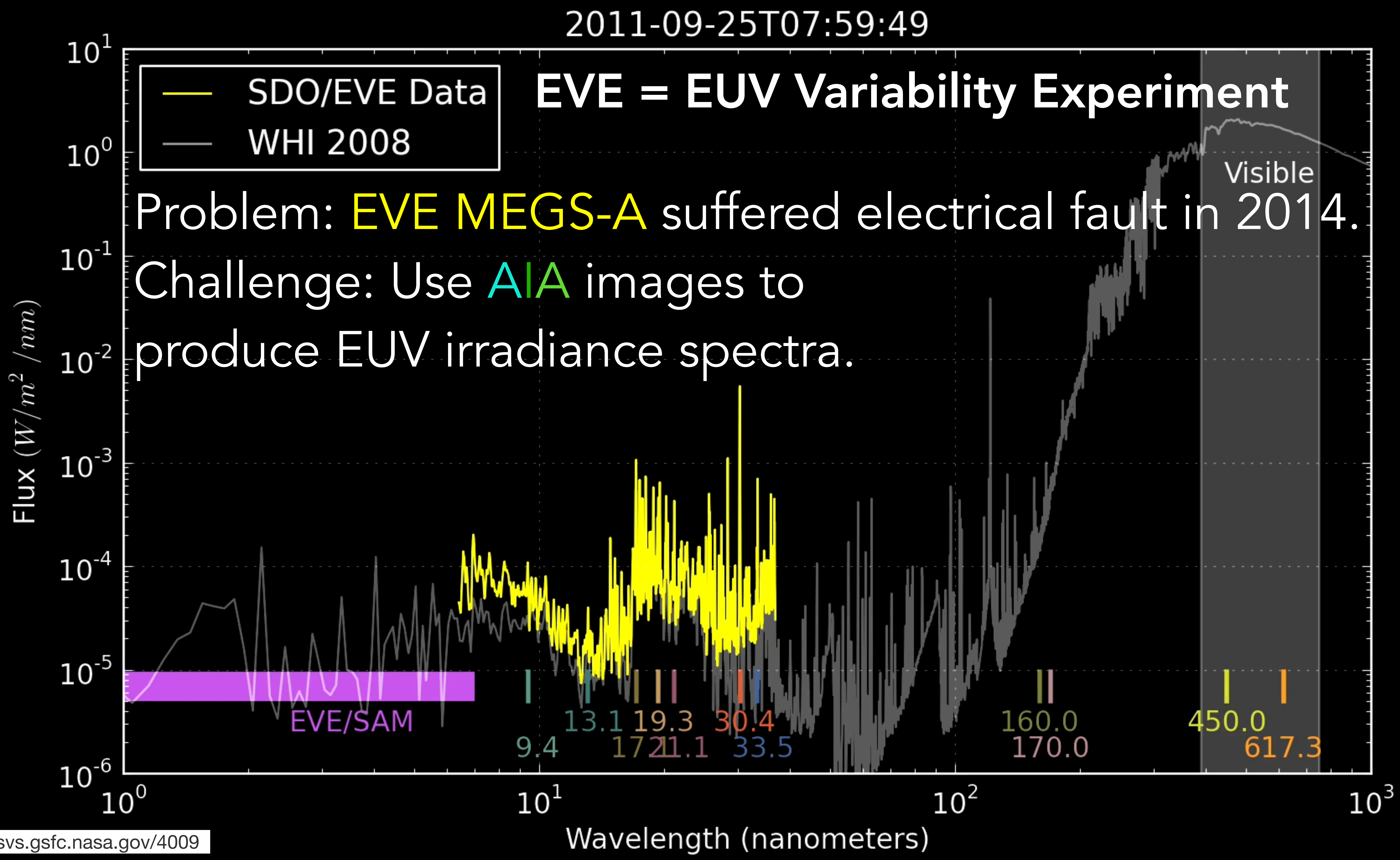
- Physics-rules not prescribed. Try to discover relations in the data.

2019 Challenges

- SDO ML (Cheung, Janvier & Jin)
- GNSS (Bhatt, continued from 2018)
- Hi resolution magnetograms over multiple cycles (Munoz-Jaramillo & Wright)

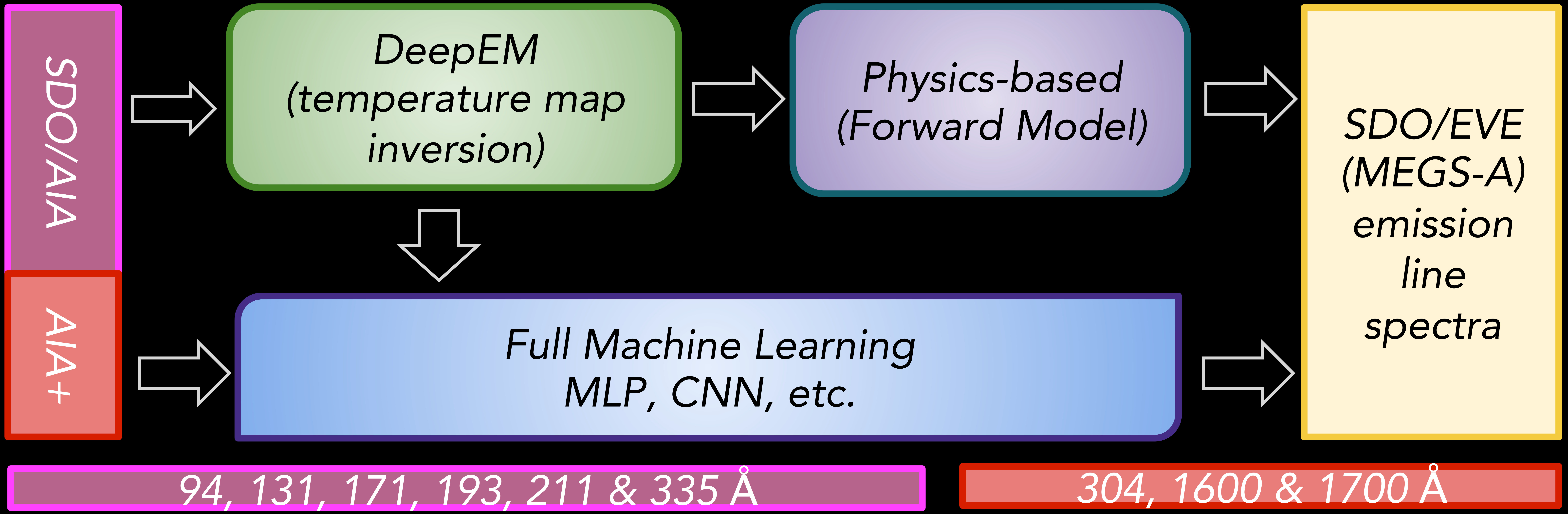
NASA Ames Research Center - Silicon Valley - 2018

NASA FRONTIER DEVELOPMENT LAB

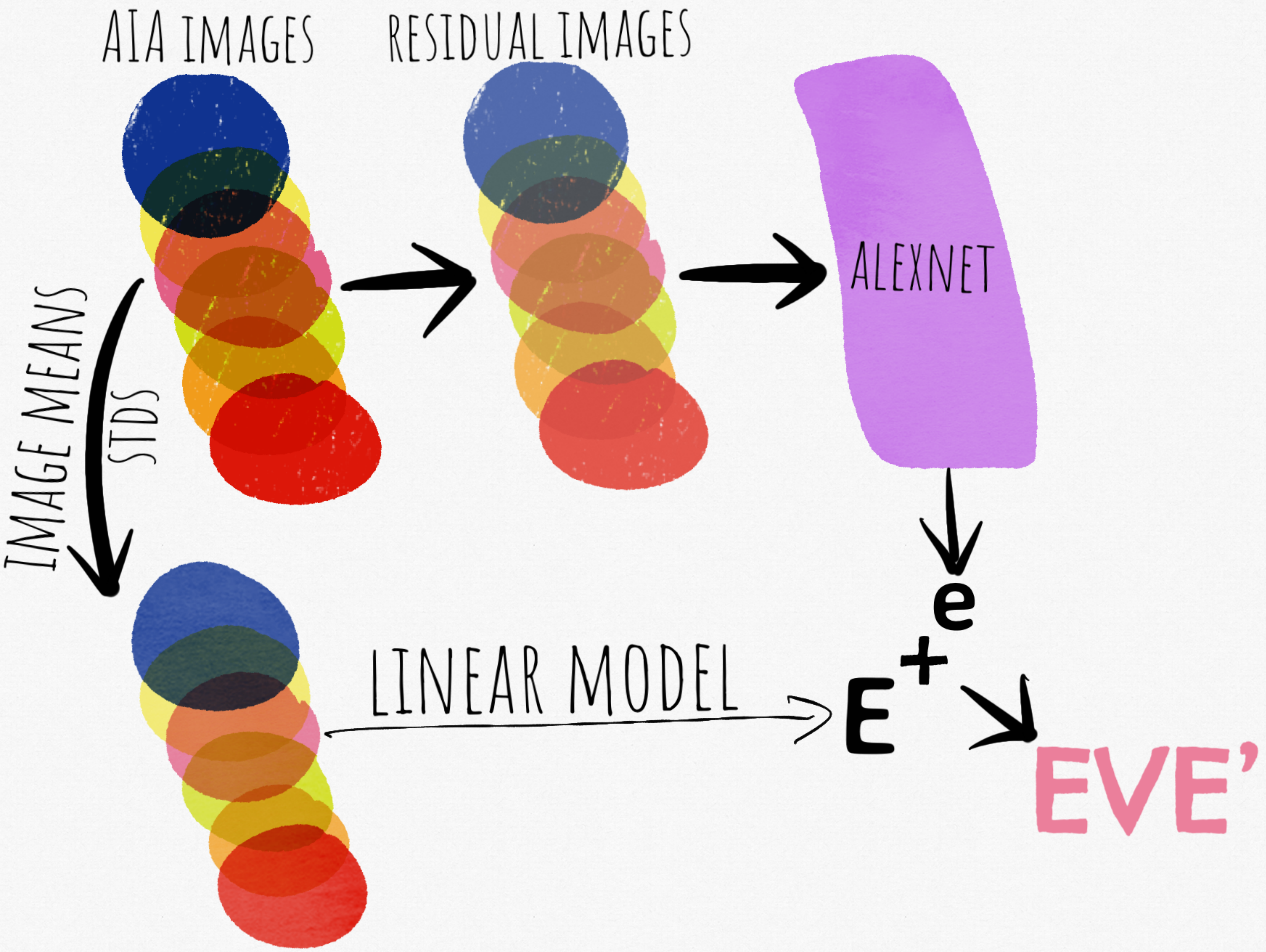


http://svs.gsfc.nasa.gov/4009

Model Approaches



Best Model: Linear Model on [AIA Means, AIA stds]
+ AlexNet on Residuals + Average Pool
<Rel Err> **< 5% for all emission lines, < 2% for most lines**


AIA IMAGES
RESIDUAL IMAGES
ALEXNET
IMAGE MEANS
STDS
e
+
E
LINEAR MODEL
EVE'

The Astrophysical Journal Supplement Series, 242:7 (11pp), 2019 May https://doi.org/10.3847/1538-4365/ab1005

A Machine-learning Data Set Prepared from the NASA *Solar Dynamics Observatory* Mission

Richard Galvez[1], David F. Fouhey[2], Meng Jin[3,4], Alexandre Szenicer[5], Andrés Muñoz-Jaramillo[6], Mark C. M. Cheung[3,7], Paul J. Wright[8], Monica G. Bobra[7], Yang Liu[7], James Mason[9], and Rajat Thomas[10]
[1] Center for Data Science, New York University, New York, NY 10011, USA; richard.galvez@nyu.edu
[2] University of Michigan, Ann Arbor, MI 48109, USA
[3] Lockheed Martin Solar & Astrophysics Laboratory, Palo Alto, CA, USA
[4] SETI Institute, Mountain View, CA 94043, USA
[5] University of Oxford, Oxford OX1 2JD, UK
[6] Southwest Research Institute, San Antonio, TX 78238, USA
[7] Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305, USA
[8] SUPA School of Physics & Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
[9] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[10] University of Amsterdam, 1012 WX Amsterdam, Netherlands
Received 2019 January 19; revised 2019 March 11; accepted 2019 March 13; published 2019 May 8

Ready for use with numpy, sklearn, python DL frameworks.

Abstract

In this paper, we present a curated data set from the NASA *Solar Dynamics Observatory* (*SDO*) mission in a format suitable for machine-learning research. Beginning from level 1 scientific products we have processed various instrumental corrections, down-sampled to manageable spatial and temporal resolutions, and synchronized observations spatially and temporally. We illustrate the use of this data set with two example applications: forecasting future extreme ultraviolet (EUV) Variability Experiment (EVE) irradiance from present EVE irradiance and translating Helioseismic and Magnetic Imager observations into Atmospheric Imaging Assembly observations. For each application, we provide metrics and baselines for future model comparison. We anticipate this curated data set will facilitate machine-learning research in heliophysics and the physical sciences generally, increasing the scientific return of the *SDO* mission. This work is a direct result of the 2018 NASA Frontier Development Laboratory Program. Please see the Appendix for access to the data set, totaling 6.5TBs.

Key words: astronomical databases: miscellaneous – catalogs – editorials, notices – miscellaneous – surveys

ML reveals systematic accumulation of electric current in lead-up to solar flares

Table 4. Average values of SHARP features over flaring and nonflaring AR magnetic-field observations categorized by the SVM

		Flaring ARs, >72 h from flare		Nonflaring ARs		
Symbol	Brief description	TP	FN	FP	TN	(TP − TN)/σ_{TN}
USFLUX, 10^{22} Mx	Total unsigned flux	3.30 ± 0.29	1.07 ± 0.11	2.48 ± 0.19	0.56 ± 0.03	94.71
TOTUSJH, 10^2 G^2/m	Total unsigned current helicity	23.87 ± 2.09	7.89 ± 0.81	19.45 ± 1.42	4.29 ± 0.21	91.27
TOTBSQ, 10^{10} G^2	Total Lorentz force	4.43 ± 0.41	1.53 ± 0.17	3.57 ± 0.32	0.83 ± 0.04	89.10
TOTUSJZ, 10^{13} A	Total unsigned vertical current	5.43 ± 0.45	1.86 ± 0.21	4.48 ± 0.35	1.00 ± 0.05	85.68
TOTFZ, 10^{23} dyne	Total vertical Lorentz force	−3.30 ± 0.51	−0.61 ± 0.19	−1.55 ± 0.20	−0.34 ± 0.04	78.54
SAVNCPP, 10^{13} A	Sum of net current per polarity	1.14 ± 0.11	0.39 ± 0.05	0.93 ± 0.10	0.24 ± 0.01	74.70
ABSNJZH, G^2/m	Absolute net current helicity	254.59 ± 31.95	68.37 ± 12.66	208.28 ± 28.53	40.68 ± 2.92	73.23
TOTPOT, 10^{23} erg/cm	Total magnetic free energy	5.20 ± 0.61	1.44 ± 0.30	4.80 ± 0.61	0.72 ± 0.06	71.71
AREA, Mm^2	AR area	262.17 ± 19.99	110.95 ± 11.88	222.82 ± 18.35	62.75 ± 2.81	71.00
R_VALUE, Mx	Flux near polarity inversion line	4.06 ± 0.09	2.81 ± 0.25	4.04 ± 0.08	2.09 ± 0.09	20.95
SHRGT45, %	Area with shear >45°	29.76 ± 1.85	24.87 ± 3.81	37.30 ± 2.04	20.24 ± 1.17	8.10
MEANPOT, 10^2 erg/cm^3	Mean magnetic free energy	68.34 ± 4.61	54.82 ± 10.08	81.42 ± 6.64	45.51 ± 3.04	7.51

True positives (TP) and false negatives (FN) are observations from flaring ARs which are classified as flaring and nonflaring, respectively. True negatives (TN) and false positives (FP) are observations from nonflaring ARs that are classified as nonflaring and flaring, respectively.

Dhuri, Hanasoge & Cheung (PNAS 20th May 2019)

Follows ML approach of Bobra & Couvidat (2015)

SDO/HMI Vector Magnetograms Important for Flare Prediction

Summary

https://tinyurl.com/ictp-cheung

Data-inspired, Data-constrained, Data-Driven Physics based models.

NASA's Solar Dynamics Observatory a poster child for a successful science mission that also contributes to operational space weather. How?

- Open data policy (legacy of SOHO; NASA Heliophysics leadership): Near-real time (nearly science quality) data available within minutes. Final science data available within days.
- Meta-databases (HEK) + APIs (not just some passive FTP site)
- Instruments and investigations operated by teams who care about the science.

BTW Don't take SDO for granted. No space-based Sun-Earth line solar magnetogram funded.

Open data + open source software + machine learning frameworks + (relatively) inexpensive compute will let us use our sensor networks more effectively, and make improvements to space weather predictions.

Backup slides

33

Modeling Geomagnetic Variations using a Machine Learning Framework

Mark C. M. Cheung[1,2,3] (cheung@lmsal.com), Casey Handmer[3], Burçu Kosar[4,3,*], George Gerules[3], Bala Poduval[5,3,*], Graham Mackintosh[3], Andrés Muñoz-Jaramillo[6,3], Monica Bobra[2,3], Troy Hernandez[7,3], Ryan McGranaghan[8,3]

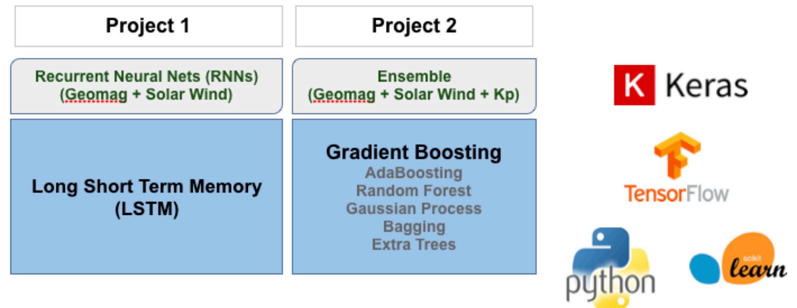
[1]Lockheed Martin Solar & Astrophysics Lab, Palo Alto, CA, USA, [2]Stanford University, CA, USA, [3]NASA Frontier Development Lab, Mountain View, CA, USA, [4]NASA Goddard Space Flight Center, MD, USA, [5]Space Science Institute, Boulder, CO, USA, [6]Southwest Research Institute, Boulder, CO, USA, [7]IBM, Chicago, IL, USA, [8]Jet Propulsion Laboratory, Pasadena, CA, USA, *presenting authors

AGU 2017 poster

NASA FRONTIER DEVELOPMENT LAB 2017

- an Artificial Intelligence (AI) R &D accelerator
- to tackle important questions in the space sciences
- an intense 8–week focused study
- topics: Planetary Defense, Space Weather and Space Resources

MACHINE LEARNING/AI



Keras: An open source neural network (NN) library written in Python.
Scikit-Learn: A free machine learning library for Python, featuring various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, etc.. It is designed to operate with the Python numerical and scientific libraries NumPy and SciPy.
TensorFlow: Another open source software library for machine learning, designed for building and training deep neural networks to detect and decipher patterns and correlations.

Kp INDEX

The K-indices quantify the disturbances in the horizontal component of geomagnetic field, represented by an integer in the range 0-9. It is derived from the maximum fluctuations of horizontal components during three–hour intervals. The planetary index Kp is the mean of standardized K-indices from 13 stations between 44° and 60° N/S geomagnetic latitude. NOAA/Space Weather Prediction Center (SWPC) makes use of the Kp index when issuing geomagnetic storm warnings.

G-Scale	Kp	Activity Level	Occurrence Frequency
G0	4 & lower	Below Storm	
G1	5	Minor Storm	1700 per cycle (900 days per cycle)
G2	6	Moderate Storm	600 per cycle (360 days per cycle)
G3	7	Strong Storm	200 per cycle (130 days per cycle)
G4	8	Severe Storm	100 per cycle (60 days per cycle)
G5	9	Extreme Storm	4 per cycle (4 days per cycle)

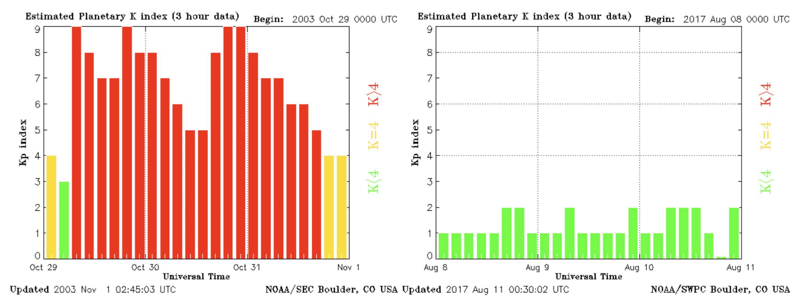


Figure : 1 Kp index during Halloween event (left) and during a very quiet period (right).

THE PROBLEM DEFINITION

(Q1) Can we apply machine learning (ML) to forecast geomagnetic variability using solar wind and ground-based measurements?

(Q2) Without imposing a priori, first-principles based, physical models of the solar wind-driven geomagnetic system, what insights can ML extract from the data?

METRIC OF ACCURACY

We obtained the mean square errors between observed and predicted Kp indices using various models. Also, we computed the p-statistics to determine the statistical significance of how well the models do compared with each other. With $> 95\%$ confidence, the models have different performance metrics.

ML method	1h ahead	3h ahead	6h ahead
Persist	0.007	0.020	0.025
Mean	0.046	0.046	0.046
Median	0.048	0.048	0.048
Gradient Boosting	0.007	0.015	0.021
Adaptive Boost	0.012	0.018	0.032
Extra Trees	0.009	0.021	0.027
Random Forest	0.015	0.015	0.026

> 95% confidence level

SPACE WEATHER EVENTS & THEIR SOCIO-ECONOMIC IMPACTS

- Space Weather: Solar-driven fluctuations in the near-Earth environment leading to disruptions and damages to our critical infrastructure and technological systems in space and on Earth.
- Space Weather events: Solar flares, coronal mass ejections (CMEs), solar energetic particle (SEP) events, solar radio bursts, geomagnetic disturbances
- Space Weather impacts: Disruptions in wireless communications, Global Positioning System (GPS), satellite operations and communication, aviation, and the electrical power grid.
- Space Weather forecast: Using physics-based and empirical models to mitigate the impacts of extreme space weather events (National Space Weather Action Plan - SWAP). Improved predictions offer better protection for space weather stakeholders.

Scale	Description	Effect	Physical measure	Average Frequency (1 cycle = 11 years)
G 5	Extreme	**Power systems:** Widespread voltage control problems and protective system problems can occur, some grid systems may experience complete collapse or blackouts. Transformers may experience damage. **Spacecraft operations:** May experience extensive surface charging, problems with orientation, uplink/downlink and tracking satellites. **Other systems:** Pipeline currents can reach hundreds of amps, HF (high frequency) radio propagation may be impossible in many areas for one to two days, satellite navigation may be degraded for days, low-frequency radio navigation can be out for hours, and aurora has been seen as low as Florida and southern Texas (typically 40° geomagnetic lat.).	Kp = 9	4 per cycle (4 days per cycle)
G 4	Severe	**Power systems:** Possible widespread voltage control problems and some protective systems will mistakenly trip out key assets from the grid. **Spacecraft operations:** May experience surface charging and tracking problems, corrections may be needed for orientation problems. **Other systems:** Induced pipeline currents affect preventive measures, HF radio propagation sporadic, satellite navigation degraded for hours, low-frequency radio navigation disrupted, and aurora has been seen as low as Alabama and northern California (typically 45° geomagnetic lat.).	Kp = 8, including a 9-	100 per cycle (60 days per cycle)
G 3	Strong	**Power systems:** Voltage corrections may be required, false alarms triggered on some protection devices. **Spacecraft operations:** Surface charging may occur on satellite components, drag may increase on low-Earth-orbit satellites, and corrections may be needed for orientation problems. **Other systems:** Intermittent satellite navigation and low-frequency radio navigation problems may occur, HF radio may be intermittent, and aurora has been seen as low as Illinois and Oregon (typically 50° geomagnetic lat.).	Kp = 7	200 per cycle (130 days per cycle)
G 2	Moderate	**Power systems:** High-latitude power systems may experience voltage alarms, long-duration storms may cause transformer damage. **Spacecraft operations:** Corrective actions to orientation may be required by ground control; possible changes in drag affect orbit predictions. **Other systems:** HF radio propagation can fade at higher latitudes, and aurora has been seen as low as New York and Idaho (typically 55° geomagnetic lat.).	Kp = 6	600 per cycle (360 days per cycle)
G 1	Minor	**Power systems:** Weak power grid fluctuations can occur. **Spacecraft operations:** Minor impact on satellite operations possible. **Other systems:** Migratory animals are affected at this and higher levels; aurora is commonly visible at high latitudes (northern Michigan and Maine).	Kp = 5	1700 per cycle (900 days per cycle)

Date	Event	Level
1 Sept 1859	Carrington Event widespread disruption of telegraph	Extreme
13 March 1989	Hydro-Quebec 9 hour black out	Severe
20/21 Jan 1994	Anik-E1 and Anik-E2 failed Disrupted TV and computer transmission	Moderate
14 July 2000	Bastille Day Event	Extreme
31 October 2003	Halloween Events Affected airlines, caused power outages, damaged transformers, led astronauts on ISS to take shelter	Extreme

DATA USED

Period of Study: 2016 (descending phase of Solar cycles 24)

- Observed solar wind properties: Multispacecraft compilation of solar wind observations at Lagrangian point 1: http://omniweb.gsfc.nasa.gov/. solar wind speed, proton density, heliospheric magnetic field (HMF) intensity, HMF B_z, etc.
- Geomagnetic field measurements - 14 US stations operated by US Geological Survey.
- Kp (planetary K) index.

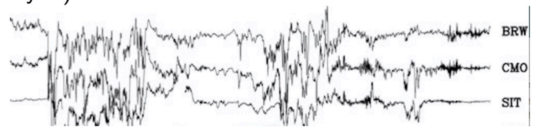


Figure : 3a Geomagnetic field data.

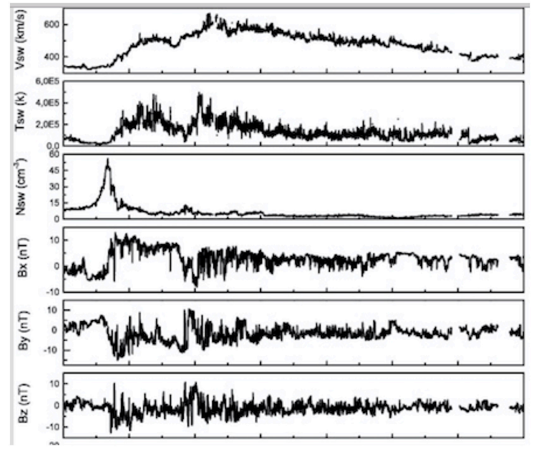


Figure : 3b Solar wind data.

PREDICTED Kp

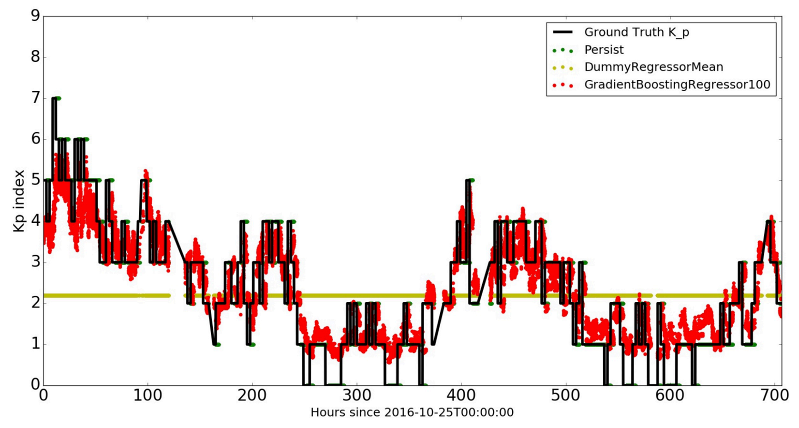
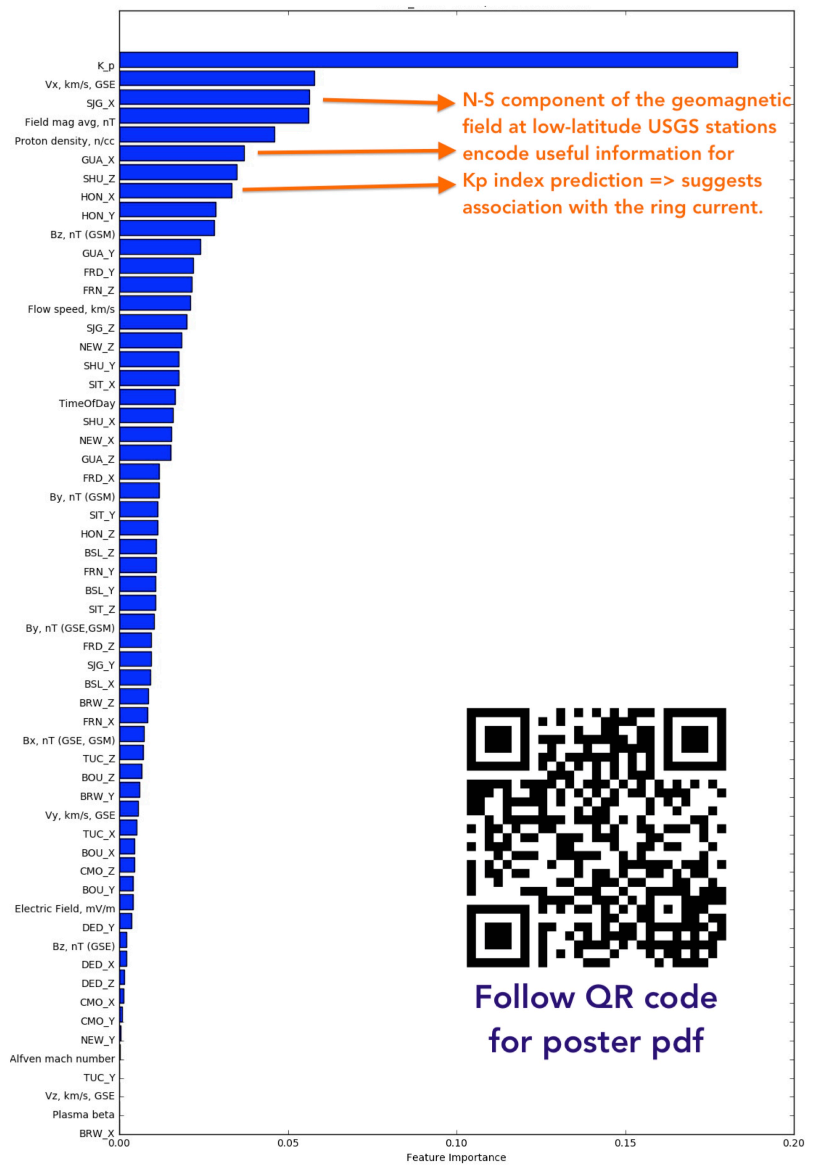


Figure : 4 Actual observed Kp (calculated from ground observations) (black dots) and corresponding values from 3-hr ahead forecast using a persistence (dark green dots), global mean (light green dots) and gradient boosting (yellow dots) models.

We used nearly 7 months of data to train the model and then tested the model by predicting the Kp indices for 3 months (Figure 4 shows a subset of the test data). The training and testing data were partitioned such that the models have not seen data with any overlap between the two sets.
The Gradient Boosting Regressor model provided the best results, consistently beating a persistence model (i.e. the current Kp index predicted not to change in the future) and various machine learning models in scikit-learn, with a confidence level $> 95\%$.
Also, the Gradient Boosting model ranks the input features by their relative importance for creating a good prediction (Figure 5).

RELATIVE IMPORTANCE OF INPUT PARAMETERS



Follow QR code for poster pdf

Figure : 5 Relative importance of input parameters in the prediction of Kp using the Gradient Boosting Regressor model.

SUMMARY & CONCLUDING REMARKS

Without prior domain knowledge, the model learned that the most important precursor is the current Kp index.
Other important factors:
- Solar wind speed and proton density.
- Solar wind magnetic field strength and Bz.

Moreover, the model suggested that the N-S component of the geomagnetic field at low latitude stations - Guam (GUA), Hawaii (HON), Puerto Rico (SJG), are also important precursors. These quantities are largely influenced by ring current and therefore, this finding implies the importance of considering the effects of ring current in the prediction of geomagnetic storm. This result came as a total surprise since the machine learning algorithm was not expected to be capable of learning such heuristics without prior knowledge!
Scope: Based on the results we feel confident that the method can be applied to address other aspects of the socio-economic impact of space weather by predicting the appropriate variable if sufficient data exist.
Ultimate goal: To couple the complex and dynamic solar-terrestrial system using AI.

ACKNOWLEDGEMENTS

Thanks are also due to IBM for providing the computing resources for the project and NVIDIA, Intel, Lockheed Martin and Kx for their partnership in the FDL 2017.
B. Poduval wishes to acknowledge Dr. J. Love and colleagues, USGS; Dr. Jennifer Ganon, CPI and Robert Steenburgh, NOAA, for the many discussions helpful for this project.

ACRONYMS

AI	Artificial Intelligence	GB	Gradient Boosting Regressor
ML	Machine Learning	NN	Neural Network
RMSE	Root Mean Square Error	Kp index	Planetary K index
K	'Kennziffer' for 'characteristic digit.'		