

# Development of Intelligent Databases and Analysis Tools for Heliophysics

Alexander Kosovichev<sup>1</sup>, Gelu Nita<sup>1</sup>, Vincent Oria<sup>1</sup>, Viacheslav Sadykov<sup>2</sup>, Egor Illarionov<sup>3</sup>, and Andrey Tlatov<sup>4</sup>

<sup>1</sup>New Jersey Institute of Technology

<sup>2</sup>Bay Area Environmental Research Institute

<sup>3</sup>Moscow State University

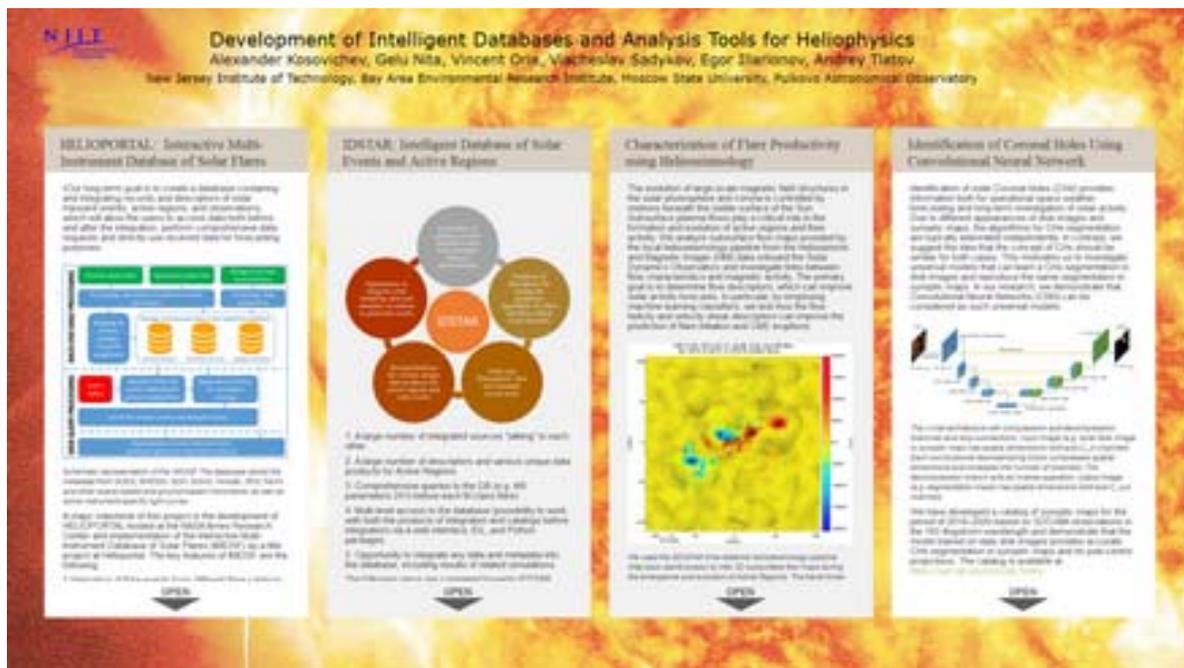
<sup>4</sup>Pulkovo Observatory

November 26, 2022

## Abstract

The project develops innovative tools to extract and analyze the available observational and modeling data in order to enable new physics-based and machine-learning approaches for understanding and predicting solar activity and its influence on the geospace and Earth systems. The heliophysics data are abundant: several terabytes of solar and space observations are obtained every day. Finding the relevant information from numerous spacecraft and ground-based data archives and using it is paramount, and currently a difficult task. The scope of the project is to develop and evaluate data integration tools to meet common data access and discovery needs for two types of Heliophysics data: 1) long-term synoptic activity and variability, and 2) extreme geoeffective solar events caused by solar flares and eruptions. The methodology consists in the development of a data integration infrastructure and access methods capable of 1) automatic search and identification of image patterns and event data records produced by space and ground-based observatories, 2) automatic association of parallel multi-wavelength/multi-instrument database entries with unique patterns or event identifiers, 3) automatic retrieval of such data records and pipeline processing for the purpose of annotating each pattern or event according to a predefined set of physical parameters inferable from complementary data sources, and 4) generation of a pattern or catalog and associated user-friendly graphical interface tools that are capable to provide fast search, quick preview, and automatic data retrieval capabilities. The Team has developed and implemented the Heliportal that provides a synergy of solar flare observations, taking advantage of big datasets from the ground- and space-based instruments, and allows the larger research community to significantly speed up investigations of flare events, perform a broad range of new statistical and case studies, and test and validate theoretical and computational models. The Heliportal accumulates, integrates and presents records of physical descriptors of solar flares, as well as the magnetic characteristic of active regions from various catalogs of observational data from different observatories and heliophysics missions.

# Development of Intelligent Databases and Analysis Tools for Heliophysics



Alexander Kosovichev, Gelu Nita, Vincent Oria, Viacheslav Sadykov, Egor Illarionov, Andrey Tlatov

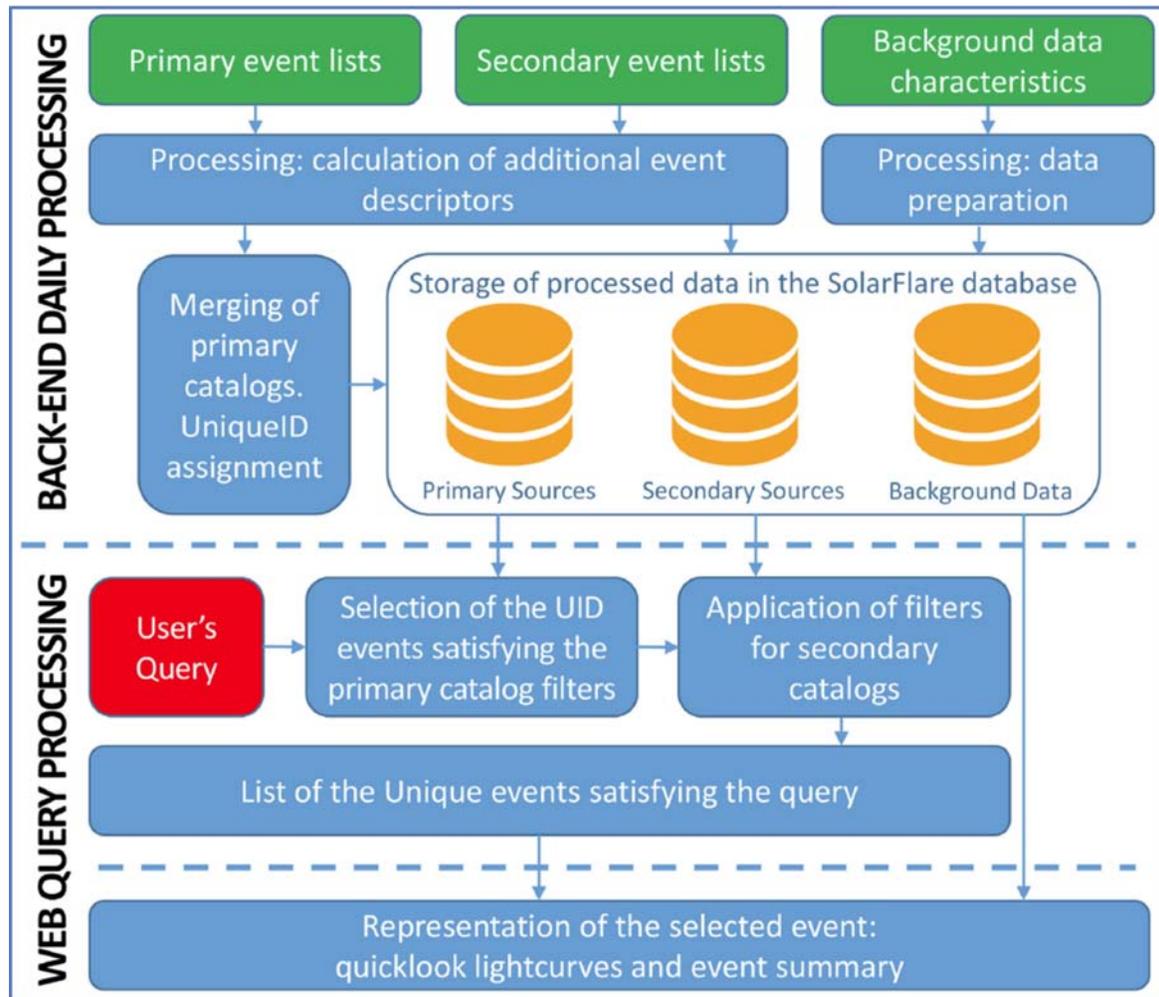
New Jersey Institute of Technology, Bay Area Environmental Research Institute, Moscow State University, Pulkovo Astronomical Observatory

PRESENTED AT:



# HELIOPORTAL: INTERACTIVE MULTI-INSTRUMENT DATABASE OF SOLAR FLARES

•Our long-term goal is to create a database containing and integrating records and descriptors of solar transient events, active regions, and observations, which will allow the users to access data both before and after the integration, perform comprehensive data requests and directly use received data for forecasting purposes.



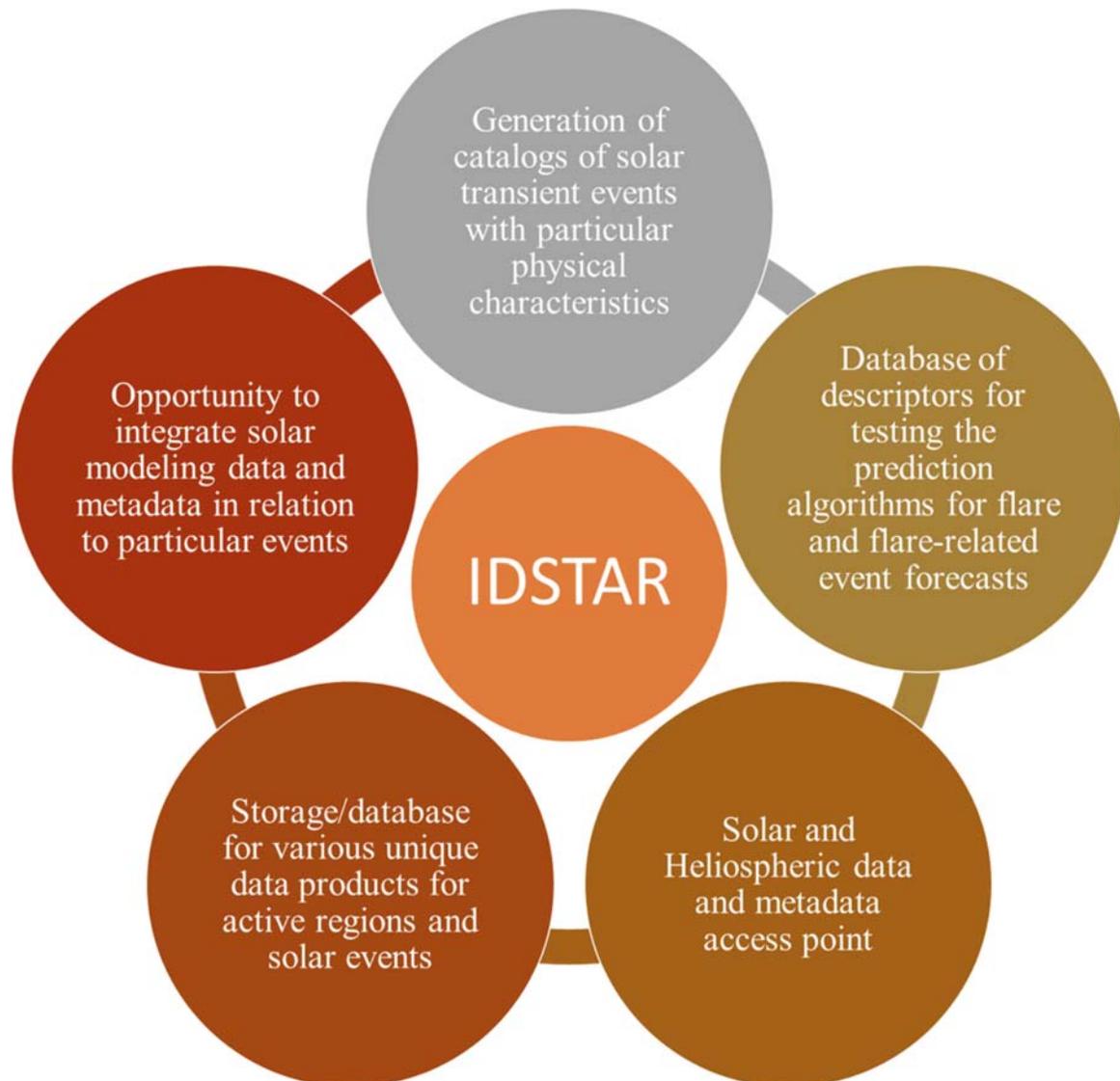
*Schematic representation of the IMDSF. The database stores the metadata from GOES, RHESSI, SDO, SOHO, Hinode, IRIS, Fermi, and other space-based and ground-based instruments, as well as some instrument-specific light curves.*

A major milestone of this project is the development of HELIOPORTAL hosted at the NASA Ames Research Center and implementation of the Interactive Multi-Instrument Database of Solar Flares (IMDSF) as a title project at Heliportal. The key features of IMDSF are the following:

1. Integration of flare events from different flare catalogs (GOES, RHESSI, HEK, Hinode, Fermi GBM, Konus-WIND, the OVSA flare catalogs). "One flare – one result" philosophy.
2. A search of the flare events based on their physical descriptors (both those stored in the catalogs and calculated), availability of observations (IRIS and Nobeyama observational filters are available), or presence of accompanying events (solar energetic particles, coronal mass ejections, etc.)
3. A detailed look at the particular event which includes quicklook light curves (GOES, NORP and ESP/EVE SDO), cross-mission summary for each event, user-defined searches for similar events.

IMDSF and Heliportal are available at <https://solarflare.njit.edu/> and <https://heliportal.nas.nasa.gov>

# IDSTAR: INTELLIGENT DATABASE OF SOLAR EVENTS AND ACTIVE REGIONS



1. A large number of integrated sources “talking” to each other
2. A large number of descriptors and various unique data products for Active Regions
3. Comprehensive queries to the DB (e.g. AR parameters 24 h before each M-class flare)
4. Multi-level access to the database (possibility to work with both the products of integration and catalogs before integration) via a web interface, IDL, and Python packages
5. Opportunity to integrate any data and metadata into the database, including results of related simulations.

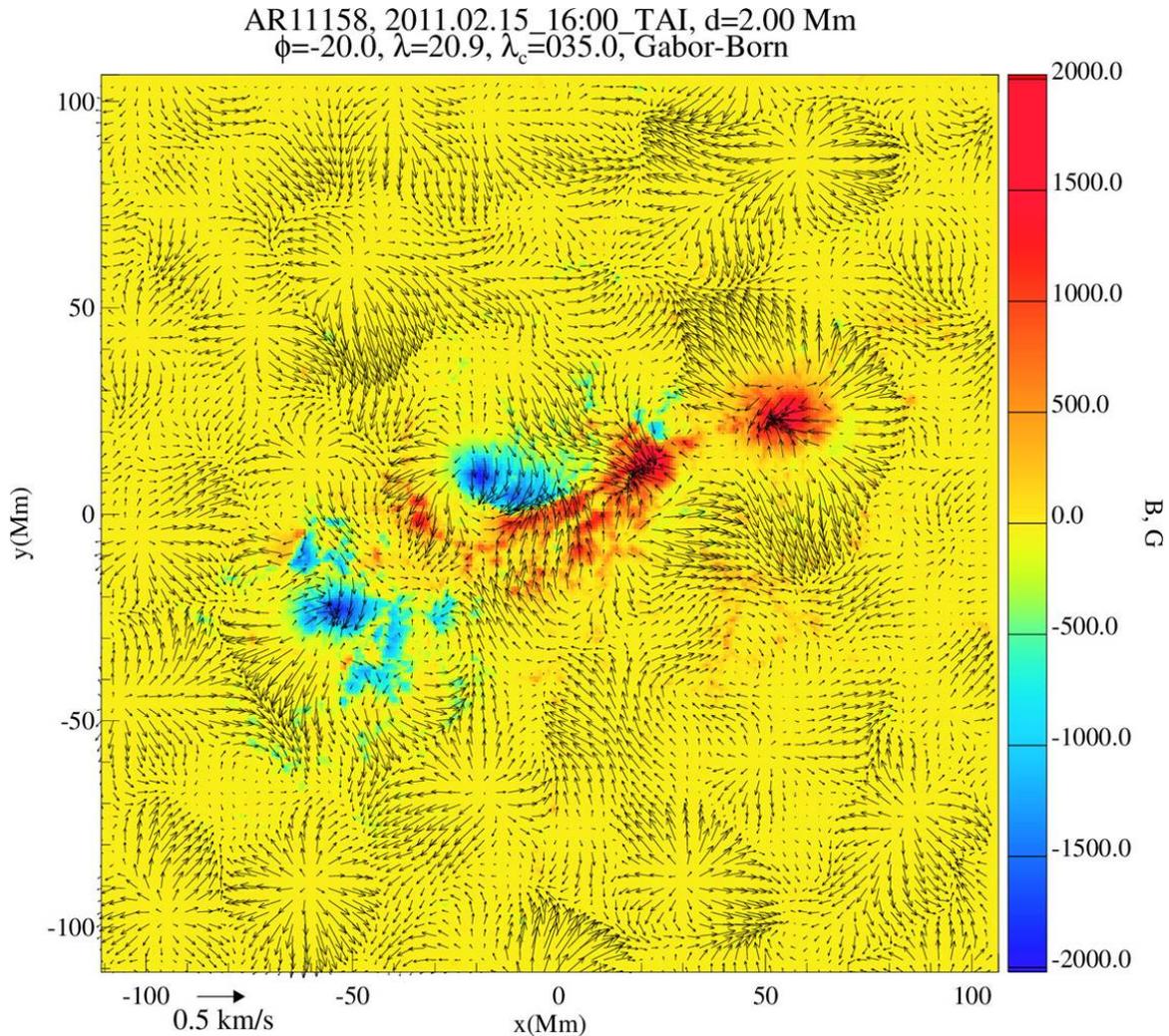
The following steps are completed towards IDSTAR implementation:

1. The relational MySQL database is set up at NJIT server. The database follows ERD structure and is indexed properly.
2. The data from IMDSF into IDSTAR is transferred.

3. The records from different flare catalogs are integrated, and unique flare identifiers (UniqueIDs) for the flare records are assigned.
4. The SHARP data are loaded from JSOC/Stanford, as well as some other AR descriptors (e.g. properties of the Polarity Inversion Line, PIL, Sadykov and Kosovichev 2017)
5. Non-Linear Force-Free Field (NLFFF) extrapolations are calculated for the flaring and non-flaring ARs. The phases of calculations are described below.
6. Other AR descriptors and data products, such as subsurface flow maps (Zhao et al. 2012) or ENHEL descriptors (Tziotziou et al. 2014) are tested

# CHARACTERIZATION OF FLARE PRODUCTIVITY USING HELIOSEISMOLOGY

The evolution of large-scale magnetic field structures in the solar photosphere and corona is controlled by motions beneath the visible surface of the Sun. Subsurface plasma flows play a critical role in the formation and evolution of active regions and their activity. We analyze subsurface flow maps provided by the local helioseismology pipeline from the Helioseismic and Magnetic Imager (HMI) data onboard the Solar Dynamics Observatory and investigate links between flow characteristics and magnetic activity. The primary goal is to determine flow descriptors, which can improve solar activity forecasts. In particular, by employing machine learning classifiers, we test how the flow helicity and velocity shear descriptors can improve the prediction of flare initiation and CME eruptions.



We used the SDO/HMI time-distance helioseismology pipeline (<http://jsoc.stanford.edu>) to infer 3D subsurface flow maps during the emergence and evolution of Active Regions. The travel times are used for reconstruction of subsurface flows in 8 subsurface layers in the following depth ranges: 0-1, 1-3, 3-5, 5-7, 7-10, 10-13, 13-17, and 17-21 Mm, and with a horizontal spatial sampling of 0.12 degrees (1.5 Mm). The horizontal flows for AR11158 at the depth of 2 Mm are shown by arrows in the figures below, and the background color images are the corresponding photospheric magnetograms.

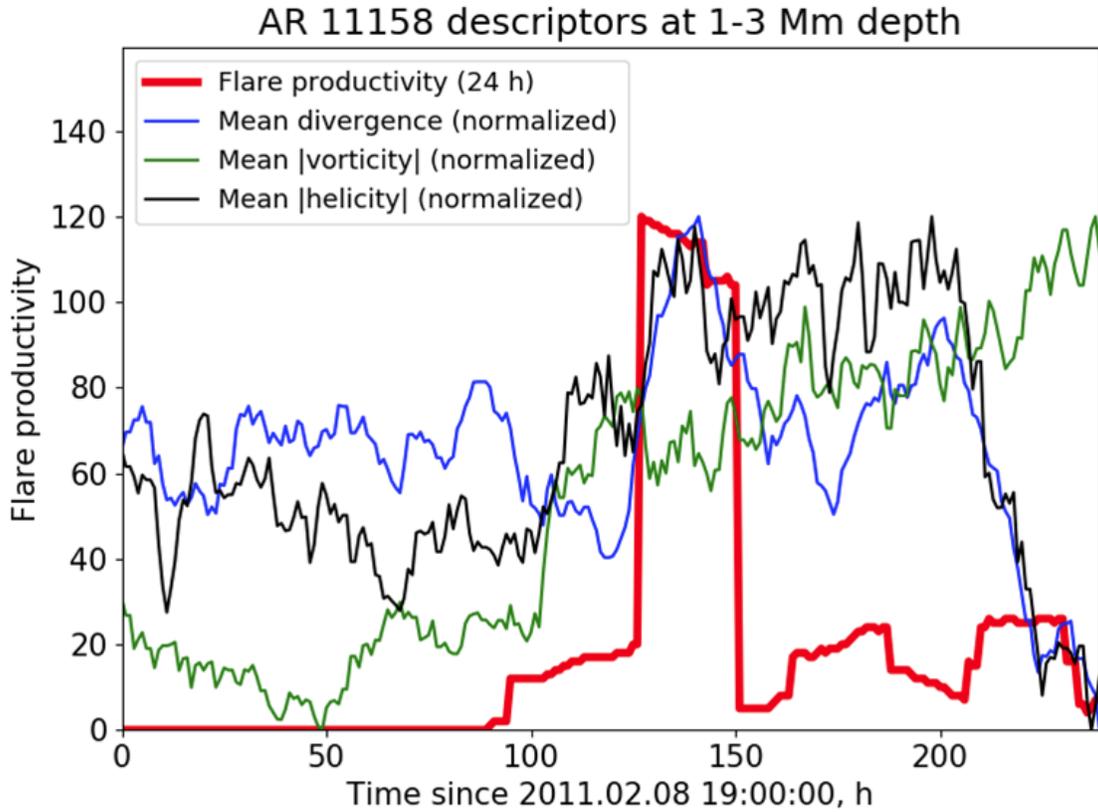
We derive the following physical descriptors to characterize the subsurface flow maps:

1. a layer-averaged horizontal divergence computed based on the horizontal velocity components;
2. a layer-averaged vertical component of the velocity curl (vorticity);
3. a proxy for the kinetic helicity defined as a product of the horizontal divergence and the vertical component of vorticity.

We define the flare productivity of an active region as  $P=NC+10 \times NM+100 \times NX$  where NC, NM, and NX are the total number of

C-class, M-class, and X-class GOES flares happening in the AR within 24 or 48 hours from the considered moment. The flare productivity is used for correlation analysis with the AR magnetic and flow descriptors.

We analyze correlations of the derived descriptors with the flare productivity of the parental active region within the next 24-hour window. In addition to classically-used Pearson's correlation coefficients, which check for linear dependence between pairs of parameters, we also calculate the non-parametric Kendall's tau correlation coefficient.



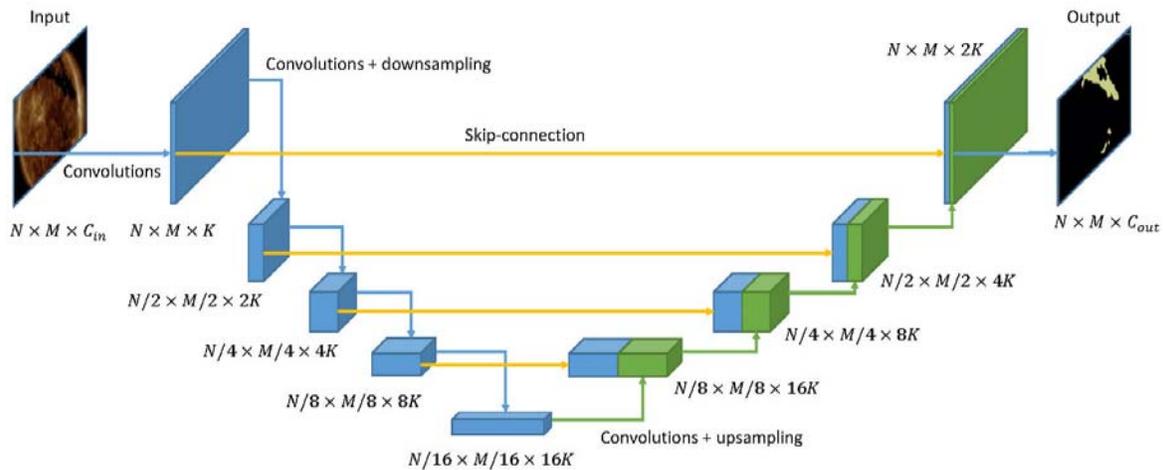
*The relationship between the flare productivity and characteristics of subsurface flows.*

*Table 1. Strongest Kendall's tau rank correlation coefficients and corresponding time lags (varying from 0 h to 24 h) found for subsurface flow map descriptors at 1-3 Mm depth and flare productivity. The correlations are obtained for 8 ARs tracked for 10 days with a 1-hour cadence. Red entries indicate cases where correlation coefficients were not statistically significant (the corresponding p-value is < 0.05).*

Active Region	Total divergence at 1-3 Mm	Total vorticity at 1-3 Mm	Total kinetic helicity at 1-3 Mm
11158	0.14 / 0 h	0.53 / 7 h	0.47 / 0 h
11283	0.61 / 0 h	-0.27 / 18 h	0.40 / 15 h
11305	0.65 / 20 h	-0.25 / 23 h	0.46 / 2 h
11618	0.30 / 24 h	-0.43 / 7 h	0.08 / 0 h
11875	-0.38 / 0 h	0.10 / 7 h	-0.16 / 22 h
11967	0.15 / 6 h	0.20 / 0 h	0.48 / 11 h
12253	0.41 / 24 h	0.54 / 24 h	0.65 / 2 h
12673	0.14 / 19 h	0.65 / 0 h	0.22 / 20 h

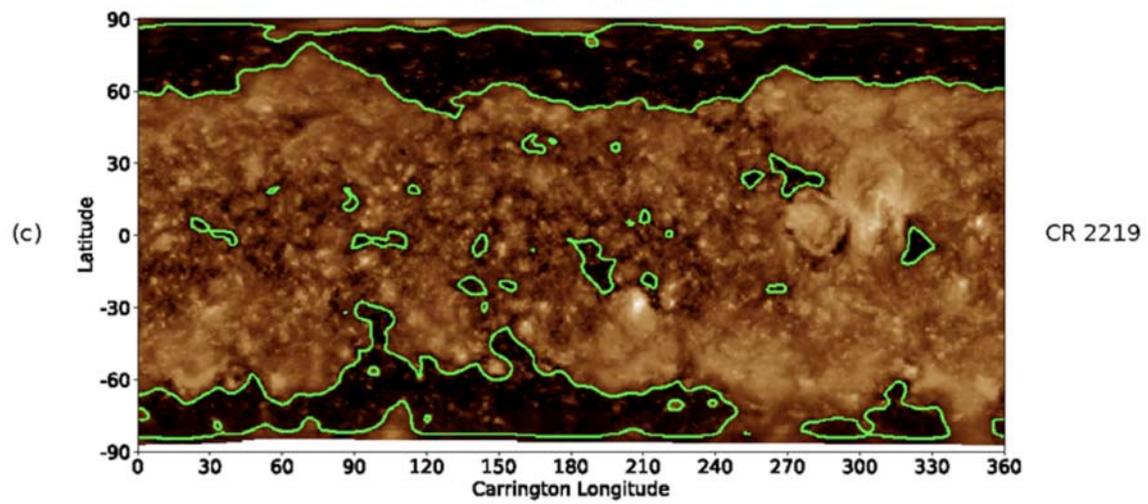
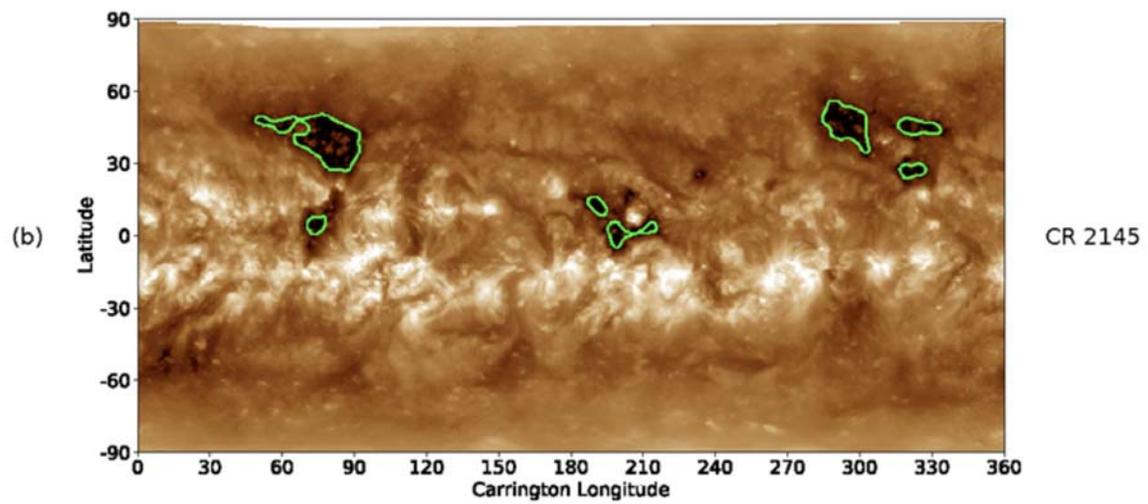
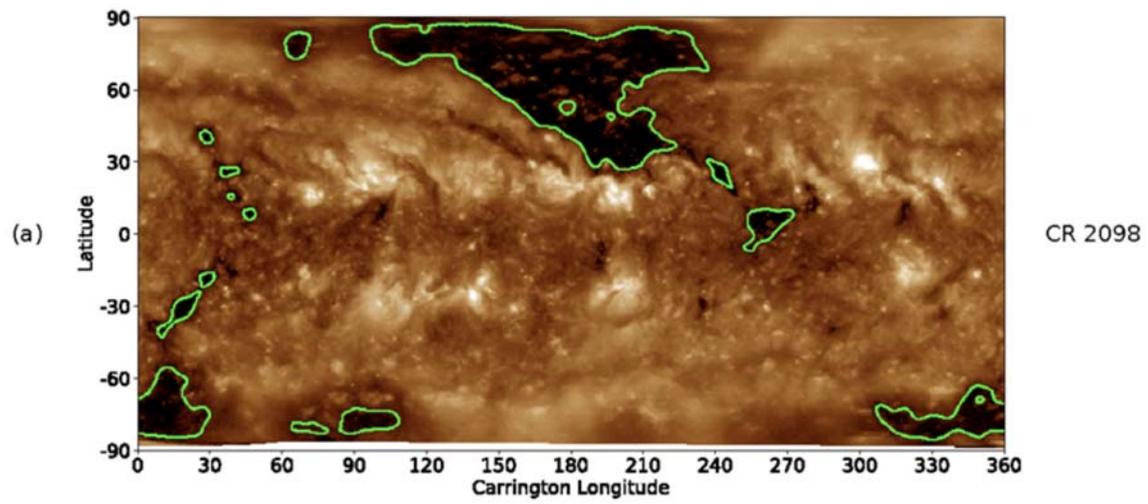
# IDENTIFICATION OF CORONAL HOLES USING CONVOLUTIONAL NEURAL NETWORK

Identification of solar Coronal Holes (CHs) provides information both for operational space weather forecasting and long-term investigation of solar activity. Due to different appearances of disk images and synoptic maps, the algorithms for CHs segmentation are typically elaborated independently. In contrast, we suggest the idea that the concept of CHs should be similar for both cases. This motivates us to investigate universal models that can learn a CHs segmentation in disk images and reproduce the same segmentation in synoptic maps. In our research, we demonstrate that Convolutional Neural Networks (CNN) can be considered as such universal models.

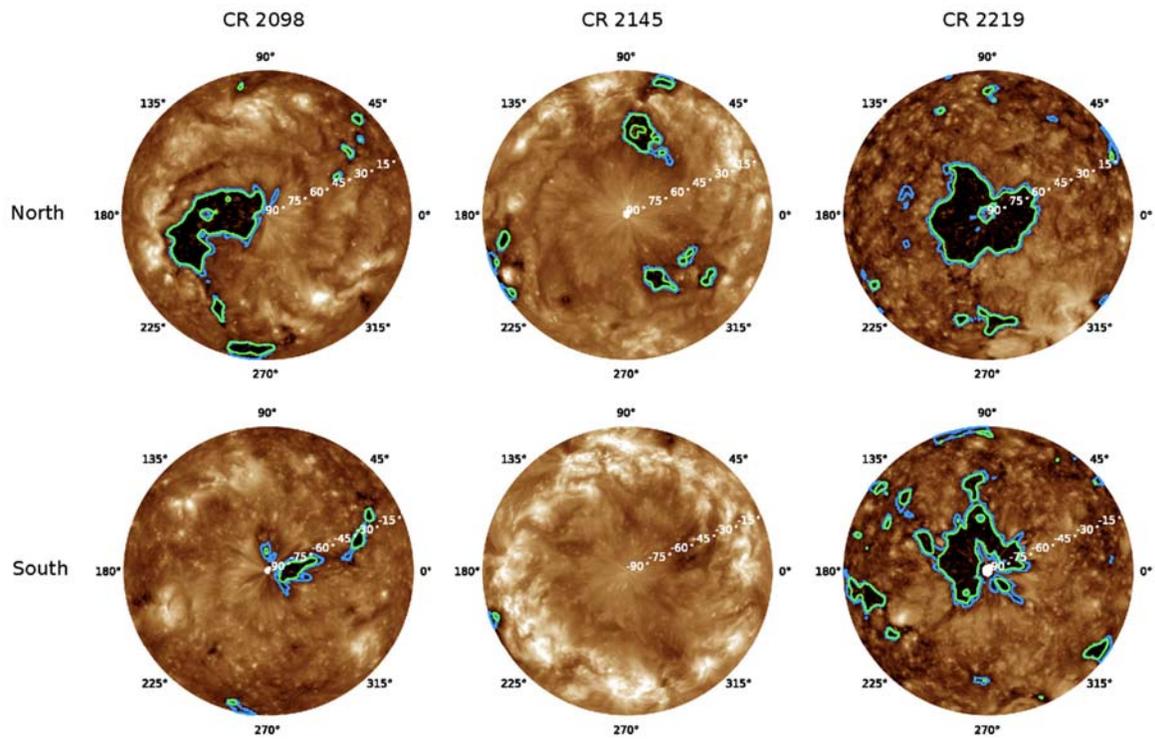


The U-Net architecture with compression and decompression branches and skip-connections. Input image (e.g. solar disk image or synoptic map) has spatial dimensions  $N \times M$  and  $C_{in}$  channels. Each convolutional-downsampling block compresses spatial dimensions and increases the number of channels. The decompression branch acts as inverse operation, output image (e.g. segmentation mask) has spatial dimensions  $N \times M$  and  $C_{out}$  channels.

We have developed a catalog of synoptic maps for the period of 2010–2020 based on SDO/AIA observations in the 193 Angstrom wavelength and demonstrate that the model trained on daily disk images provides accurate CHs segmentation in synoptic maps and its pole-centric projections. The catalog is available at [https://sun.njit.edu/coronal\\_holes/](https://sun.njit.edu/coronal_holes/). ([https://sun.njit.edu/coronal\\_holes/](https://sun.njit.edu/coronal_holes/))



Sample synoptic maps and predicted CHs boundaries overlaid for CRs 2098 (a), 2145 (b), and 2219 (c).



*CHs predicted for pole-centric input images (green lines) is comparison to pole-centric projections of CHs predicted in synoptic maps (blue lines).*

## CV

Biosketch. Alexander Kosovichev is a Professor of Physics in the New Jersey Institute of Technology and Director of Center of Computational Heliophysics. Dr. Kosovichev's research interests encompass a broad range of heliophysics, including both theory and data analysis. He worked as Associate Investigator for the NASA-ESA mission Solar and Heliospheric Observatory (SoHO) and Co-Investigator in two instrument teams, Helioseismic and Magnetic Imager (HMI) and Atmospheric Imaging Assembly (AIA) of the Solar Dynamics Observatory (SDO) mission. He served as Science Lead of the HMI project and developed the project science plan. He has been PI of many NASA-funded research grants, served in review panels, the LWS science architecture team, and organized several conferences and workshops on topics of NASA heliophysics missions, including SoHO and SDO Workshops.

# ABSTRACT

The project develops innovative tools to extract and analyze the available observational and modeling data in order to enable new physics-based and machine-learning approaches for understanding and predicting solar activity and its influence on the geospace and Earth systems. The heliophysics data are abundant: several terabytes of solar and space observations are obtained every day. Finding the relevant information from numerous spacecraft and ground-based data archives and using it is paramount, and currently a difficult task.

The scope of the project is to develop and evaluate data integration tools to meet common data access and discovery needs for two types of Heliophysics data: 1) long-term synoptic activity and variability, and 2) extreme geoeffective solar events caused by solar flares and eruptions. The methodology consists in the development of a data integration infrastructure and access methods capable of 1) automatic search and identification of image patterns and event data records produced by space and ground-based observatories, 2) automatic association of parallel multi-wavelength/multi-instrument database entries with unique patterns or event identifiers, 3) automatic retrieval of such data records and pipeline processing for the purpose of annotating each pattern or event according to a predefined set of physical parameters inferable from complementary data sources, and 4) generation of a pattern or catalog and associated user-friendly graphical interface tools that are capable to provide fast search, quick preview, and automatic data retrieval capabilities.

The Team has developed and implemented the Heliportal that provides a synergy of solar flare observations, taking advantage of big datasets from the ground- and space-based instruments, and allows the larger research community to significantly speed up investigations of flare events, perform a broad range of new statistical and case studies, and test and validate theoretical and computational models. The Heliportal accumulates, integrates and presents records of physical descriptors of solar flares, as well as the magnetic characteristic of active regions from various catalogs of observational data from different observatories and heliophysics missions.