Building an Extensible Data Ecosystem for Solar Magnetic Polarity Inversion Lines

Nick Murphy
Georgia State University
Atlanta, Georgia, USA
nmurphy11@student.gsu.edu

Ziba Khani
Georgia State University
Atlanta, Georgia, USA
zkhani1@student.gsu.edu

Berkay Aydin
Georgia State University
Atlanta, Georgia, USA
byadin2@gsu.edu

Abstract

Forecasting the relevant characteristics of central space weather events such as coronal mass ejections and solar flares is of utmost interest due to their potential near-Earth impact on our technological infrastructure. A key solar feature that is successfully utilized for predicting these solar events is magnetic polarity inversion lines (MPILs). Derived from magnetic field rasters, MPILs represent the shear layers between two opposing (i.e., positive and negative) polarity regions, and the complexity of these separating lines is shown to be highly relevant precursors for these solar events. In this paper, we present our data ecosystem for detecting and serving metadata for MPILs, along with important spatial, temporal and spatial-temporal search capabilities. The MPILs are detected using our detection framework and are served through our public APIs. Our APIs provide active region-based, spatial, and temporal search capabilities. The MPIL metadata consists of a series of binary rasters and metadata parameters. The rasters show the polarity inversion lines, regions of polarity inversion, the unsigned negative and positive polarity regions (both positive and negative) and convex hull of polarity inversion lines, while metadata features include physical and shape-based image parameters. We organize the metadata series as spatial and temporal time series derived from active region trajectories. This data resource is heterogeneous in nature and is designed to be easily extensible. MPIL-derived features are currently used in various deployed operational systems. We envision that our near-real time detection module and API service will be used as the backbone for operational space weather forecasting tools.

CCS Concepts

• **Applied computing** → *Astronomy*; • **Information systems** → **Spatial-temporal systems**; *Information integration.*

Keywords

data integration, object detection, solar astroinformatics

ACM Reference Format:
Nick Murphy, Ziba Khani, and Berkay Aydin. 2024. Building an Extensible Data Ecosystem for Solar Magnetic Polarity Inversion Lines. In *Proceedings of The 32nd ACM International Conference on Advances in Geographic Information Systems Proceedings (SIGSPATIAL '24).* ACM, New York, NY, USA, 4 pages. https://doi.org/3678717.3691210

SIGSPATIAL '24, October 29 - November 1, 2024, Atlanta, GA

ACM ISBN 979-8-4007-1107-7/24/10
https://doi.org/3678717.3691210

1 Introduction

Extreme space weather events caused by interaction of the Earth with emissions from the Sun can potentially impact our technological infrastructure, both on the Earth's surface and its surrounding orbit. Robust prediction of these space weather events is vital to many stakeholders in various branches of governments and across many sectors of commercial enterprises. The shape and complexity of magnetic polarity inversion lines (MPIL) have long been considered a strong precursor of these central space weather events such as solar flares and coronal mass ejections. MPILs are essentially shear layers that separate sufficiently intense positive and negative magnetic polarity regions. The complexity of MPILs and related features in a particular solar region is an indicator of energy build-ups among opposing magnetic regions. The recent findings from simulation-based and observational studies show that topological features of MPILs from a solar active region are strongly associated with the flare and eruption productivity [6][11]; therefore, providing large-scale MPIL data and metadata that are integrated with existing resources is of great importance both for enabling data-driven heliophysics research and for operational space weather forecasting. That said, current MPIL-related datasets are practically scattered, come from non-standardized (and often publicly unavailable) detection modules and various raw magnetogram data, and are created for task-specific instances [7]. Given the limited accessibility, we believe it is important to provide a data service capable of generating integrated, discoverable, and accessible MPIL data with proper information on data lineage, parametric configurations and relevant metadata.

The few publicly available MPIL datasets, due to the fact that they are used in analytical studies, are stored in flat files and raster format under standard file system directories. However, the nature of the MPIL data is spatial and temporal, where these separating lines continuously change their shape-based characteristics over time. Operational users (such as space weather forecasting researchers) need to locate and identify spatio-temporal characteristics within the MPIL data. To do this, spatial and temporal query support as well as timely generation and access to the data are necessary. In this paper, we present an end-to-end data service to provide this support, consisting of two main features: (1) A data generation framework that downloads and processes magnetogram data to create properly annotated MPIL data and (2) A public-facing application programming interface (API) for querying and retrieving information from the generated near-real time and historical datasets. The data generation framework is built upon an extended version of MPIL generation algorithm presented in [5]. The historical and near-real time datasets are generated mainly using HMI Active Region Patch (HARP) data series [3], which are produced by the Solar Dynamics Observatory's Helioseismic and Magnetic Imager (SDO/HMI) instrument. HARP data series, available in definitive and near-real-time versions, are essentially spatio-temporal trajectories of solar active regions that show the evolving physical characteristics of various magnetic field observations. Using our generation framework, we create MPIL-related metadata for solar active regions and integrate them to widely used HARP identifiers. The created data is then served using the public API. Users can pass URL-based parameters to request the MPIL rasters and metadata by active region identifiers or within specific temporal or spatio-temporal windows. This interface increases the accessibility of the dataset to researchers and other third parties. The web service is currently available at https://dmlab.cs.gsu.edu/mpil-api/ and is under active development.

The rest of the paper is organized in the following manner. Section 2 discusses relevant research and applications related to this work. Sections 3 and 4 describe the architecture of the service and its utility to provide temporal, spatio-temporal, and active region-based information to clients. Section 5 presents conclusions and introduces the next steps for future development using this service.

2 Related Work

Various methods have been proposed recently for detecting MPILs from solar magnetograms in order to provide insights into their properties and their potential application in predicting solar flares. Some methods focus on pixel-level analysis, while others employ segmentation and optimization procedures. For instance, Toriumi and Takasao[11] used the horizontal gradient of the magnetic field to detect and analyze the properties of MPILs in numerical simulations of solar active regions. Vasantharaju et al.[12] studied the magnetic properties of solar flares by analyzing PIL observations and investigating their relationship with flare productivity. Sharykin et al.[10] examined the energy release in the lower solar atmosphere near the PIL during a flare event. Kim et al.[8] proposed a new mechanism for solar eruptions based on preflare eruptions occurring before major solar flares.

To leverage the properties of PILs for solar flare prediction, machine learning models have been applied. Wang et al.[13] utilized features of MPILs extracted from HMI vector magnetograms and employed a random forest model for classifying solar active regions into flaring and non-flaring categories. Sadykov and Kosovichev[9] derived multiple MPIL-related physical parameters from line-of-sight magnetograms and employed a Support Vector Machine (SVM) for flare forecasting. These studies demonstrate the potential of MPIL detection and analysis in enhancing our understanding of solar activity and its effects on the near-Earth environment.

There are existing protocols for retrieving Earth-based geospatial data such as Web Map Services [2]. However, most existing PIL detection tools and datasets are not stored in formats that can be easily integrated with these services. The data is not readily available for widespread use in data-intensive forecasting studies or operational modules, limiting their accessibility and hindering progress in this field and operational space weather forecasting. To address this gap, we present an extensible ecosystem that will serve as critical solar metadata cyberinfrastructure for operational space weather forecasting and support heliophysics research.

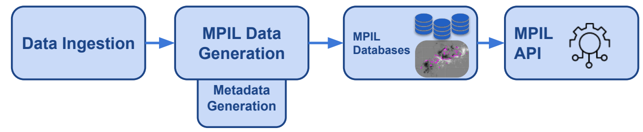

Figure 1: A general overview of MPIL data generation and public API

3 Structure

A general overview of our MPIL data service is illustrated in Fig. 1. The input magnetogram data series are downloaded and stored as flat files. Then the magnetogram files are fed to our MPIL generation framework to create MPIL rasters with physical, spatial, and shape-based features. These features eventually form multivariate time series data for each active region trajectory. In addition we generate various identifying metadata information. The resulting data products are stored in a spatially extended database and then are served via MPIL API module. We present the structure of the MPIL data products and related database and public API in Fig. 2.

The MPIL API serves as a public interface for clients to access the MPIL datasets. The front-end was developed using the Node.js framework. Client requests are sent via RESTful queries for spatial, temporal, and region-based information. Relevant metadata is then retrieved from a PostgreSQL database with PostGIS extension and returned to the requester in JSON format. Rasters are stored outside the database and accessed from links provided in the returned metadata. The service exists inside a Docker container for portability and ease of deployment.

Our service currently provides access to MPILs from both definitive and near-real time HARP series and covers active region trajectories starting from May 2010. The MPIL data is created with the extended version of our MPIL detection framework [1]. The flat files and rasters generated from the MPIL detection framework are further processed with custom scripts to gather data lineage and configuration metadata. The data lineage metadata includes

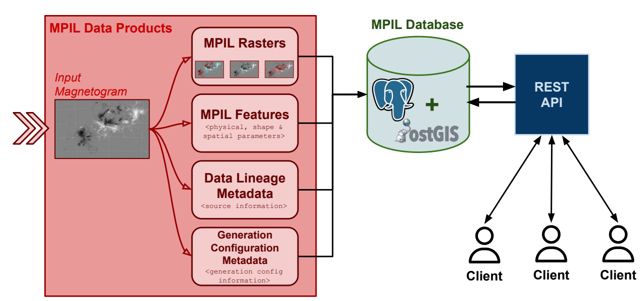

Figure 2: Structure of the MPIL data products and the architecture of its accompanying public API

information about the HARP data series (e.g., data series identifiers, projections for magnetogram rasters, type of the longitudinal magnetogram, etc.). The configuration metadata consists of information about the code version and repository links used to generate MPILs as well as the parametric settings used (such as the magnetic field strength threshold or size filters). Additional information about widely used solar active region identifiers (HARP identifiers – HARPNUM for both definitive and near-real time series and NOAA active region numbers) are also integrated to provide access. We use spatio-temporal co-occurrence-based integration to identify corresponding active region identifiers[4]. The MPIL features and metadata values are then ingested into a MPIL database instance, which is currently served in PostgreSQL. The MPIL raster files are copied to a separate public repository and the relative URLs for the rasters are also stored in the database.

4 Features

MPIL Rasters and Features: We briefly described our generation procedure in Section 3. In addition to the original magnetogram cutout raster, each active region patch is associated with six binary masks showing positive and negative polarity regions, unsigned polarity regions, the region of polarity inversion (RoPI), a thinned MPIL, and the convex hull of the MPIL. These masks were used during the MPIL detection framework to generate a rich collection of metadata for each active region. Some of the MPIL features generated include coordinate-based shape of the active region patch, the convexity, eigenvalues, fractal dimensions, and magnetic field strength covered by regions of polarity inversion, as well as additional details about the binary masks such as their size and number of connected line components. These values are all recorded with a timestamp for use in active region trajectory analysis. We provide an example set of raster data products in Fig. 3.

Query Capabilities: The information stored in MPIL database can be interacted with via the front-end interface through three main types of queries: active-region-based, temporal, and spatio-temporal. Active region-based queries assume that the user already knows the identity of a particular area and wish to retrieve all relevant data regarding the active region. A benefit provided by the API is that multiple well known active region identifiers can be correlated to the same raster and metadata resources. For example, if a HARP number and a NOAA Active Region (NOAA AR) identified region are related, requesting data from either format will contain the same results.

The temporal and spatio-temporal queries can be used when the time frame and location are the relevant search parameters. For temporal window queries, clients provide the start and end time in the standard YYYY-MM-DD-HH:MM:SS format. Spatio-temporal queries also include the bounding box coordinates defined by the lower-left and upper-right latitude and longitude values in Heliographic coordinates. Both queries allow clients to take advantage of convenience features provided by the API. The geometric aspect of active region data is stored as spatial objects using the PostGIS extension, and the spatial and spatio-temporal queries from the API are performed using spatial overlap operations to determine which active regions intersect with the bounding box. Combined with the date-range matching provided by traditional relational database structure, clients can easily obtain active region trajectory information without having to build additional custom infrastructure.

All queries are passed to an active Node.js instance as HTTP GET requests. The query type and parameters undergo initial verification to ensure they are in the correct format. If correct, the data is sent to the appropriate stored procedure in the Postgres database and the query is performed. Upon receiving the results, the Node.js instance converts the SQL data into JSON format and returns the information to the client.

For error handling and notification purposes, query responses begin with request status and current version of the API. All successful queries include configuration metadata about how the raster data was originally captured and processed to provide for outside reproduction of results. Then the records are returned as a series of objects grouped by an internal active trajectory region identifier (in addition to any other relevant active region identifier). The overall time range and bounding box for the series are listed first. Following that, an array of raster data records is sent. Public URLs for all binary mask rasters are presented along with the raster type and timestamp. Finally, the MPIL features described in Section 4 are listed in one JSON string.

5 Conclusion

In this paper we have demonstrated our MPIL data service, which consists of data generation, processing, storage, and serving tools. The data service provides access to a rich metadata source for use in heliophysics and space weather forecasting. The integrated dataset serves as a critical infrastructure that can be used by various forecasting methods. The public API demonstrated in this work provides an effective access mechanism for researchers to query and analyze these datasets. Given that various aspects of MPILs are used in a wide range of predictive analytics studies, we envision that our API will further facilitate the development of machine learning-based prediction tools for space weather forecasting. In the future, we plan to include and integrate other historical magnetogram data sources such as the cutouts from SOHO/MDI (the predecessor of SDO/HMI). We also plan to implement visual support and analytical search functions for metadata features.

6 Acknowledgement

This work is supported in part by a grant from NSF (Award #2104004).

References

[1] Berkay Aydin, Anli Ji, Nigar Khasayeva, Rafal A. Angryk, Petrus Martens, and Manolis K. Georgoulis. 2022. Exploratory Analysis of Magnetic Polarity Inversion Line Metadata and Eruptive Characteristics of Solar Active Regions. In *44th COSPAR Scientific Assembly. Held 16-24 July*, Vol. 44. 3223.

[2] Veenendaal Bert. 2015. Developing a map use model for web mapping and GIS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-4/W7. https://doi.org/10.5194/isprsarchives-XL-4-W7-31-2015

[3] M. G. Bobra, X. Sun, J. T. Hoeksema, M. Turmon, Y. Liu, K. Hayashi, G. Barnes, and K. D. Leka. 2014. The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs – Space-Weather HMI Active Region Patches. *Solar Physics* 289, 9 (April 2014), 3549–3578. https://doi.org/10.1007/s11207-014-0529-3

[4] Xumin Cai, Berkay Aydin, Manolis K. Georgoulis, and Rafal Angryk. 2019. An Application of Spatio-temporal Co-occurrence Analyses for Integrating Solar Active Region Data from Multiple Reporting Modules. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. https://doi.org/10.1109/bigdata47090.2019.9006185

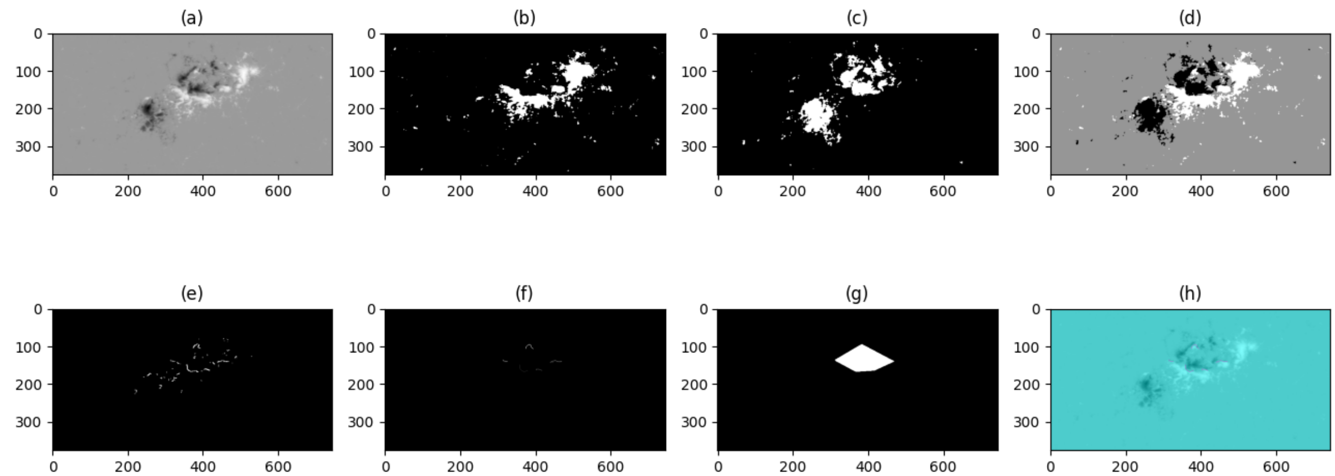

Figure 3: The example raster data products generated from PIL detection tool from an active region patch (shown in (a) HARP 377 (NOAA AR 11158) at 2011-02-13 16:00:00 UT): (b) Positive polarity region, (c) Negative polarity region, (d) Union of positive and negative polarity regions (colored for illustration – white areas for positive and black for negative regions), (e) regions of polarity inversion, (f) refined polarity inversion lines (PILs) from the regions of polarity inversion, (g) convex hull of PILs, (h) PILs overlaid on original magnetogram raster.

[5] Xumin Cai, Berkay Aydin, Anli Ji, Manolis K. Georgoulis, and Rafal Angryk. 2020. A Framework for Detecting Polarity Inversion Lines from Longitudinal Magnetograms. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE. https://doi.org/10.1109/bigdata50022.2020.9377808

[6] Domenico Cicogna, Francesco Berrilli, Daniele Calchetti, Dario Del Moro, Luca Giovannelli, Federico Benvenuto, Cristina Campi, Sabrina Guastavino, and Michele Piana. 2021. Flare-forecasting Algorithms Based on High-gradient Polarity Inversion Lines in Active Regions. *The Astrophysical Journal* 915, 1 (July 2021), 38. https://doi.org/10.3847/1538-4357/abfafb

[7] Anli Ji, Xumin Cai, Nigar Khasayeva, Manolis K. Georgoulis, Petrus C. Martens, Rafal A. Angryk, and Berkay Aydin. 2023. A Systematic Magnetic Polarity Inversion Line Data Set from SDO/HMI Magnetograms. *The Astrophysical Journal Supplement Series* 265, 1 (March 2023), 28. https://doi.org/10.3847/1538-4365/acb43a

[8] Sujin Kim, Y.-J. moon, Y.-H. Kim, Y.-D. Park, K.-S. Kim, Gwangson Choe, and Khan-Hyuk Kim. 2008. Preflare Eruption Triggered by a Tether-cutting Process. *The Astrophysical Journal* 683 (2008), 510 – 515.

[9] Viacheslav M. Sadykov and Alexander G. Kosovichev. 2017. Relationships between Characteristics of the Line-of-sight Magnetic Field and Solar Flare Forecasts. *The Astrophysical Journal* 849, 2 (Nov. 2017), 148. https://doi.org/10.3847/1538-4357/aa9119

[10] I. N. Sharykin, V. M. Sadykov, A. G. Kosovichev, S. Vargas-Dominguez, and I. V. Zimovets. 2017. Flare Energy Release in the Lower Solar Atmosphere near the Magnetic Field Polarity Inversion Line. *The Astrophysical Journal* 840, 2 (may 2017), 84. https://doi.org/10.3847/1538-4357/aa6dfd

[11] Shin Toriumi and Shinsuke Takasao. 2017. Numerical Simulations of Flare-productive Active Regions: -sunspots, Sheared Polarity Inversion Lines, Energy Storage, and Predictions. *The Astrophysical Journal* 850, 1 (nov 2017), 39. https://doi.org/10.3847/1538-4357/aa95c2

[12] N. Vasantharaju, P. Vemareddy, B. Ravindra, and V. H. Doddamani. 2018. Statistical Study of Magnetic Nonpotential Measures in Confined and Eruptive Flares. *The Astrophysical Journal* 860, 1 (jun 2018), 58. https://doi.org/10.3847/1538-4357/aac272

[13] Shuhui Wang, King-Fai Li, Diana Zhu, Stanley P. Sander, Yuk L. Yung, Andrea Pazmino, and Richard Querel. 2020. Solar 11-Year Cycle Signal in Stratospheric Nitrogen Dioxide—Similarities and Discrepancies Between Model and NDACC Observations. *Solar Physics* 295, 9 (Sept. 2020). https://doi.org/10.1007/s11207-020-01685-1