# An Application of Spatio-temporal Co-occurrence Analyses for Integrating Solar Active Region Data from Multiple Reporting Modules

Xumin Cai*, Berkay Aydin†, Manolis K. Georgoulis‡, Rafal Angryk§

Department of Computer Science, Georgia State University*†§
Department of Physics & Astronomy, Georgia State University‡
Email: xcai3@student.gsu.edu*, baydin2@cs.gsu.edu†, manolis.georgoulis@phy-astr.gsu.edu‡, rangryk@cs.gsu.edu§

***Abstract*—Spatio-temporal co-occurrence analysis captures the spatial and temporal relations between events that occur at the same time and location. In this paper, we utilize spatio-temporal co-occurrence relations to integrate solar active region (AR) data detected and reported by three feature recognition methods, namely, human labeling by forecasters at National Oceanic and Atmospheric Administration (NOAA), Spaceweather HMI Active Region Patch (SHARP) detection pipeline, and Spatial Possibilistic Clustering Algorithm (SPoCA). We determine the associations between individual reports by identifying the spatio-temporal co-occurrences among the reports from these modules. We compare our findings with the data from the Joint Science Operations Center (JSOC), analyzing the discrepancies in different circumstances. We found 105 SHARP series not properly associated with the NOAA-labeled ARs. In the end, we provide detailed movement analyses for the AR trajectories, create an updated SHARP-to-NOAA AR associations, that is crucial for space weather predictions utilizing magnetic field information, and make the ternary associations between SHARP, NOAA ARs, and SPoCA ARs available to the public.**

## I. Introduction

Solar active regions are areas of intense magnetic flux in the Sun. These regions frequently generate various types of explosive solar activity, including eruptions such as solar flares and coronal mass ejections. They are the main source of information for predicting these potentially catastrophic solar events. Various reporting modules detect active regions (ARs) using their own methods. This work is mainly concerned with integrating and further cleaning of heterogeneous active region data using their reported locations and occurrence times.

Our goal, in this work, is to associate the active regions' spatio-temporal vector data reported from NOAA, SHARP, and SPoCA and integrate these resources for further space weather analytics applications. For each active region instance from these reporting modules, we unify their identification and provide a detailed view of the data.

While identifying the same phenomena, active region reports are fairly heterogeneous, primarily due to the aim of each method and the data source used for detection. For instance, NOAA ARs are detected from relatively low-resolution continuum images of the solar photosphere, used to identify sunspots and then magnetograms to track helpful magnetic characteristics of sunspots that are useful for forecasting. SHARPs, on the other hand, are defined from high-resolution vector magnetograms and are used to understand the magnetic field characteristics of active regions. There are also a variety of different metadata attributes each module detects or calculates. Attributes common in all data sources, at least from a semantic point of view, are spatial and temporal attributes. However, the spatial data types of each module are different and the temporal cadence of the reports also varies significantly. We believe that integrating these data resources and cleaning them can provide comprehensive reference for future research and enable more robust space weather analytics. This is because these AR data resources are practically indispensable and are utilized, often independently, in flare prediction and characterization studies [1]–[4].

To integrate, we used the spatial and temporal attributes of active region reports and discovered the spatio-temporal co-occurrence relationships [5] to measure the correlation among different AR instances. In this work, we used time intersection as a basis for temporal co-existence and topological and distance-based spatial relationships [6] as the approach to associate co-occurring AR instances. To identify significance of these spatio-temporal co-occurrences, we employed two measures of significance, which can be considered as trajectory similarity measures of trajectories with mixed-spatial extents.

Active region data from all three resources are tracked, namely, individual event reports are organized as spatio-temporal trajectories using a unique identifier for each trajectory specific to the reporting module. SHARPs use HARPNUMs (HMI Active Region Patch Number), NOAA uses the NOAA Active Region numbers (NOAA AR#), and SPoCA uses the SPoCA AR ID [7]–[9]. SHARP records also provide one or more NOAA AR# for each trajectory, if available. Note that some SHARPs do not have an associated NOAA AR#. Furthermore, SHARPs present a detailed spatio-temporal view for each segment (i.e., individual event) in high temporal and spatial resolution, providing a bounding box and, for strong magnetic fields, a region representated as bitmaps (we use bounding boxes in this study). Here, we first clean the NOAA AR reports via an outlier analysis, update the SHARP-to-NOAA AR associations, and extend the co-occurrence analysis to discover SHARP-to-SPoCA AR associations and provide a clean ternary relation among the AR reports.

The rest of this paper is organized as follows. In Section II, we present the background information on the solar active region reporting modules: NOAA, SHARP, and SPoCA. Section

978-1-7281-0858-2/19/$31.00 © 2019 IEEE

III describes our spatio-temporal co-occurrence integration methodology. In Section IV, we provide a comparison between our SHARP-to-NOAA AR pairs and the association information by the Joint Science Operations Center (JSOC) and also individually analyze the discrepancies. We also detect NOAA AR outliers and update event records to provide improved association results. We briefly describe our conclusions and envisioned future work in Section V.

## II. Background

### A. NOAA Active Regions

The forecasters from Space Weather Prediction Center (SWPC) of NOAA [7] analyze and detect active regions associated with sunspot groups in the Sun's photosphere each day. They use ground-based observations to detect visible sunspot groups and evaluate the size and complexity based on the McIntosh classification, namely a modified Zurich classification scale and Mount Wilson magnetic classification system [10]. The SWPC assigns an identifier, called the NOAA AR number (NOAA AR#), for each AR when they are first detected on the earthward solar disk.

The SWPC also generates a report named the Solar Region Summary (SRS) [7] that includes active regions detected during the previous day. The report records the information including the NOAA number, central location of the ARs and a few other attributes such as number of sunspots and McIntosh classification.

### B. Spaceweather HMI Active Region Patches (SHARPs)

The SHARP detection pipeline [8] processes the data coming from the Helioseismic and Magnetic Imager (HMI) onboard the Solar Dynamics Observatory (SDO) and generate HMI Active Region Patches (HARPs), which show the spatial extent of detected active regions. HARPs are further processed to create SHARPs, which also include various space weather-related quantities calculated from photospheric vector magnetograms. SHARP series contain the bounding boxes of tracked AR patches with a 12-minute cadence.

The module also assigns an identification number for each HMI Active Region Patch (HARP), encoded as HARPNUM. The detection identifies AR patch on the CCD image represented by a bounding box and create a bitmap inside the box. The bitmap which identifies the strong-field pixels is within this bounding box. These bounding boxes are large enough to include the maximum heliographic extent magnetic structures, often neighboring active regions. Each HARP data series provides a piece of consecutive geometric information about the AR patch every 12 minutes.

### C. Spatial Possibilistic Clustering Algorithm (SPoCA)

Spatial Possibilistic Clustering Algorithm (SPoCA) [9] is essentially a multi-channel, salient event detection framework utilizing an unsupervised clustering method. The SPoCA applies image segmentation methods on solar extreme ultraviolet (EUV) images at Sun's chromosphere. These images are obtained from the Atmospheric Imaging Assembly (AIA) onboard SDO. SPoCA provides a number of attributes for each detected AR event such as the center location or area. The SPoCA detected ARs are reported to the Heliophysics Event Knowledge Base (HEK) generally every four hours. Each report contains the centroid location, bounding box, and chain code representation of identified ARs.

## III. Integration Methodology

In this section, we describe the primary methodology for integrating the heterogeneous AR data from NOAA, SHARP, and SPoCA. To achieve this, we treat the timestamps and time intervals of ARs as temporal attributes, whereas we treat centroids, bounding boxes, and chain codes of ARs as spatial attributes. We, in principle, consider that if reports from two modules co-occur in space and time, then they represent the same phenomena. To implement the co-occurrence relationship, we use time interval intersection as the temporal predicate to find the coexistence relationships. Moreover, we further refine the coexisting instances using spatial intersect predicate as the primary co-location rule to establish the co-occurrence relationship between AR reports.

### A. Temporal Coexistence

Basic temporal relationships [11] between the two time-points are straightforward. We could describe them as follows: A time-point can occur (1) before, (2) after, or (3) at the same time with another time point.

A pair of ordered time-points constitute a time-interval, which represents the continuous time range between these two time-points. In 1983, Allen introduced a set of relationships that can occur between the two time-intervals [12], which was widely adopted as the methodology for discovering temporal patterns. There is a total of thirteen relationships between two time-intervals formalized in Allen's interval algebra, including seven primary relationships [12]: *before, meets, overlaps, during, starts, finishes, equals*; and six inverse relationships: *after, met by, overlapped by, contains, starts by, finished by*. Nine of these relations essentially represent the coexistence relationship, which, for two time intervals T1 and T2, can be formalized as follows:

$$\textit{Coexists(T1,T2)} = \begin{cases} \textit{True} & \textbf{if } T1.end > T2.start \textbf{ AND} \\ & \quad T1.start < T2.end \\ \textit{False} & \textbf{otherwise} \end{cases} \quad (1)$$

where $T_i.start$ and $T_i.end$ represents the start and end times of the intervals.

Each reporting module provides the temporal attributes for the AR reports; however, with great differences in terms of cadence. SPoCA reports the AR locations approximately every 4 hours. SHARP detection module reports the locations with a cadence of 12 minutes, while NOAA ARs are reported daily. To account for these differences, for each daily NOAA AR event report, we applied a temporal buffer of $\pm 12$ hours. By doing this, we extrapolated the approximate locations of

**(a)**

**(b)**

**Fig. 1:** Extrapolation process of NOAA AR 11939. Shown in (a), we apply a temporal buffer of $\pm 12$ hours to daily reports of NOAA AR 11939. (b) shows the extrapolation of individual event reports using the time delays and longitudinal displacement towards east and west, matching the timestamps of SHARPs (i.e., every 12 minutes).

NOAA AR centroids 12 hours backwards and forwards, with 12-minute cadence. For the SPoCA AR, we extrapolated the approximate locations of bounding box between the SPoCA AR start time and end time, with 12-minute cadence.

This extrapolation process makes use of known differential rotation of the Sun [13]. We use the angular velocity of the sun ($\omega$), shown in Eq. 2, provided in [13] to estimate the approximated displacement of NOAA AR and SPoCA at HARP segment time point.

$$\omega = A + B\sin^2(\varphi) + C\sin^4(\varphi) \tag{2}$$

where $\varphi$ is the latitude, A, B, and C are approximated constants (A=14.11, B=-1.7, and C=-2.35) [14]. Then, given $\Delta T$ – the time difference (in days) between the observation and the AR location to be extrapolated, the longitudinal displacement ($\Delta\lambda$ is calculated as:

$$\Delta\lambda = \omega \times \Delta T \tag{3}$$

We, first, calculate the time difference $\Delta T$ between the original AR observation's timestamp ($t_{Obs}$) and the template HARP segment's time point ($t_{Ex}$). Then, using $\Delta T$, we apply the solar differential rotation formula presented in Eq. 2 and calculate the longitudinal displacement $\Delta\lambda$ between $t_{Obs}$ and $t_{Ex}$. Finally, we add $\Delta\lambda$ to our observation's longitude and determine the approximate AR location at $t_{Ex}$. Note that, we do not alter the observation's latitude when extrapolating, because, as shown in our outlier analysis (in Fig. 4.b), expected daily latitudinal displacement for ARs is 0 degrees.

In Fig. 1, we demonstrate an example extrapolation of NOAA AR 11939 between 2013-12-30 and 2013-12-31 using 12-hour backward and forward temporal buffers. Note here, we extrapolate for every 12-minute interval to match the cadence of SHARP records.

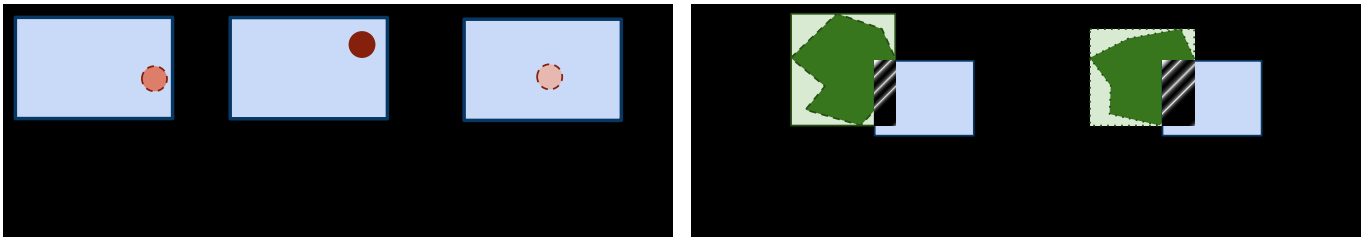
Following the location extrapolation, our first step in identifying the spatio-temporal co-occurrences is finding co-existing AR reports from different reporting modules. We use SHARP records (i.e., individual event locations at a particular time) as the template, and search the list of 12-hour extrapolated NOAA AR list as well as the SPoCA AR list for coexistences. The list of candidate NOAA and SPoCA ARs for each SHARP are temporarily recorded. Based on these two lists, we then search for spatial co-locations, which we will describe in the next subsection.

### *B. Spatial Co-locations*

Spatial relationships characterize how a given spatial object is located in space with respect to another spatial object. There are various types of spatial relationships that may appear among spatial objects such as topological, directional/orientation-based, shape-based, or distance-based relations [15]. A spatial co-location is an abstract spatial


(a) (b)

**Fig. 2:** Topological spatial relationships between AR reports with various spatial extents. (a) NOAA AR centroid (point) is within SHARP bounding box, (b) SPoCA AR bounding box overlaps with SHARP bounding box.



(a) (b)

**Fig. 3:** Spatio-temporal co-occurence relationship between the trajectories of (a) SHARP and NOAA AR using spatial within relation as co-location predicate and (b)SHARP and SPoCA AR using spatial intersect as the co-location predicate.

relation which represents the association among two or more spatial objects that are close-by or at the same location. While there are a plethora of spatial co-location mining literature [16], in our work, we mainly focus on the topological and distance-based spatial relationships among the AR event reports.

The distance-based spatial relationship between two spatial objects is based on a well-defined metric, defined over the space. The relationship specifies how far is the object from a given reference object. The Manhattan or Euclidean distances [17], for instance, between two spatial objects, are two widely used metrics to identify the spatial relationship. Topological spatial relationships are qualitative properties that characterize the relative positions of spatial objects toward each other and they are preserved under certain continuous transformations including all affine transformations [18]. The space can be stretched, compressed, twisted, bent, and transformed into different forms with topological relations between two objects remaining unchanged, which is essential for spatial objects with highly varying characteristics in terms of size, shape, and data types. The topological relationships are well formalized by the standard Dimensionally Extended nine-Intersection Model (DE-9IM) [19].

Due to different spatial data types of AR records, we employ both topological and distance-based relationships. To discover the colocations between coexisting NOAA ARs (reported as points) and SHARPs (reported as bounding box), we check (1) if NOAA AR centroid is within the bounding box of SHARP, namely spatial intersection, and (2) the minimum distance between NOAA AR and SHARP bounding box (which is 0 if *within*). Note here that SHARPs also provide a region representation for strong magnetic fields; however, we use the bounding box. We use the spatial intersection as the primary form of colocation and distances are calculated to quantify the discrepancies in tracked spatio-temporal trajectories. Figure 2a shows an example illustration for the spatial topological relationship we use between NOAA AR centroid and SHARP bounding box. For the SPoCA ARs, we use bounding box which is derived from the polygon, then check whether the bounding box of SPoCA AR overlaps with the bounding box of SHARP. Figure 2b shows an example of this relationship between a SPoCA AR and a SHARP AR.

### C. Spatio-temporal Co-occurrences

Our goal is to essentially discover the associations between the AR reports by discovering the spatio-temporal co-occurrences among them. Using the temporally coexisting SHARP-to-NOAA and SHARP-to-SPoCA associations as a template, we further refine these pairs based on whether they are co-located in space as well. Fig. 3 shows the spatio-temporal co-occurences between SHARP-to-NOAA AR in (a) and SHARP-to-SPoCA AR in (b). To determine the spatio-temporal co-occurrences, we used the spatial intersection as the primary spatial relationship to identify the co-occurrences. We also identified the distances between the AR reports to better quantify and highlight potential problems in the JSOC provided SHARP-to-NOAA AR associations.

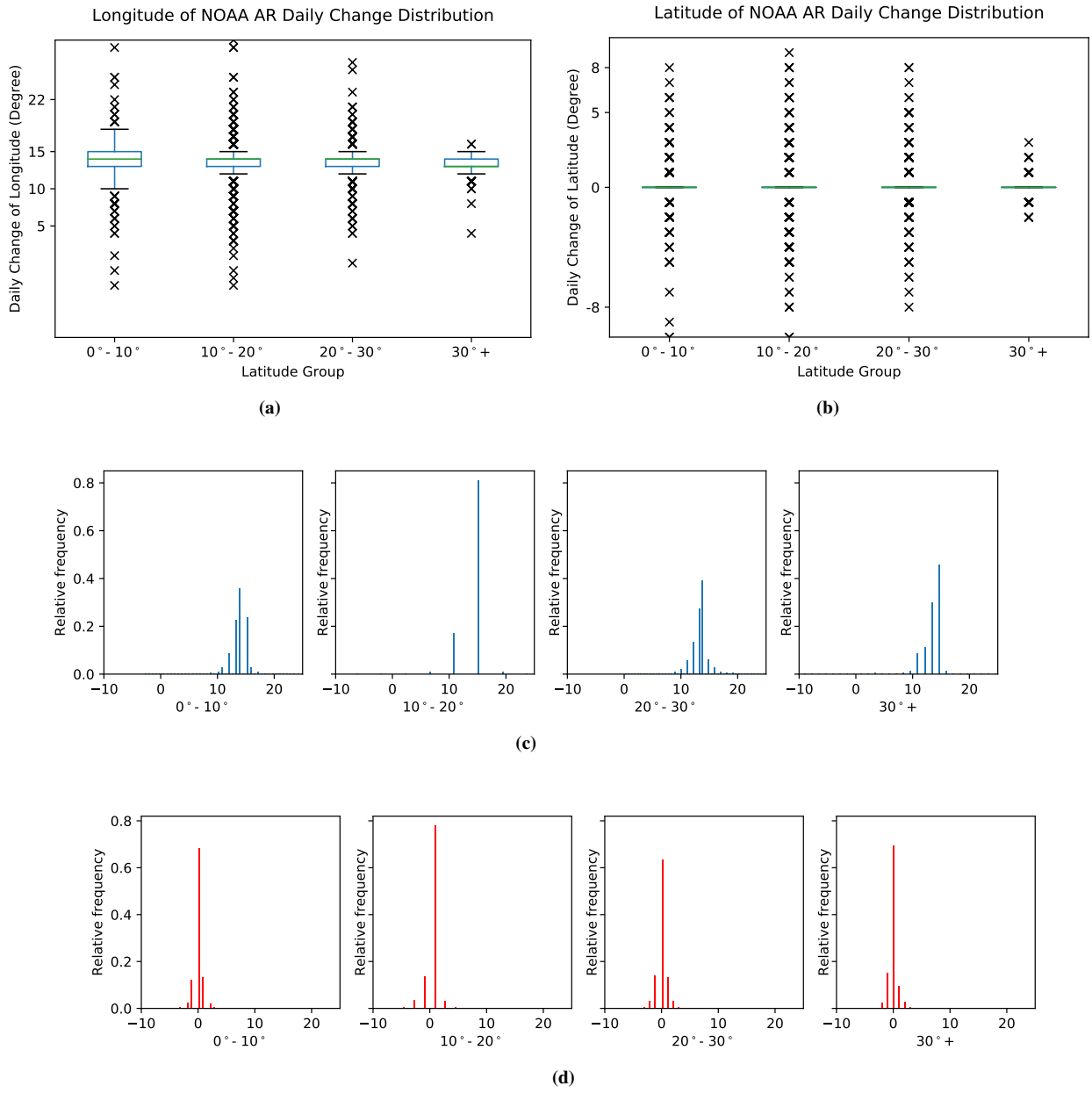
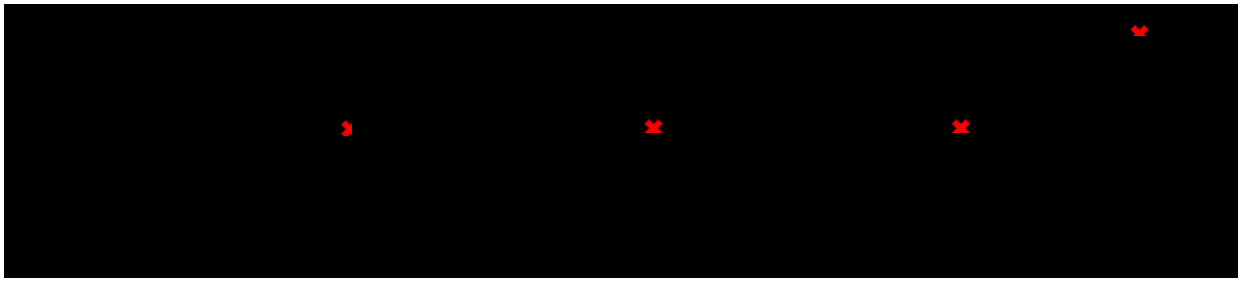



**Fig. 4:** Distributions of daily NOAA AR longitudinal and latitudinal displacements. The NOAA ARs are binned into four groups based on their absolute average latitude that are 0–±10 degrees, ±10–±20 degrees, ±20–±30 degrees, and ±30+ degrees. (a) Boxplots of longitudinal displacements, centered around 13 to 14 degrees. (b) Boxplots of latitudinal displacements, centered around 0 degrees. (c) Histograms of longitudinal displacement distributions for each group. (d) Histograms of latitudinal displacement distributions for each group.

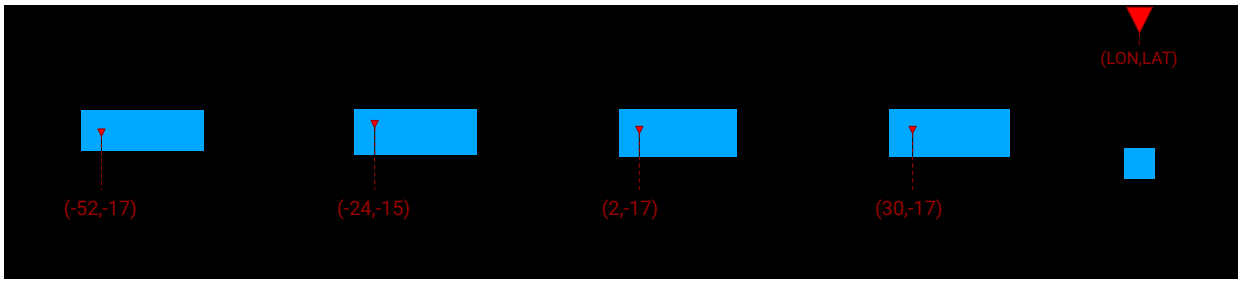
## IV. Evaluation and Analysis

To conduct our analysis, we use NOAA AR data retrieved from [20]. The source code for our analysis is also publicly available[1]. We included the plage regions to NOAA AR reports to track the movement of individual regions better. As mentioned earlier, we also manually updated the NOAA AR outliers, which we will discuss in the next subsection. Locations of SHARP series are obtained from JSOC, from header keywords [21]. SPoCA ARs are obtained from the Heliophysics Event Knowledge Base (HEK) database [22]. For all three data resources, the time range is between May 2010 and September 2018. Within this interval, we have 1,661 NOAA AR trajectories and 15,742 individual daily reports. In the same period, we have 14,375 SPoCA ARs (100,030 individual reports) and 4,075 SHARPs (2,613,321 individual reports). For each individual report in 4,075 SHARP trajectories, we annotate the corresponding NOAA AR# and SPoCA AR ID to each segment. We compared our findings with the SHARP-to-

[1]https://bitbucket.org/gsudmlab/st_cooccurrence_ar/src/master/



**(a)**




**(b)**

**Fig. 5:** Two example discrepancies: (a) SHARP 3686 and NOAA AR 11975, which is reported to be associated by JSOC provided SHARP-to-NOAA AR association list. (b) SHARP 2511 and NOAA AR 11688, which is not reported by JSOC's list.

NOAA AR associations provided in SHARP header keywords. Some SHARPs do not have a corresponding NOAA AR# while others may have one or more associated NOAA ARs. It is not expected that NOAA ARs are mapped to multiple SHARPs, although some of them are.

### *A. Updates to NOAA Active Region Reports*

During our analysis, we explored the daily movement patterns of NOAA ARs, primarily, to verify the correctness of NOAA-reported locations. We have found a few outliers in NOAA AR daily reports, which had unexpected longitudinal and latitudinal daily displacements which could not be explained with occasional flux emergence and enlargement of active regions.

To quantitatively identify these unexpected locations, we essentially performed an outlier analysis, where we checked the daily longitude and latitude changes. We binned the active regions based on their latitudes and created for groups of latitudes that are $0^\circ$ to $\pm 10^\circ$, $\pm 10^\circ$ to $\pm 20^\circ$, $\pm 20^\circ$ to $\pm 30^\circ$, and $\pm 30^\circ$ to $\pm 90^\circ$. We found the median longitude displacement for each of these and distributions can be seen in Fig 4. On average, the latitude displacement of the events stay within $\pm 8^\circ$ even for strongly emerging flux regions. We discovered only 19 NOAA ARs whose latitude change over $8^\circ$ daily. Similarly, the mean longitudinal displacement is $+13^\circ$ to $+14^\circ$ depending on the latitude and most ARs ($\sim$97%) stay within $+5^\circ$ to $+22^\circ$. Nevertheless, these location outliers for NOAA ARs potentially cause mismatches between the NOAA AR#s and HARPNUMs. We give one such example in Fig. 6, which shows the NOAA AR 11490 having an erroneous location on 2012-05-27. We observe that NOAA AR 11490 moves to the east ($+30^\circ$) and to the south ($-14^\circ$) and then moves back to its expected route. We considered NOAA AR movements that are within $\pm 8^\circ$ in latitude and $+5$ to $+22^\circ$ in longitude as the expected maximum limits of daily movements. If displacements are beyond these limits, we considered and identified them as anomalous movements. For each of those active regions which had anomalous daily movements, we manually updated the locations by examining the full disk AIA and HMI images. We believe that mismatches between JSOC provided and our SHARP-to-NOAA AR associations (presented in the next subsection) are caused by these movement anomalies. Therefore, we use the updated NOAA AR locations, practically to improve SHARP-to-NOAA AR associations. We made the trajectories of anomalous NOAA ARs publicly available together with the updated locations. We also share the full list of updated NOAA ARs in [23].

### *B. Discrepancy Analysis with JSOC Data*

In this part of our analysis, we compare SHARP-to-NOAA AR associations provided by JSOC to the co-occurrence-based associations we discovered. In JSOC-provided associations, there are 1314 SHARP series matched to NOAA ARs, 172 of which matched to two NOAA ARs, and 71 of which matched to three or more NOAA ARs. 2,761 SHARP series do not have any NOAA AR associations. In JSOC-provided associations, most of the NOAA ARs (1,517 of total 1,591) are matched to a single SHARP series, there are 74 NOAA ARs, which are matched to multiple SHARP series.

3970 of 4,075 SHARP-to-NOAA AR associations provided by JSOC are in accordance with our results. However, for 105 SHARP series,these associations do not fully match, i.e., for these SHARP series, either we could not find the associated NOAA AR#, we found an extra NOAA AR#, or a combination

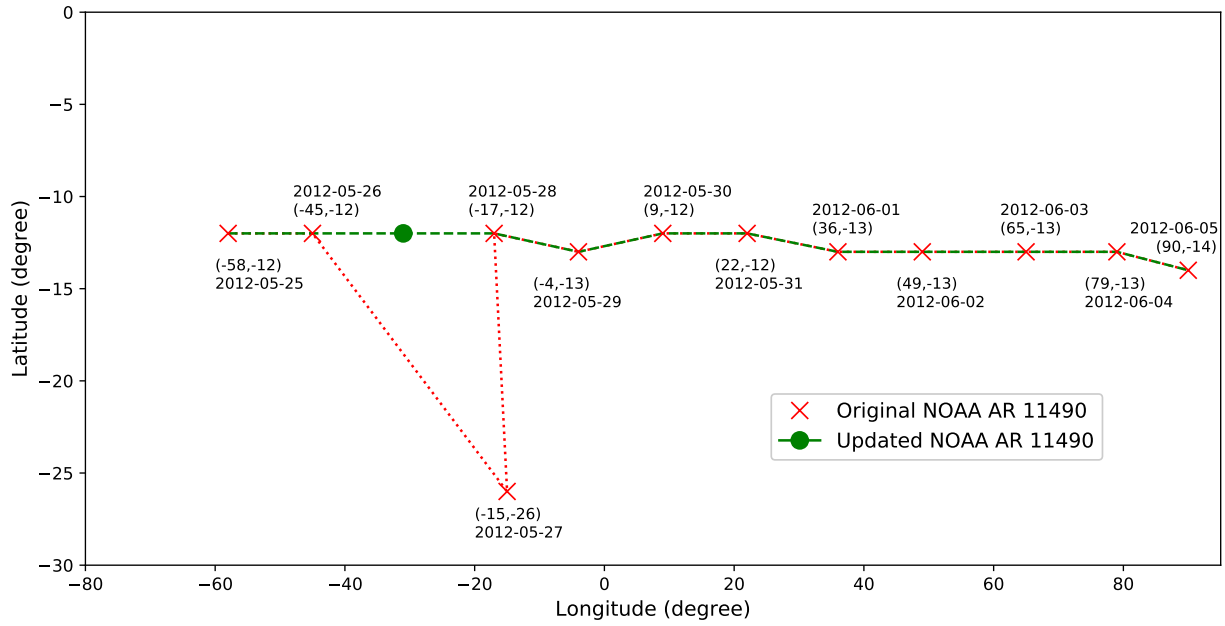



**Fig. 6:** Anomalous movement of NOAA AR 11490 on 2012-05-27 and we have updated the records for this day with (-31, -12) as shown with the green point in the scatter.

of these two. For the case of missing NOAA AR# (i.e., the ones we could not find the associations), some NOAA ARs and SHARPs do not coexist, while some others do coexist; but do not co-occur (meaning they are not colocated). For instance, while they are shown to be associated by JSOC, NOAA AR 11522 (lifespan between 2012-07-13 to 2012-07-18) and SHARP 1844 (lifespan between 2012-07-06 17:36 to 2012-07-08 20:00) do not have any temporal coexistence. On the other hand, NOAA AR 11975 and SHARP 3686 coexist but they do not spatially intersect. The relative locations of NOAA AR 11975 and SHARP 3686 are illustrated in Fig.5a. For the case of extra NOAA AR# detected, we essentially found a spatio-temporal co-occurrence not reported by JSOC. While some of these can be relatively weak co-occurrences, there are a few instances where we observed strong co-occurrences. Shown in Fig. 5b, NOAA AR 11688 and SHARP 2511 co-occur over 98% of their coexisting lifespans; however, their associations are not reported. To quantify the strength of the co-occurrence while avoiding misreportings caused by the inaccurate locations, we also calculated measures of trajectory similarity, which are co-occurrence factor ($cof$) and average minimum distance ($\mu_{mindist}$). Co-occurrence factor is calculated as the ratio between the lengths of co-occurrence and co-existence time intervals, while average minimum distance is calculated as the mean of the minimum Euclidean distance between individual NOAA AR and SHARP geometries at time $t_i$ for each timestamp they co-exist.

$$cof(\text{NOAA-AR, SHARP}) = \frac{\text{Length of co-occurrence time interval between NOAA-AR and SHARP}}{\text{Length of co-existence time interval between NOAA-AR and SHARP}} \quad (4)$$

$$\mu_{mindist} = \frac{1}{N}\sum_{t_i}^{N} mindist_{(NOAA-AR.g_i, SHARP.g_i)} \quad (5)$$

An example of weaker spatio-temporal co-occurrence is observed for SHARP 4390 and NOAA AR 12127, where they only co-occur $\sim$1.1% of their coexistence time intervals and on average they are they are over $9^\circ$ far apart.

Discrepancy analysis with JSOC Data is conducted in two steps. Firstly, based on the co-occurrence associations, we have 156 individual NOAA AR#-to-HARPNUM pair differences with JSOC. In the second step, we refine the difference pairs from 156 to 116 by manually examining each individual NOAA AR#-to-HARPNUM difference pairs on the full disk AIA and HMI image. We provide detailed explanation for each difference pairs in [24]. Finally, there are 116 individual NOAA AR#-to-HARPNUM pair differences (subdivided from 105 SHARP-to-NOAA AR discrepancies with JSOC), which are listed in Table I. Table I shows the lengths of NOAA AR and SHARP trajectories, co-existence and co-occurrence time intervals, as well as the co-occurrence factor and average minimum distance for each of these discrepancies. The full list of co-occurrence based SHARP-to-NOAA AR associations is also shared in [25].

## V. Conclusion

Active regions are responsible for most of the major, potentially threatening, explosive space weather events such as flares and CMEs. We believe that enhancing data quality enhances our ability to understand and predict these events, at the same time enabling the unobstructed use of machine

and deep learning techniques to this purpose. Magnetic field-based parameters, derived from SHARPs, are frequently used in flare prediction studies. Similarly, NOAA ARs are used for associating solar flares with active regions properties, more often than not with SHARPs. In case of spurious associations between NOAA ARs and SHARPs, we unintentionally feed incorrect data to learning models. Here, we have attempted to integrate the AR data from multiple feature recognition modules with using spatio-temporal co-occurrence analysis. We simply matched SHARPs, which have high temporal resolution, with both NOAA and SPoCA ARs. By doing that, (1) we provide a detailed spatio-temporal integration of solar active regions, (2) we discovered and updated a number of erroneous locations of NOAA ARs, and (3) using updated NOAA ARs, we further updated SHARP-to-NOAA AR associations.

In the future, we plan to extend our work to integrate not only active regions, but also other solar event types to our analysis, such as sunspot class information and sigmoids, which have flare-predictive attributes.

## VI. Acknowledgments

This project has been supported by funding from the Division of Advanced Cyberinfrastructure within the Directorate for Computer and Information Science and Engineering, the Division of Astronomical Sciences within the Directorate for Mathematical and Physical Sciences, and the Division of Atmospheric and Geospace Sciences within the Directorate for Geosciences, under NSF award #1443061.

## References

[1] M. J. Aschwanden and T. Shimizu, "MULTI-WAVELENGTH OBSERVATIONS OF THE SPATIO-TEMPORAL EVOLUTION OF SOLAR FLARES WITH AIA/SDO. II. HYDRODYNAMIC SCALING LAWS AND THERMAL ENERGIES," *The Astrophysical Journal*, vol. 776, no. 2, p. 132, oct 2013. [Online]. Available: https://doi.org/10.1088%2F0004-637x%2F776%2F2%2F132

[2] M. G. Bobra and S. Couvidat, "SOLAR FLARE PREDICTION USINGSDO/HMI VECTOR MAGNETIC FIELD DATA WITH a MACHINE-LEARNING ALGORITHM," *The Astrophysical Journal*, vol. 798, no. 2, p. 135, jan 2015. [Online]. Available: https://doi.org/10.1088%2F0004-637x%2F798%2F2%2F135

[3] E. Jonas, M. Bobra, V. Shankar, J. Todd Hoeksema, and B. Recht, "Flare prediction using photospheric and coronal image data," *Solar Physics*, vol. 293, no. 3, p. 48, Feb 2018. [Online]. Available: https://doi.org/10.1007/s11207-018-1258-9

[4] McCloskey, Aoife E., Gallagher, Peter T., and Bloomfield, D. Shaun, "Flare forecasting using the evolution of mcintosh sunspot classifications," *J. Space Weather Space Clim.*, vol. 8, p. A34, 2018. [Online]. Available: https://doi.org/10.1051/swsc/2018022

[5] B. Aydin, A. Kucuk, R. A. Angryk, and P. C. Martens, "Measuring the significance of spatiotemporal co-occurrences," *ACM Trans. Spatial Algorithms Syst.*, vol. 3, no. 3, pp. 9:1–9:35, Nov. 2017. [Online]. Available: http://doi.acm.org/10.1145/3139351

[6] B. Aydin and R. A. Angryk, *Modeling Spatiotemporal Relationships Among Trajectories*. Cham: Springer International Publishing, 2018, pp. 17–27. [Online]. Available: https://doi.org/10.1007/978-3-319-99873-2_3

[7] NOAA / NWS Space Weather Prediction Center, "Solar region summary," https://www.swpc.noaa.gov/products/solar-region-summary, (Accessed on 09/23/2019).

[8] M. G. Bobra, X. Sun, J. T. Hoeksema, M. Turmon, Y. Liu, K. Hayashi, G. Barnes, and K. D. Leka, "The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: Sharps – space-weather hmi active region patches," *Solar Physics*, vol. 289, no. 9, pp. 3549–3578, Sep 2014. [Online]. Available: https://doi.org/10.1007/s11207-014-0529-3

[9] Barra, V., Delouille, V., Kretzschmar, M., and Hochedez, J.-F., "Fast and robust segmentation of solar euv images: algorithm and results for solar cycle 23," *A&A*, vol. 505, no. 1, pp. 361–371, 2009. [Online]. Available: https://doi.org/10.1051/0004-6361/200811416

[10] P. S. McIntosh, "The classification of sunspot groups," *Solar Physics*, vol. 125, no. 2, pp. 251–267, Sep 1990. [Online]. Available: https://doi.org/10.1007/BF00158405

[11] F. Moerchen, "Tutorial cidm-t temporal pattern mining in symbolic time point and time interval data," in *2009 IEEE Symposium on Computational Intelligence and Data Mining*, March 2009, pp. xiv–xiv.

[12] J. F. ALLEN, "Maintaining knowledge about temporal intervals," in *Readings in Qualitative Reasoning About Physical Systems*, D. S. Weld and J. de Kleer, Eds. Morgan Kaufmann, 1990, pp. 361 – 372. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B978148321447450033X

[13] R. F. Howard, J. W. Harvey, and S. Forgach, "Solar surface velocity fields determined from small magnetic features," *Solar Physics*, vol. 130, no. 1, pp. 295–311, Dec 1990. [Online]. Available: https://doi.org/10.1007/BF00156795

[14] B. J. LaBonte, M. K. Georgoulis, and D. M. Rust, "Survey of magnetic helicity injection in regions producing x-class flares," *The Astrophysical Journal*, vol. 671, no. 1, pp. 955–963, dec 2007. [Online]. Available: https://doi.org/10.1086%2F522682

[15] S. Shekhar and H. Xiong, *Qualitative Spatial Reasoning*. Boston, MA: Springer US, 2008, pp. 934–934. [Online]. Available: https://doi.org/10.1007/978-0-387-35973-1_1058

[16] Z. Jiang and S. Shekhar, *Spatial and Spatiotemporal Big Data Science*. Cham: Springer International Publishing, 2017, pp. 15–44. [Online]. Available: https://doi.org/10.1007/978-3-319-60195-3_2

[17] S. Shekhar and H. Xiong, *Encyclopedia of GIS*. Springer Science & Business Media, 2007.

[18] M. Schneider and T. Behr, "Topological relationships between complex spatial objects," *ACM Trans. Database Syst.*, vol. 31, no. 1, pp. 39–81, Mar. 2006. [Online]. Available: http://doi.acm.org/10.1145/1132863.1132865

[19] C. Strobl, *Dimensionally Extended Nine-Intersection Model (DE-9IM)*. Cham: Springer International Publishing, 2017, pp. 470–476. [Online]. Available: https://doi.org/10.1007/978-3-319-17885-1_298

[20] ftp://ftp.swpc.noaa.gov/pub/warehouse/, (Accessed on 09/26/2019).

[21] J. T. Hoeksema, Y. Liu, K. Hayashi, X. Sun, J. Schou, S. Couvidat, A. Norton, M. Bobra, R. Centeno, K. D. Leka, G. Barnes, and M. Turmon, "The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: Overview and performance," *Solar Physics*, vol. 289, no. 9, pp. 3483–3530, Sep 2014. [Online]. Available: https://doi.org/10.1007/s11207-014-0516-8

[22] N. Hurlburt, M. Cheung, C. Schrijver, L. Chang, S. Freeland, S. Green, C. Heck, A. Jaffey, A. Kobashi, D. Schiff, J. Serafin, R. Seguin, G. Slater, A. Somani, and R. Timmons, *Heliophysics Event Knowledgebase for the Solar Dynamics Observatory (SDO) and Beyond*. New York, NY: Springer US, 2012, pp. 67–78. [Online]. Available: https://doi.org/10.1007/978-1-4614-3673-7_5

[23] X. Cai, "Updated noaa ar trajectory," http://dmlab.cs.gsu.edu/wp-content/uploads/appendix_noaa_ars_plages_n_updated1.csv, Nov 2019.

[24] X. Cai, "SHARP to NOAA Discrepancies," http://dmlab.cs.gsu.edu/wp-content/uploads/appendix_sharp_to_noaa_discrepancy_remarks.csv, Nov 2019.

[25] X. Cai, "Full list of HARPNUM-to-NOAA AR Number Associations," http://dmlab.cs.gsu.edu/wp-content/uploads/MU_all_harps_with_noaa_ars.txt, Nov 2019.

**TABLE I:** 116 individual NOAA AR number to HARPNUM discrepancies compared to JSOC-provided list

| No | HARPNUM | NOAA AR# | Co_existance (hours) | Co_Occurence (hours) | Cof_rate (%) | HARP_records (hours) | NOAA_records (hours) | avg_min_distance (deg) | min_distance (deg) | Is Extra | Is NOAA Location Updated |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 45 | 11073 | 134.2 | 0 | 0 | 134.2 | 240 | 5.779099964 | 1.660613549 | | |
| 2 | 700 | 11253 | 66.4 | 0 | 0 | 226.4 | 144 | 5.50657413 | 5.112789491 | | |
| 3 | 714 | 11256 | 199 | 0 | 0 | 309 | 216 | 4.056030007 | 2.906088 | | |
| 4 | 734 | 11253 | 7.4 | 0 | 0 | 7.4 | 144 | 11.0565582 | 9.877950294 | | |
| 5 | 877 | 11293 | 11.8 | 7.4 | 62.71 | 24.6 | 144 | 0.169046148 | 0 | extra | |
| 6 | 1124 | 11373 | 91.2 | 0 | 0 | 322 | 96 | 3.746813623 | 1.8571 | | |
| 7 | 1133 | 11365 | 155.4 | 0 | 0 | 182.8 | 192 | 9.184393267 | 6.611815826 | | update |
| 8 | 1165 | 11367 | 159.8 | 0 | 0 | 159.8 | 240 | 7.287003759 | 5.506166 | | update |
| 9 | 1295 | 11394 | 132.2 | 24 | 18.15 | 142 | 240 | 2.620807277 | 0 | extra | |
| 10 | 1493 | 11437 | 53.2 | 0 | 0 | 53.2 | 192 | 3.297733319 | 2.926298 | | |
| 11 | 1535 | 11442 | 8.4 | 7.2 | 85.71 | 8.4 | 264 | 0.065225881 | 0 | extra | |
| 12 | 1578 | 11464 | 166 | 0 | 0 | 258 | 168 | 3.136260794 | 2.245369 | | |
| 13 | 1633 | 11468 | 11.8 | 11.4 | 96.61 | 11.8 | 192 | 0.008860237 | 0 | extra | |
| 14 | 1662 | 11487 | 145.4 | 0 | 0 | 318.4 | 168 | 2.918219503 | 1.660772064 | | |
| 15 | 1697 | 11489 | 217.4 | 0 | 0 | 292.4 | 240 | 12.72621982 | 6.968548208 | | update |
| 16 | 1724 | 11502 | 195.2 | 0 | 0 | 303.8 | 240 | 5.779507335 | 3.82651544 | | |
| 17 | 1844 | 11522 | — | — | — | 50 | 144 | — | — | | |
| 18 | 1845 | 11522 | 131.2 | 128.4 | 97.87 | 259.6 | 144 | 0.010065763 | 0 | extra | |
| 19 | 1907 | 11545 | 60 | 0 | 0 | 337.6 | 120 | 5.684938152 | 0.586471842 | | |
| 20 | 1998 | 11557 | 68.2 | 0 | 0 | 68.2 | 240 | 20.91493134 | 20.2544632 | | |
| 21 | 2155 | 11594 | 23.8 | 7 | 29.41 | 23.8 | 288 | 0.525051408 | 0 | extra | |
| 22 | 2360 | 11655 | 156.2 | 121.2 | 77.59 | 275.6 | 168 | 1.927847647 | 0 | extra | |
| 23 | 2375 | 11642 | 15.2 | 0 | 0 | 15.2 | 312 | 7.232228477 | 6.650084336 | | |
| 24 | 2403 | 11652 | 3.4 | 0 | 0 | 3.4 | 312 | 6.011314655 | 5.464216783 | | |
| 25 | 2432 | 11666 | 130.6 | 10.4 | 7.96 | 168.6 | 288 | 1.697249223 | 0 | extra | |
| 26 | 2439 | 11667 | 286.4 | 286.4 | 100 | 307.4 | 288 | 0 | 0 | extra | |
| 27 | 2442 | 11665 | 9.2 | 0 | 0 | 9.2 | 312 | 4.192045503 | 3.341770472 | | |
| 28 | 2469 | 11674 | 168 | 18.6 | 11.07 | 328.8 | 168 | 6.109594474 | 0 | extra | |
| 29 | 2492 | 11679 | 203.4 | 72 | 35.4 | 282.2 | 264 | 2.040053432 | 0 | extra | |
| 30 | 2511 | 11688 | 197 | 194.6 | 98.78 | 297.4 | 264 | 0.013973413 | 0 | extra | |
| 31 | 2546 | 11694 | 261 | 203.4 | 77.93 | 299.8 | 288 | 0.236215101 | 0 | extra | |
| 32 | 2546 | 11701 | 53.8 | 53.8 | 100 | 299.8 | 72 | 0 | 0 | extra | |
| 33 | 2597 | 11708 | 280 | 216.2 | 77.21 | 332.2 | 288 | 2.214304008 | 0 | extra | |
| 34 | 2666 | 11717 | 7.6 | 0 | 0 | 7.6 | 288 | 3.914317406 | 3.746405927 | | |
| 35 | 2790 | 11761 | 237.8 | 0 | 0 | 319.8 | 264 | 4.822431514 | 2.445867973 | | |
| 36 | 2878 | 11780 | 151.6 | 0 | 0 | 315.2 | 192 | 7.831242186 | 4.939963571 | | |
| 37 | 2948 | 11789 | 74.8 | 0 | 0 | 74.8 | 216 | 8.255289803 | 6.248741788 | | update |
| 38 | 3115 | 11826 | 103.6 | 0 | 0 | 103.6 | 192 | 2.83170128 | 1.047685673 | | |
| 39 | 3190 | 11844 | 1 | 1 | 100 | 16 | 72 | 0 | 0 | extra | |
| 40 | 3212 | 11845 | 146 | 0 | 0 | 146 | 288 | 4.588168768 | 2.363810335 | | |
| 41 | 3252 | 11866 | 41.4 | 0 | 0 | 217.4 | 48 | 7.8100391 | 2.93310817 | | |
| 42 | 3263 | 11861 | 301.8 | 0 | 0 | 305.8 | 312 | 6.335951837 | 1.281207 | | |
| 43 | 3520 | 11935 | 226.6 | 0 | 0 | 356.6 | 240 | 5.283393071 | 1.936610134 | | |
| 44 | 3520 | 11939 | 48 | 0 | 0 | 356.6 | 48 | 4.800767896 | 3.911222 | | |
| 45 | 3557 | 11938 | 158.2 | 0 | 0 | 158.2 | 288 | 4.638900558 | 1.880400641 | | |
| 46 | 3602 | 11938 | 0.6 | 0 | 0 | 0.6 | 288 | 0.878634167 | 0.643649091 | | |
| 47 | 3686 | 11975 | 45.4 | 0 | 0 | 340.8 | 120 | 17.99319668 | 10.15517622 | | |
| 48 | 3744 | 11979 | 3.6 | 3.6 | 100 | 21.8 | 96 | 0 | 0 | extra | |
| 49 | 4252 | 12094 | 261.4 | 0 | 0 | 293 | 288 | 3.129049832 | 0.595583 | | update |
| 50 | 4296 | 12108 | 259.8 | 0 | 0 | 338.8 | 312 | 5.142132787 | 2.061919536 | | |
| 51 | 4346 | 12103 | 38 | 20.8 | 54.74 | 38 | 264 | 1.770064002 | 0 | extra | |
| 52 | 4424 | 12140 | 70.4 | 70.4 | 100 | 335.6 | 72 | 0 | 0 | extra | |
| 53 | 4432 | 12136 | 56 | 18.8 | 33.57 | 175.6 | 96 | 1.336574963 | 0 | extra | |
| 54 | 4438 | 12137 | 126.6 | 126.6 | 100 | 232.8 | 144 | 0 | 0 | extra | |
| 55 | 4440 | 12135 | 278.4 | 278.4 | 100 | 281.2 | 288 | 0 | 0 | extra | |
| 56 | 4447 | 12144 | 135.4 | 133.2 | 98.38 | 238.6 | 144 | 0.010318564 | 0 | extra | |
| 57 | 4448 | 12139 | 283.4 | 273.4 | 96.47 | 308.8 | 312 | 0.028469395 | 0 | extra | |
| 58 | 4450 | 12138 | 45.2 | 45.2 | 100 | 53.6 | 48 | 0 | 0 | extra | |
| 59 | 4454 | 12143 | 266.4 | 266.4 | 100 | 300.4 | 288 | 0 | 0 | extra | |
| 60 | 4455 | 12141 | 274.8 | 274.8 | 100 | 287.4 | 288 | 0 | 0 | extra | |
| 61 | 4455 | 12142 | 263 | 40.6 | 15.44 | 287.4 | 264 | 2.983788402 | 0 | extra | |
| 62 | 4460 | 12145 | 71.4 | 36.8 | 51.54 | 81.6 | 192 | 0.923697125 | 0 | extra | |
| 63 | 4466 | 12146 | 310 | 310 | 100 | 340.6 | 312 | 0 | 0 | extra | |
| 64 | 4466 | 12148 | 233.6 | 209.6 | 89.73 | 340.6 | 240 | 0.103752561 | 0 | extra | |
| 65 | 4469 | 12147 | 263.8 | 234.2 | 88.78 | 271 | 288 | 0.22025427 | 0 | extra | |
| 66 | 4477 | 12149 | 290.6 | 290.6 | 100 | 298.6 | 312 | 0 | 0 | extra | |
| 67 | 4478 | 12150 | 294.6 | 289.4 | 98.23 | 312 | 312 | 0.091014047 | 0 | extra | |
| 68 | 4478 | 12151 | 280 | 280 | 100 | 312 | 312 | 0 | 0 | extra | |
| 69 | 4502 | 12152 | 235.6 | 232.8 | 98.81 | 274.8 | 288 | 0.011602093 | 0 | extra | |
| 70 | 4505 | 12153 | 176.6 | 175.6 | 99.43 | 226 | 216 | 0.00149461 | 0 | extra | |
| 71 | 4523 | 12154 | 176.6 | 167.4 | 94.79 | 218.6 | 240 | 0.210376027 | 0 | extra | |
| 72 | 4523 | 12161 | 93 | 93 | 100 | 218.6 | 168 | 0 | 0 | extra | |
| 73 | 4530 | 12155 | 269 | 217 | 80.67 | 284.4 | 312 | 1.076426807 | 0 | extra | |
| 74 | 4530 | 12157 | 257.4 | 257.4 | 100 | 284.4 | 312 | 0 | 0 | extra | |
| 75 | 4532 | 12156 | 149.4 | 119.2 | 79.79 | 170.2 | 264 | 0.180614175 | 0 | extra | |
| 76 | 4536 | 12158 | 279.6 | 279.6 | 100 | 282.6 | 336 | 0 | 0 | extra | |
| 77 | 4539 | 12159 | 237.6 | 168.2 | 70.79 | 240.2 | 288 | 1.157205748 | 0 | extra | |
| 78 | 4540 | 12162 | 108 | 98.6 | 91.3 | 124.4 | 168 | 0.041813918 | 0 | extra | |

*Continued on next page*

TABLE I – *Continued from previous page*

| No | HARPNUM | NOAA AR# | Co_existance (hours) | Co_Occurence (hours) | Cof_rate (%) | HARP_records (hours) | NOAA_records (hours) | avg_min_distance deg) | min_distance (deg) | Is Extra | Is NOAA Location Updated |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 79 | 4541 | 12160 | 62.2 | 62.2 | 100 | 72.8 | 72 | 0 | 0 | extra | |
| 80 | 4543 | 12163 | 212.4 | 210.2 | 98.96 | 235.2 | 288 | 0.002252748 | 0 | extra | |
| 81 | 4549 | 12164 | 190 | 188 | 98.95 | 218.6 | 216 | 0.000752433 | 0 | extra | |
| 82 | 4552 | 12165 | 214.2 | 210.2 | 98.13 | 236 | 288 | 0.047969633 | 0 | extra | |
| 83 | 4556 | 12166 | 262.8 | 234.4 | 89.19 | 285 | 288 | 1.463206985 | 0 | extra | update |
| 84 | 4556 | 12167 | 227.2 | 202 | 88.91 | 285 | 264 | 0.968023839 | 0 | extra | |
| 85 | 4574 | 12169 | 252.8 | 231.6 | 91.61 | 272.6 | 288 | 0.312876686 | 0 | extra | |
| 86 | 4574 | 12170 | 252.8 | 246.8 | 97.63 | 272.6 | 288 | 0.013181776 | 0 | extra | |
| 87 | 4576 | 12168 | 176 | 153.8 | 87.39 | 189.4 | 240 | 0.155703105 | 0 | extra | |
| 88 | 4579 | 12174 | 32.2 | 32.2 | 100 | 81.2 | 72 | 0 | 0 | extra | |
| 89 | 4580 | 12171 | 281.6 | 274.8 | 97.59 | 353.2 | 288 | 0.156997532 | 0 | extra | |
| 90 | 4580 | 12172 | 310.4 | 298.6 | 96.2 | 353.2 | 336 | 0.031699436 | 0 | extra | |
| 91 | 4580 | 12173 | 307.6 | 299.8 | 97.46 | 353.2 | 312 | 0.044374366 | 0 | extra | |
| 92 | 4591 | 12175 | 161.6 | 161.6 | 100 | 271.4 | 168 | 0 | 0 | extra | |
| 93 | 4603 | 12176 | 235.4 | 219.2 | 93.12 | 281.6 | 264 | 0.028341344 | 0 | extra | |
| 94 | 4616 | 12181 | 208.2 | 0 | 0 | 295.2 | 216 | 4.038821339 | 0.310715 | | |
| 95 | 4619 | 12180 | 80.4 | 17.6 | 21.89 | 142 | 144 | 3.927349468 | 0 | extra | |
| 96 | 4661 | 12184 | 224.4 | 0 | 0 | 274.4 | 288 | 14.62560661 | 8.554640144 | | update |
| 97 | 4862 | 12224 | 204.2 | 0 | 0 | 337 | 216 | 3.389071133 | 0.512933542 | | |
| 98 | 5342 | 12309 | 166.2 | 0 | 0 | 318 | 168 | 5.815746584 | 0.646579455 | | |
| 99 | 5387 | 12312 | 217.4 | 0 | 0 | 261.2 | 288 | 17.62617339 | 15.58457491 | | update |
| 100 | 5434 | 12317 | 33.6 | 0 | 0 | 33.6 | 144 | 8.019439155 | 0.274898 | | |
| 101 | 5596 | 12352 | 129.6 | 0 | 0 | 201.6 | 168 | 6.506478185 | 4.793330721 | | |
| 102 | 5708 | 12372 | 9.4 | 9.4 | 100 | 15.2 | 48 | 0 | 0 | extra | |
| 103 | 5738 | 12383 | 149 | 0 | 0 | 282 | 168 | 3.782208893 | 0.826734 | | |
| 104 | 5750 | 12378 | 189 | 0 | 0 | 189 | 288 | 6.372000005 | 0.343558 | | |
| 105 | 6063 | 12447 | 172.2 | 0 | 0 | 314 | 216 | 4.363646865 | 2.482430238 | | |
| 106 | 6148 | 12461 | 7.8 | 7.8 | 100 | 14.4 | 96 | 0 | 0 | extra | |
| 107 | 6155 | 12467 | 141 | 0 | 0 | 292.2 | 192 | 2.041529343 | 0.658325 | | |
| 108 | 6178 | 12474 | 71.6 | 0 | 0 | 333 | 72 | 1.249175067 | 0.688929 | | |
| 109 | 6361 | 12503 | 154.2 | 0 | 0 | 221.2 | 168 | 4.707918139 | 1.479921576 | | |
| 110 | 6523 | 12540 | 199.2 | 0 | 0 | 265.2 | 264 | 4.654517664 | 3.056984589 | | |
| 111 | 6688 | 12572 | 9.2 | 0 | 0 | 153.2 | 96 | 3.990382908 | 3.615620253 | | |
| 112 | 6777 | 12593 | 12 | 0 | 0 | 12 | 192 | 0.997705522 | 0.485976309 | | |
| 113 | 6893 | 12622 | 7.6 | 7.6 | 100 | 203.8 | 72 | 0 | 0 | extra | |
| 114 | 6901 | 12623 | 7.4 | 7.4 | 100 | 32.2 | 144 | 0 | 0 | extra | |
| 115 | 7078 | 12667 | 9 | 9 | 100 | 109 | 48 | 0 | 0 | extra | |
| 116 | 7123 | 12675 | 84.6 | 0 | 0 | 84.6 | 144 | 6.634176599 | 0.562046 | | |

$^{a}$ extra represent extra NOAA AR# in our co-occurrence based SHARP-to-NOAA AR association, $^{b}$ update represent NOAA AR event records have been updated by outlier detection