J. Space Weather Space Clim., **6**, A22 (2016)
DOI: 10.1051/swsc/2016015

RESEARCH ARTICLE OPEN ACCESS

A large-scale dataset of solar event reports from automated feature recognition modules

Michael A. Schuh[1,*], Rafal A. Angryk[2], and Petrus C. Martens[3]

[1] Dept. of Computer Science, Montana State University, Bozeman, MT 59717, USA
[*]Corresponding author: michael.schuh@cs.montana.edu
[2] Dept. of Computer Science, Georgia State University, Atlanta, GA 30303, USA
[3] Dept. of Physics and Astronomy, Georgia State University, Atlanta, GA 30303, USA

Received 1 February 2015 / Accepted 18 March 2016

ABSTRACT

The massive repository of images of the Sun captured by the Solar Dynamics Observatory (SDO) mission has ushered in the era of Big Data for Solar Physics. In this work, we investigate the entire public collection of events reported to the Heliophysics Event Knowledgebase (HEK) from automated solar feature recognition modules operated by the SDO Feature Finding Team (FFT). With the SDO mission recently surpassing five years of operations, and over 280,000 event reports for seven types of solar phenomena, we present the broadest and most comprehensive large-scale dataset of the SDO FFT modules to date. We also present numerous statistics on these modules, providing valuable contextual information for better understanding and validating of the individual event reports and the entire dataset as a whole. After extensive data cleaning through exploratory data analysis, we highlight several opportunities for knowledge discovery from data (KDD). Through these important prerequisite analyses presented here, the results of KDD from Solar Big Data will be overall more reliable and better understood. As the SDO mission remains operational over the coming years, these datasets will continue to grow in size and value. Future versions of this dataset will be analyzed in the general framework established in this work and maintained publicly online for easy access by the community.

Key words. Solar activity – Data analysis – Data mining – Validation – Statistics and probability

1. Introduction

The era of Big Data is here for Solar Physics. With the Solar Dynamics Observatory (SDO) mission capturing over 150,000 high-resolution full-disk images of the Sun per day (Pesnell et al. 2012), never before has there been such a massive volume of solar images available. Given this deluge of data that will likely only increase with future missions, it is infeasible to continue traditional brute-force human analysis and labeling of solar phenomena in every image. In response to this issue, general research in automated detection analysis is becoming increasingly popular in Solar Physics, utilizing algorithms from computer vision, image processing, and machine learning.

Automated event detection algorithms of the SDO Feature Finding Team (FFT) (Martens et al. 2012) process the continuous stream of image data and report event findings to the public Heliophysics Event Knowledgebase (HEK; Hurlburt et al. 2012). With an abundance of generated event reports, we are able to analyze large-scale statistics and facilitate knowledge discovery from data (KDD). This better picture of solar phenomena (events) through direct large-scale observations has the unprecedented potential of further advancing scientific understanding and possible predictions of such events and related space weather processes. The importance of better understanding and prediction of space weather cannot be understated. Many modern technological conveniences are affected by space weather, including: radio and GPS communications, air and space flight health and safety, and energy grid and power infrastructures. It has been estimated that the damage from a severe solar storm could exceed $2 trillion (USD) in total economic impact and a full recovery could take years (Council 2008).

The dataset presented here[1] combines the metadata from all HEK-available FFT automated detection modules that run continuously in a dedicated SDO data pipeline. Much like the need to automate the detection and reporting of events from these massive image repositories, the fundamental analyses of the event reports should also be a formally structured and semi-automated process that allows rapid and up-to-date data curation from continuously running modules. This work builds upon our initial investigation into creating large-scale datasets with SDO data products (Schuh et al. 2013; Schuh & Angryk 2014). We revisit the data collection process with more advanced and detailed analyses, statistics, and visualizations to better standardize (and automate) the dataset creation process. The main contribution of this work is to present an overview of the entire SDO FFT event data present and available to the public and outline a general framework for large-scale data cleaning and validation of the diverse module data products to be assimilated together into a singular standardized benchmark dataset for advanced research in the future.

In Section 2, we provide an overview of the SDO mission and the specific data sources used. Section 3 presents the general data curation and analysis processes, including specific cleaning and validation steps. Then we highlight several more advanced spatial and temporal analyses that are now capable in

[1] More information at http://dmlab.cs.montana.edu/solar/

Section 4, and we briefly discuss our future work and conclusions in Section 5.

2. Background

The SDO mission is the first mission of NASA's Living With a Star (LWS) program, a long-term project dedicated to studying aspects of the Sun that significantly affect human life, with the goal of eventually developing a scientific understanding sufficient for prediction of solar activity (Withbroe 2000). Launched on February 11, 2010, the SDO is a 3-axis stabilized spacecraft designed to continuously monitor the Sun in a variety of ways. It contains three independent instruments, with the FFT modules mostly using the Atmospheric Imaging Assembly (AIA) instrument, which captures images in ten separate wavebands selected to highlight specific elements of solar activity (Lemen et al. 2012) – seven of these are in extreme ultraviolet (EUV), two are in ultraviolet (UV), and one is in the visible-light spectrum. Images from the Helioseismic and Magnetic Imager (HMI) instrument, which was designed to study variabilities of the magnetic field of the photosphere (Scherrer et al. 2012), are also used by several FFT modules.

The SDO Feature Finding Team (FFT)[2] is an international consortium of independent groups selected by NASA to produce a comprehensive set of automated feature recognition modules (Martens et al. 2012). The SDO FFT modules operate through the SDO Event Detection System (EDS) at the Joint Science Operations Center (JSOC) of Stanford and Lockheed Martin Solar and Astrophysics Laboratory (LMSAL), as well as the Harvard-Smithsonian Center for Astrophysics (CfA), Montana State University (MSU), and NASA's Goddard Space Flight Center (GSFC). While all SDO data is made publicly accessible in a timely fashion, because of the overall size, only a small window of original image data is available for on-demand access, while tapes provide long-term archival storage. Therefore, the majority of FFT modules utilize direct access to the raw data pipeline for stream-like data analysis and event detection in a near real-time manner. The mission recently marked five years of operations, and at an approximate data rate of 1.5 TB per day, that equates to almost 2.75 Petabytes (PB) of data already processed by the SDO mission.

The purpose of the SDO mission is to gather knowledge about the mechanics of solar magnetic activity, from the generation of the solar magnetic field to the release of magnetic energy in the solar wind, solar flares, coronal mass ejections (CMEs), and other events (Pesnell et al. 2012). Through the use of broad, long-term, and large-scale event datasets, we can apply data mining techniques to greatly aid in the pursuits of these goals through data-driven knowledge discovery. This includes many avenues of interdisciplinary research, such as well-founded event statistics leading to image parameter feature selection (Banda et al. 2010) and event classification (Schuh et al. 2014), visualization techniques (Schuh et al. 2015) and information retrieval (Banda et al. 2014) of similar events or image regions of interest, and spatiotemporal co-occurrence rules (Pillai et al. 2014) and pattern mining (Aydin et al. 2015) of related solar events.

While there exist many other event detection algorithms, we limit our scope to focus solely on the SDO FFT modules for several reasons. First, all modules were deployed in a unified strategy and should therefore share many similar overarching implementation decisions and goals (Martens et al. 2012). To the best of our knowledge, no broad (multi-module) and comprehensive (all reported metadata and event attributes) follow-up investigations have been performed on the data submitted to the HEK by the FFT modules after becoming operational. Second, because each solar phenomenon is reported by only one SDO FFT module, we can avoid the need for cross-module comparison of the same event types, which while beyond the scope of this work, is a very important topic of research for data calibration and event verification. Lastly, and most importantly, since these dedicated FFT modules run continuously and automatically on newly received SDO data, the available event reports will continue to increase and the assimilated dataset can grow without additional human or module efforts. Such automated processes still require periodic validation and assessment, and this work is the first introduction of such a framework.

3. The data

Here we present the overarching data collection and analysis process. The goal is to determine exactly what data is available and what sort of quality or trustworthiness it exhibits, while raising questions about possible outliers, trends, or anomalies discovered along the way. While loose guidelines for the FFT modules' products and proper HEK reporting are available, we stress the need for empirical verification, as real-world data rarely fully adheres to expectation. We first look at basic statistics and distributions on the event reporting of each module (metadata), followed by a similar statistical analysis of each of the available and pertinent module and event attributes. We also visualize several spatial and temporal attributes for easy large-scale validation and comparison between different modules and event types. Due to an abundance of raw statistics and charts, we limit the discussion here to a selection of the most interesting results for each topic presented. Our website[3] contains all results and datasets for this work and future updates.

3.1. Collection

Events are collected from all available SDO FFT modules over the entire current mission lifetime of five years, from early 2010 through the end of 2014. These events are reported to the Heliophysics Event Knowledgebase (HEK), which is a centralized archive of solar event reports publicly accessible online (Hurlburt et al. 2012). While event metadata can be downloaded manually through the official web interface,[4] for efficient and automated large-scale retrieval, we previously developed an open-source and publicly available software application named "Query HEK", or simply QHEK.[5] Using QHEK, we retrieve all FFT module reports for seven types of solar events (phenomena): active region (AR), coronal hole (CH), emerging flux (EF), filament (FI), flare (FL), sigmoid (SG), and sunspot (SS). For reference, the names and abbreviations of these event types, as well as the name of each reporting module, can be found in Table 1.

There are two important things to understand about how the FFT modules report solar activity. First, these reports provide a required start and end time as well as a location of the event at a moment within that time frame. Depending on the

[2] http://solar.physics.montana.edu/sol_phys/fft/

[3] http://dmlab.cs.montana.edu/solar/

[4] http://www.lmsal.com/isolsearch

[5] http://dmlab.cs.montana.edu/solar/qhek

Table 1. An overview of event types and their automated detection modules.

Event name	Event label	FFT module name	Reference
Active region	AR	SPoCA	Verbeeck et al. (2014)
Coronal hole	CH	SPoCA	Verbeeck et al. (2014)
Emerging flux	EF	Emerging flux region module	Martens et al. (2012)
Filament	FI	AAFDCC	Bernasconi et al. (2005)
Flare	FL	Flare detective – trigger module	Martens et al. (2012)
Sigmoid	SG	Sigmoid sniffer	Martens et al. (2012)
Sunspot	SS	EGSO_SFC	Zharkov et al. (2005)

Table 2. A summary of event types grouped by unique source metadata.

Event type	Observatory	Instrument	Channel	Reports
AR	SDO	AIA	AIA 171, AIA 193	65,566
CH	SDO	AIA	AIA 193	47,963
EF	SDO/HMI	HMI_FRONT2	LOS magnetograms	31,078
FI	BBSO	BBSOHa	H-alpha	11,674
FI	KANZELHOEHE OBSERVATORY	HA2	H-alpha	21,774
FL	SDO	AIA	131	36,270
FL	SDO	AIA	171	12,500
FL	SDO	AIA	211	9,570
FL	SDO	AIA	193	6,787
FL	SDO	AIA	304	3,279
FL	SDO	AIA	94	1,837
FL	SDO	AIA	335	1,668
SG	SDO	AIA_1	131_THIN	11,522
SG	SDO	AIA_1	131_THICK	4
SG	SDO	AIA_1	94_THIN	9,189
SG	SDO	AIA_1	94_THICK	2
SS	SDO	HMI	SDO/HMI	11,525

module and type of phenomenon, a single event report to the HEK may represent only a snapshot of a solar phenomenon over its entire life-span. Therefore, the number of event reports is in most cases much larger than the number of actual solar phenomena that existed. Second, given the difference just discussed between describing these solar events, the term "event" itself is overloaded, referring to either a solar phenomenon or a module report. To avoid ambiguous usage, and given the scope of this work, we typically refer to an event as a single module report, which may be one of many reports over the lifetime of the single phenomenon. Given these clean and validated event report datasets, one direction of current research involves tracking phenomena over time and linking event reports together that refer to the same physical solar phenomenon (Kempton et al. 2014; Kempton & Angryk 2015). We also note that several modules have begun to provide preliminary tracking metadata, but the use of this information is beyond the scope of this work.

A summary of all retrieved event reports can be found in Table 2, which states the identifying source information (metadata) of reported events by the observatory, instrument, and channel attributes. We also include the total count of event reports for each unique source of events. In Figure 1, we show examples for all seven types of events on an AIA 171Å image from January 20, 2012. While each type of event is detected by one specific module, we see some modules can operate on more than one data source – namely the FI, FL, SG event detection modules. We chose to include FI events because the automated module is part of the SDO FFT and reports to HEK, although we see here it operates on two distinct sources of H-α images from the ground-based observatories of Big Bear Solar Observatory (BBSO) in the United States and Kanzelhöehe Observatory in Austria. From the perspective of data standardization, we see several discrepancies between module reporting standards on the metadata source fields, e.g., instrument "HMI" vs. "HMI_FRONT2" or channel "AIA 193" vs. "193", which could directly impact attempts at source-based event retrieval. We also see the SG events have entirely unique instrument and channel values, and the module reports only six events in the "thick" channels compared to thousands of events in the "thin" channels. According to Lemen et al. (2012), this thick channel can be used for additional attenuation of the AIA 131Å channel during very bright solar flares.

All event reports are inserted into a PostgreSQL relational database, where each report is a single record. This allows for more efficient access and much easier manipulation of complex sets of records retrieved with SQL statements. A generalized event table contains the event report details that are required and common for all event types, which includes: event type, start time, end time, observatory, instrument, channel, center, bounding box (bbox), and chain code (ccode). The final three fields describe the event's spatial location and will be discussed later. Through the use of a unique event ID for each record and the event type field, all event-specific attributes for each report (record) are linked in a separate table for the specific event type. For convenience, we provide SQL files to use the dataset in this well-structured database format.

3.2. Cleaning checks

We perform several checks to ensure the integrity of the database of collected events. These checks focus on valid data attribute values that do not require extensive domain expertise, so

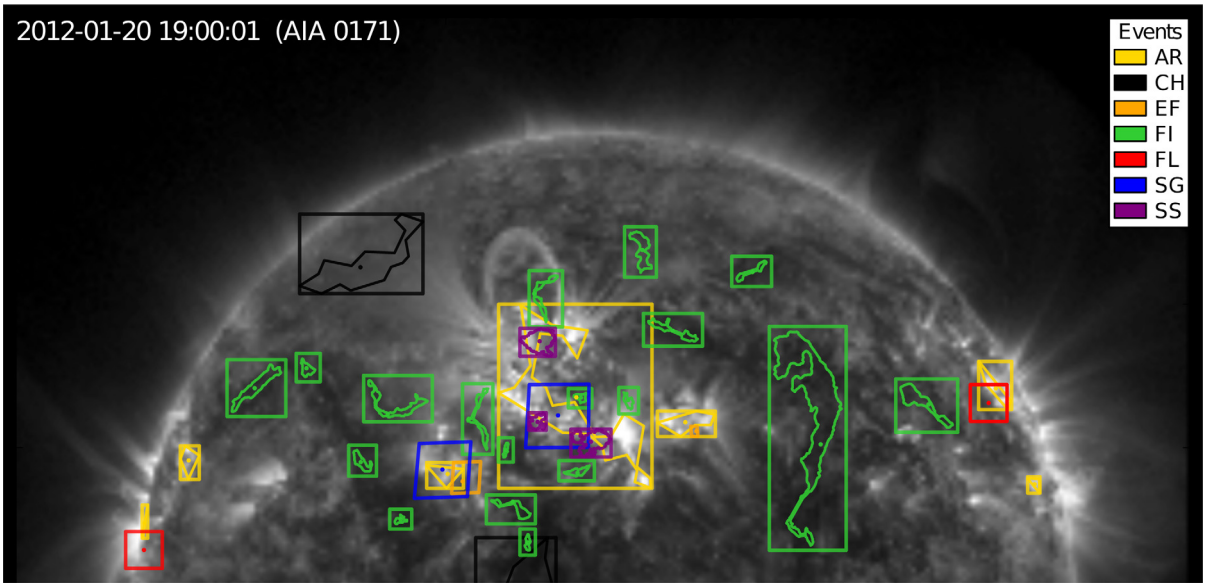

Fig. 1. All seven types of solar events co-occurring in time and overlaid on an AIA 171 Å image.

we perform them to remove bad records before beginning any other analyses which they may skew. We note that all records were successfully inserted into the database, and therefore all attribute values must already conform to the correct data type. Currently, four initial cleaning checks are tested:

1. Event start time and end time are within the data collection period.
2. Event start time is before or equal to the event end time.
3. Duration of an event report is less than two days.
4. Duplicate event reports.

We find a total of 12 event reports that violate the first two date-time checks, 8 event reports with a duration over two days, and another 6 event reports that are exact duplicates of others. Table 2 presents the totals before these events have been removed. While two days is an arbitrary limit for duration checks, we empirically observe most of the removed events have an impossible duration (greater than 78 days) given the nature of the phenomena and reporting modules. For proper large-scale analysis, we also chose to remove the 6 SG events reported in the "thick" channels due to the extremely small set size compared to other types of reports. In total, we retrieved 282,208 events, and after removing only 32 events (0.01%), we are left with a dataset of 282,176 events. While we don't further investigate these removed events, an approximate error/loss of only 0.01% is quite small and, as we will show next, reasonably negligible.

3.3. Reporting statistics

Now we investigate the reporting of modules and frequency of event reports. This provides perspective on the expectations of each module, as well as an easy assessment of standard operating conditions. First, we visualize the varying volume of total reports over time by event type. This can be seen in Figure 2, where we plot the total event instances reported for each unique timestamp of each event type over the entire time period. Note the variability across all modules, which is partly due to the nature of the solar phenomena as well as the methods of reporting. For example, given domain knowledge, we know that active regions and coronal holes are long-lasting and slowly evolving events with many present for long periods of time, while emerging flux and flares are relatively small and short-lived events that occur more sporadically.

In a similar manner, we can plot the time (in hours) between module reports for each event type, as shown in Figure 3. Again we see a wide variability across all modules, including their approximate operational start dates. Notice how this better shows the steady cadence of some modules (AR, CH, SS) vs. the sporadic (or triggered) reporting of other modules (EF, FL, FI), which again relates to the occurrence of the specific types of solar events being detected. In the case of FI events, the module typically only reports once or twice per day when H-α images are available at the two observatories listed in Table 2. This variability in reporting frequency is likely also affected by the 11-year solar cycle, which has periods of higher and lower solar activity, although it may not be apparent yet.

We also highlight possible module outages, indicated by yellow highlighted regions of time in Figure 3 if the difference between event reports exceeds 24 h (or 48 h for the FI module). Note the large reporting gap in the EF module, and the many sizable reporting gaps from the FI and SS modules. The most important matter here is the general trends in event module intervals caused by static reporting cadences and large outages that could affect the reliability of unreported (but otherwise expected) events through increased false negatives. Specifically, we can see that the AR and CH modules run at a 240 min cadence, and the SS module runs at a 360 min cadence – with much less than 1% of the intervals significantly longer or shorter than these durations.

In Figure 4, we present the aggregated box plots of the reporting counts and time intervals for each event type. This offers an excellent comparative look across modules and helps highlight any trends and possible concerns. Note that unless otherwise specified, we use standard box plots, which feature a rectangular box encompassing all values between the 1st quartile (Q1) and 3rd quartile (Q3), which are the 25th and 75th percentiles of the data, respectively. Vertical dotted lines

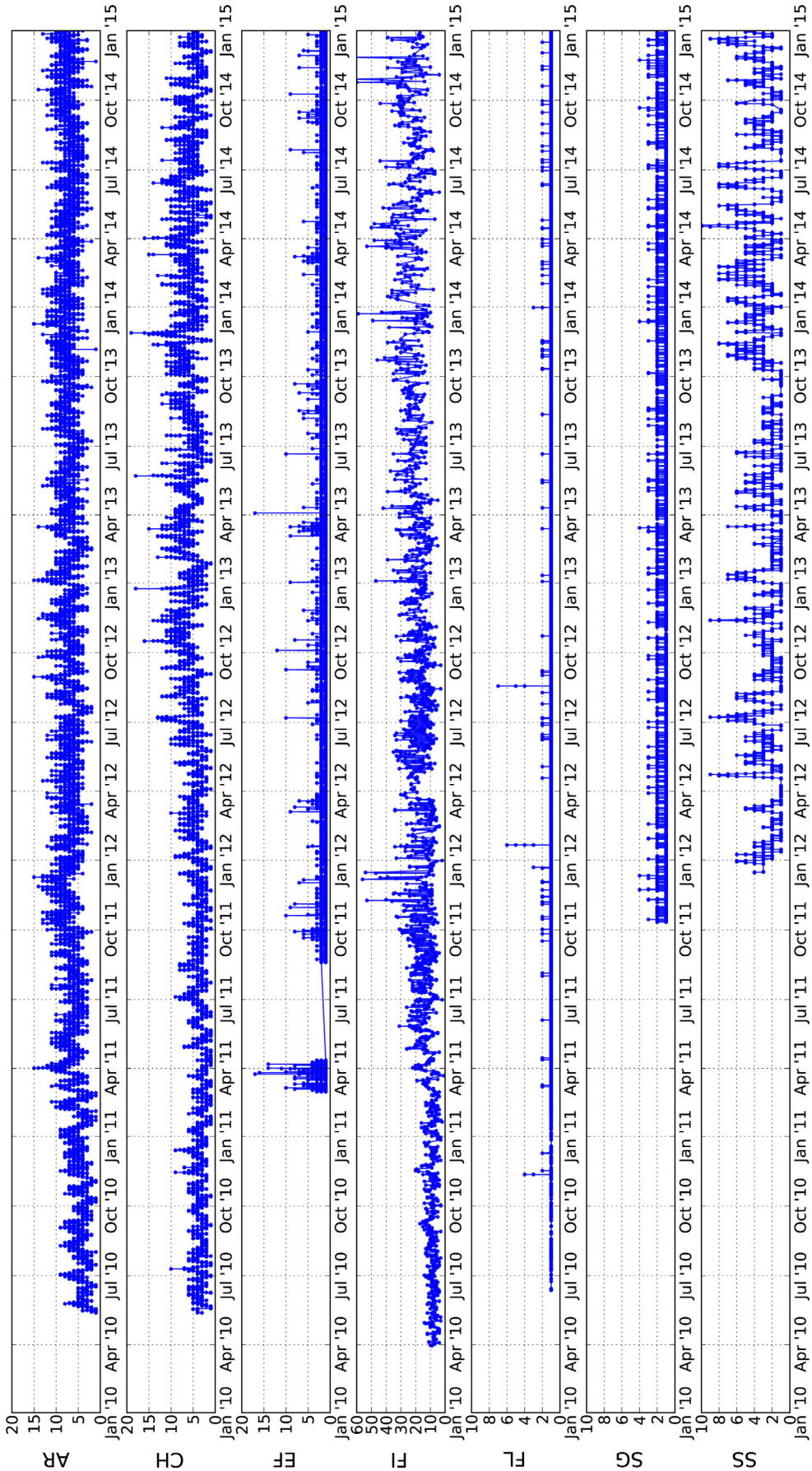

Fig. 2. Frequency of total event reports at each unique reporting time for each event type.

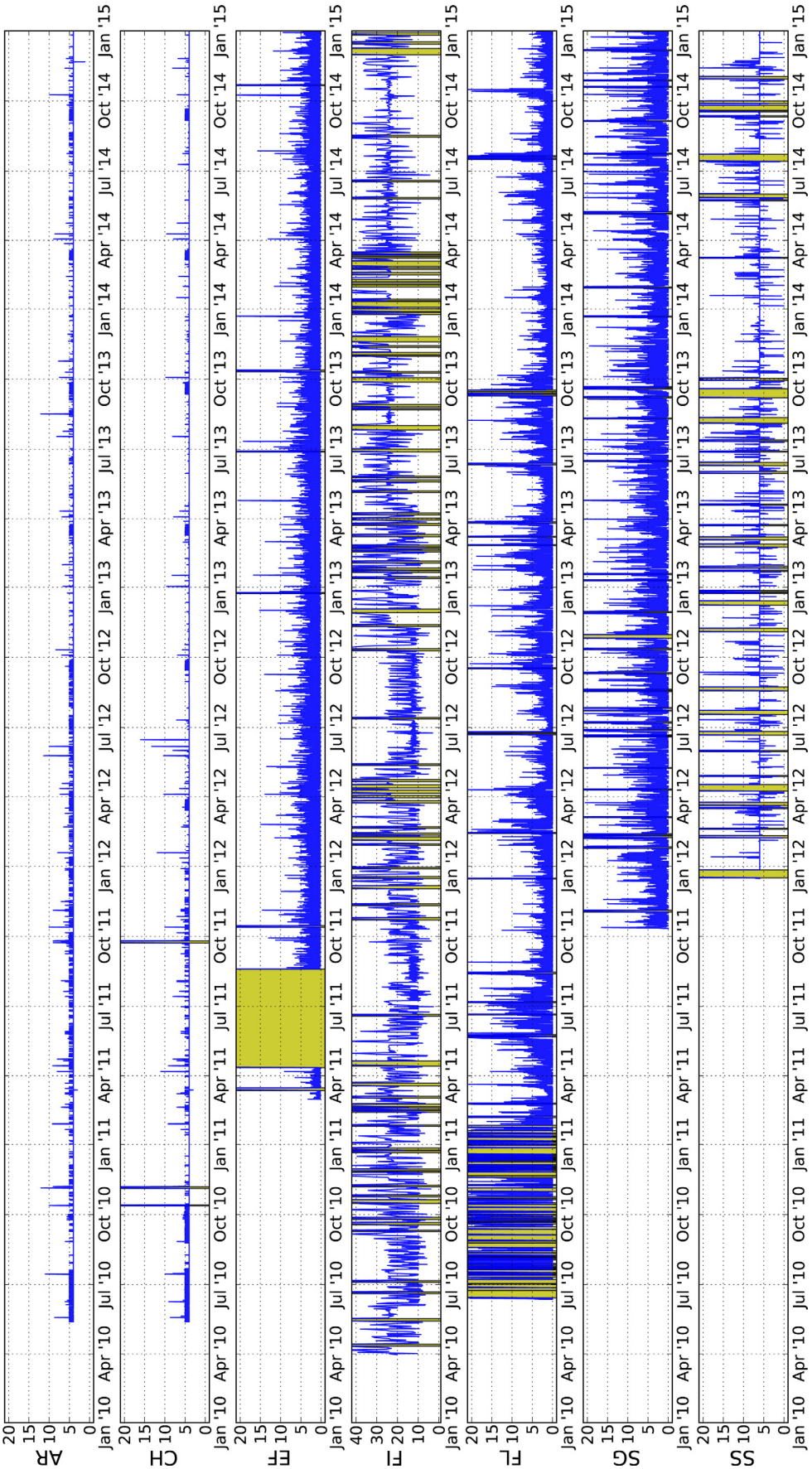

Fig. 3. Time (in hours) between each unique reporting time for each event type. A yellow highlighted box indicates a reporting gap greater than 24 h, which may suggest a data or module outage.

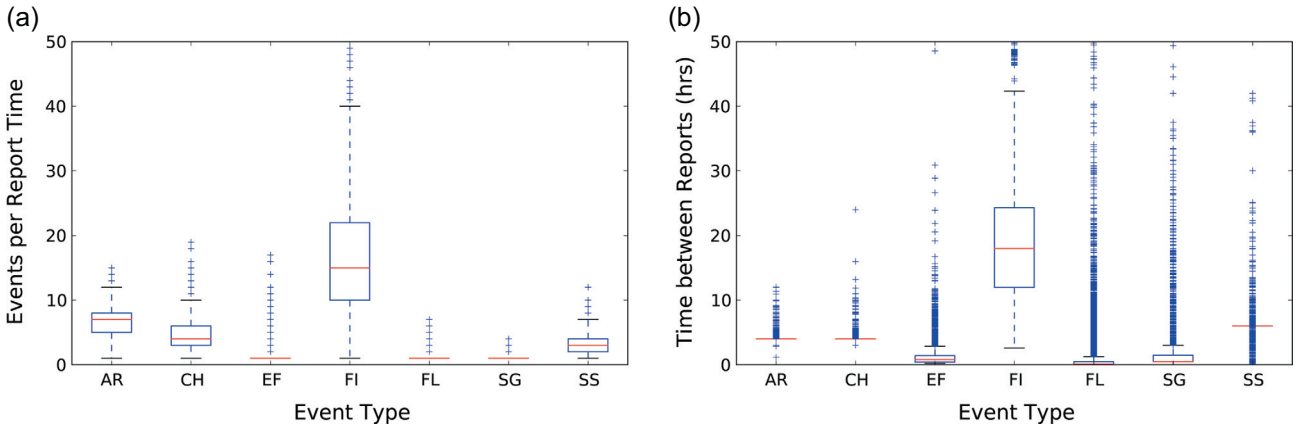

Fig. 4. Box plots of (a) the total events at each unique reporting time and (b) the interval (in hours) between module reports.

Table 3. A summary of reporting statistics for each event type, including the number of reporting gaps greater than 24 h and their cumulative percentage of the total module operational time.

Event	Start date	Total reports	Total times	Total gaps	Gap time (%)
AR	2010-05-13	65,562	9,856	0	0.00
CH	2010-05-13	47,960	9,922	3	0.32
EF	2011-03-01	31,076	26,487	7	10.01
FI	2010-03-30	33,443	1,994	84	10.32
FL	2010-06-11	71,903	71,759	80	11.70
SG	2011-10-11	20,716	18,790	35	4.22
SS	2011-12-15	11,525	3,904	37	11.70

(with solid horizontal-lined end caps) extend 1.5 times this interquartile range (IQR), which is the difference between Q3 and Q1. Beyond the whiskers, individual data outliers are plotted as "+" marks, and a horizontal line within each IQR box is used to indicate the mean data value. Here we can clearly see that nearly all EF, FL, and SG events are reported individually, i.e., at unique start times. While FI events are reported in significantly larger groups than all other event types, there exist no concerning outliers for any event types. Furthermore, we now clearly see the systematic 240 minute cadence of the AR and CH modules and the 360 min cadence of the SS module (with slightly more variance). Again, the FI module is the most unlike the others, with many intervals well beyond the 50-hour y-axis limit. Here again, while there exist outliers beyond the box plot whiskers for all event types, many can likely be attributed to the natural frequency of the solar phenomenon.

In Table 3, we provide a summary of event reports, including the total number of reporting gaps (possible outages) greater than 24 h for each event type (48 h for FI). Again, we can see that some modules never (or rarely) go longer than a day without reporting. We also show the module start date, total event reports, and total unique report times. The gap time is a percentage of total operational time missed based on start date and summed time of all indicated gaps (minus the mean reporting interval time that would be otherwise expected from these reports). These can be important metadata statistics about the overall likelihood of module reporting tendencies, possible lack of module reports over a period of time, and general frequency of types of events present at any given time. For example, we see that the EF module has seven gaps accounting for approximately 10.01% of lost time since HEK reporting began on May 1, 2011. In this case however, we note that the majority of this time is in the initial large gap, which we suspect was after an initial test run of the module due to the significantly different reporting frequencies seen after the outage.

Table 4. A summary of total event reports for varying module version numbers over their operational time periods.

Event type	Module version	Start date	End date	Reports
AR	1	2010-05-13	2014-12-31	65,562
CH	1	2010-05-13	2014-12-31	47,960
EF	0.15	2011-03-12	2011-04-11	2,131
EF	0.27	2011-08-19	2011-08-24	89
EF	0.29	2011-08-24	2014-12-31	27,642
EF	1	2011-03-01	2011-03-12	1,214
FI	5.5	2010-03-30	2010-04-15	111
FI	5.7	2010-04-15	2010-05-05	204
FI	6	2010-05-06	2010-05-07	13
FI	6.1	2010-05-08	2010-10-27	1,932
FI	6.3	2010-10-28	2014-12-29	31,183
FL	0.5	2010-06-11	2010-12-05	585
FL	0.51	2010-11-04	2014-12-31	71,316
SG	1	2011-10-11	2014-12-31	20,709
SS	0.2	2011-12-15	2014-12-31	11,525

In an effort to explore the large gap in the EF module reports, and our possible explanation of module testing or calibration issues, we also include the code versioning attribute provided by each module in each event report. Shown in Table 4, we present the earliest and latest event start dates for each unique module version number over all data and event types. As expected, we can see several version changes in the EF module around the time period in question (March 2011). We can also see several version changes in other modules at varying times of operation. While we do not discard any of these event reports for this study, we note that advanced analyses may need to account for these module changes and possible effects or biases imparted on the reports. Therefore, we include the module version number as an event-specific attribute for each report in the provided dataset.

Lastly, for a broader scope and trend over time, total reports for each event type are aggregated over each month and shown in Figure 5. This best indicates the comparative report volumes over time, and again the great variability among modules. We note the FL module appears to be most

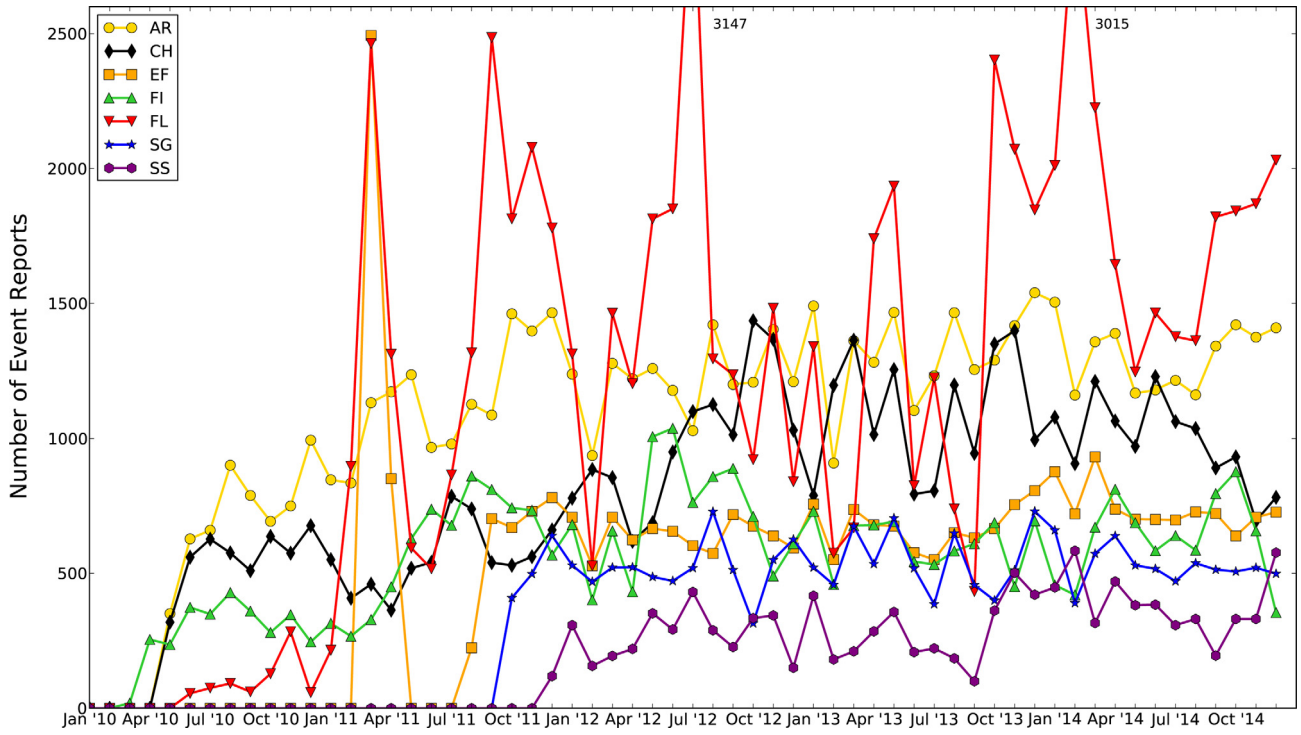

Fig. 5. The total event reports for each event type aggregated over each month of operations.

volatile, but given the large number of sources for which flare events are reported on, a general increase in solar activity could be compounded here more so than with other types of events. In other words, a single flare may be associated with multiple event reports due to each AIA channel it was reported in. Also note the large spike in EF events before the large gap previously mentioned.

3.4. Spatial and temporal attributes

Next we look at the spatial and temporal attributes provided in each event report, which are standard across all modules. Spatial locations of events are available in a variety of world coordinate systems (Thompson 2006) to the HEK, and we use the Helioprojective Cartesian (HPC) coordinate system in arcseconds from disk center for all plots and discussions. While measurements in this Earth-centered system can be affected by Earth's orbit, we found it is used as the default location attributes of reported events and therefore chose to use it for our analysis herein. Each event is required to have a center (x, y) point and a bounding box, which is a closed polygon of five points that represents a (not necessarily minimal) bounding rectangle. Optionally, an event may include a chain code, which in this context is simply an arbitrary number of sequential coordinate points that represents a detailed polygonal outline of the event boundary. We include all five-statistic data summaries (min, mean, median, max, standard deviation) for these attributes in Table A1.

In Figure 6a, we show box plots of the total number of points in the chain code polygons for each event type. We note that all EF, FL, and SG events do not have chain codes, and all other event types do, except for only a few outlier reports (16 AR and 4 CH). A clear observation here is the significantly larger number of points used for SS chain codes compared to others, especially since the SS events are relatively much smaller than an average CH event. Note the break in the y-axis (and rescaling) to better represent both extremes in a single plot. Shown in Figure 6b, we can estimate the varying relative event size through the derived areas of the bounding boxes as box plots (width times height in arcsec) for each event type. Note however, this does not account for the area of the box strictly occupied by the event, which may be deceptive for long and narrow events, such as many filaments or near-limb coronal holes. As a point of reference (verified later), the mean area of FL events shown here is approximately 1/1,024 of the total image.

We now look at the duration of events provided by their start and end time, shown as box plots in Figure 7. Much like the intervals between module reports, event durations are partly characterized by the module tendencies as well as the specific solar phenomenon being reported on. We note that FI events are always given an instantaneous time (zero duration), SG events are always given a 30 min duration, SS events are always given a 360 min duration, and AR and CH events are almost always given a 240 min duration. Clearly these are reporting conventions and not indicative of solar phenomena characteristics. Also note that the spatial coordinates given for an event report can only occur at one moment in time, although we note the bounding box may by chance encompass the event for longer. Therefore, the duration attribute derived from these event reports is not a very valuable statistic for most modules or applications.

We begin looking at spatial attributes by presenting box plots of the center location values (in HPC coordinates) of all events for each event type in Figure 8. We limit the axis to ±1,250 arcsec, and note that all event centers are within this valid range. It is also clear to see from the IQR that there is much variation in center locations across different event types. Next we look at 2D scatter plots of the location coordinates (in HPC) provided by each event report. These plots are tremendously beneficial not only for data validation and cleanliness checks, but also for module reporting biases and event location

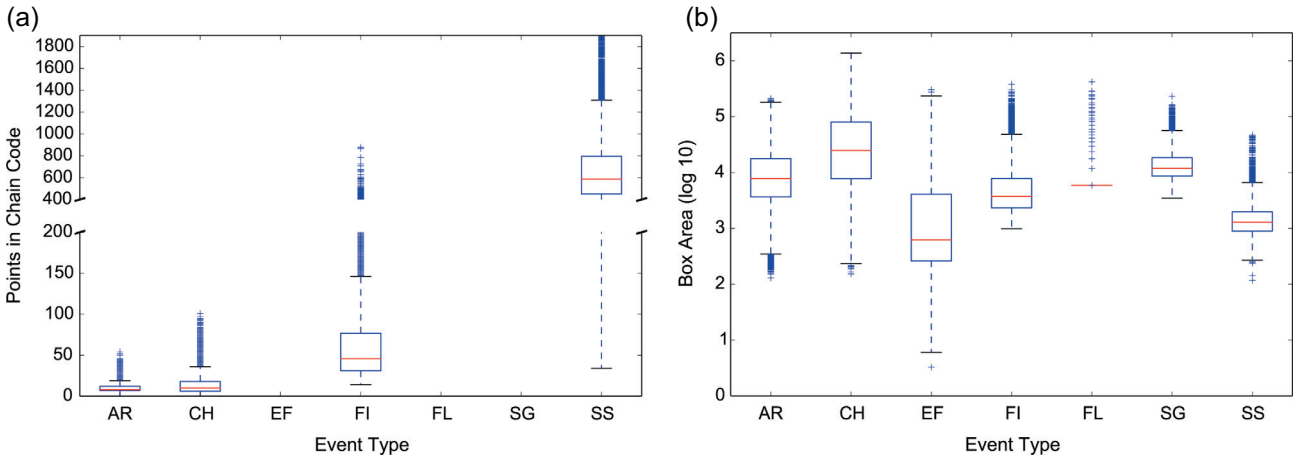

Fig. 6. Box plots for each event type of (a) the number of chain code points and (b) the derived bounding box areas in arcseconds squared.

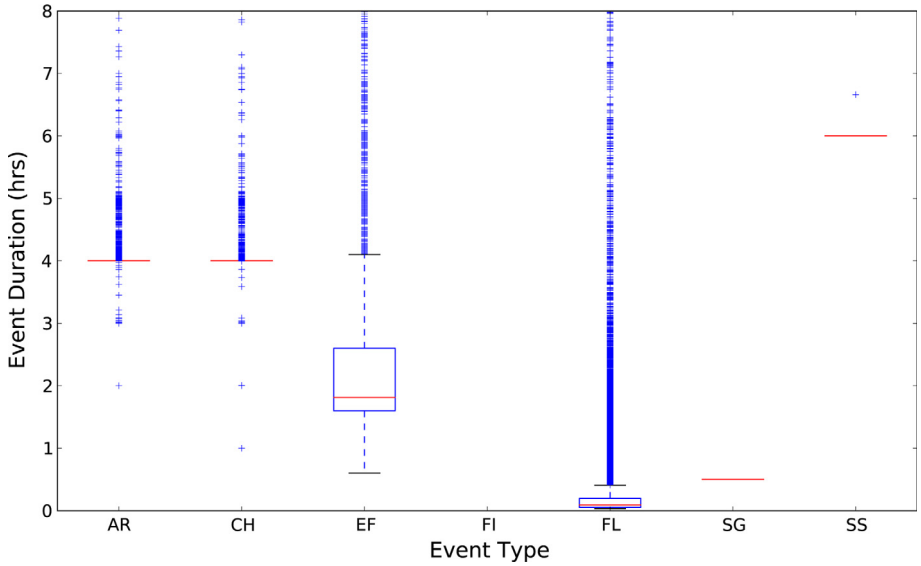

Fig. 7. Box plots for each event type of the event duration.

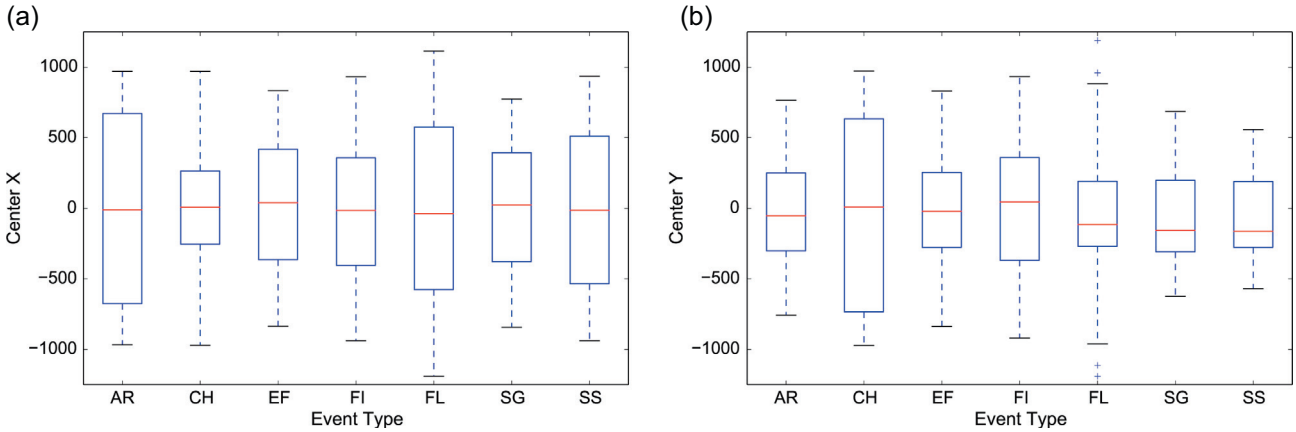

Fig. 8. Box plots for the locations (in HPC coordinates) of each event type, split into (a) center x-axis and (b) center y-axis arcsecond values.

likelihoods. Here we again standardized all plot axes limits to ±1,250 arcsec, and we also placed a circle at the center (0, 0) with an arbitrary radius of 1,000 arcsec from disk center, which is slightly larger than the solar disk radius at any point during yearly observation. In Figure 9, we see the centers, bounding boxes, and chain codes plotted for all AR events over the entire time period. While the centers are the most informative for exact event location occurrences, by visualizing the bounding boxes and chain codes we quickly and easily verify there are no off-disk outliers or anomalies.

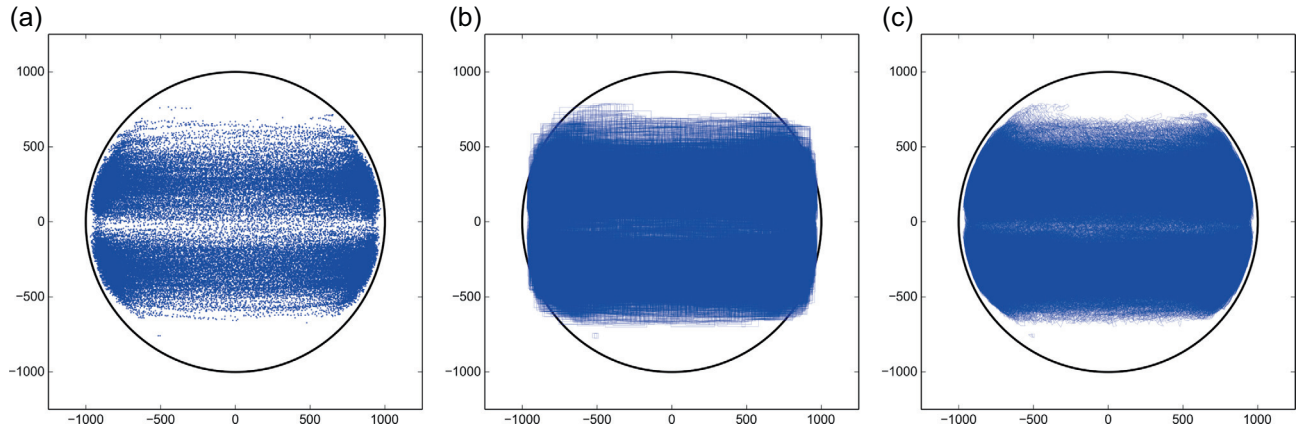

Fig. 9. The spatial locations of AR events based on their: centers (a), bounding boxes (b), and chain codes (c) in the HPC coordinate system using arcseconds.

In Figures 10 and 11, we show (from left-to-right) the same three 2D spatial plots of event centers, bounding boxes, and chain codes of all other event types. From top-to-bottom, Figure 10 shows CH, EF, and FI events, and Figure 11 shows FL, SG, and SS events. If an event type does not have chain codes, we still include the blank plot for presentation consistency. These visual aids make cross-event comparison much easier and help showcase a potentially useful set of statistics for event co-location probabilities. For example, notice clearly vacant areas of the solar disk for CH events and well-known bands of activity for AR and SS events. Also notice that FL events are identified over a discretized grid of 32×32 image cells (refer back to Fig. 1 for examples). While this is by design (Martens et al. 2012), we point out it was originally stated to be only 16×16 image cells. Regardless, this means that FL events are only reported from a finite set of known location points, some of which are off-disk due to solar limb flares. While a density map could shed further light on the frequency of these reported locations, we can preliminarily see that they all conform to the expected image cell grid. Lastly, we see a hint of possible symmetry (reflection about the origin) in the lower-left and upper-right of the bounding boxes of EF and SG events. While this could be a module bias or phenomena-specific characteristic, it is unclear from our initial findings.

3.5. Event-specific attributes

Next we take a detailed look at all event-specific attributes included in the FFT module reports. According to the HEK API documentation,[6] each type of event can have numerous optional attributes for characterizing that specific event type (Hurlburt et al. 2012). Unfortunately, real-world data does not always conform to documented standards, so here we first focus on what attributes are actually available and what sort of data they provide. We also point out several attributes that can be used to distinguish event sub-types, such as the chirality of a filament event.

Each type of event has an optional section in the reported XML files that includes all available optional attributes for that event type. We note that the alternative JSON-formatted files that we use to collect events do not have this hierarchical attribute partitioning, so all fields are first noted manually by XML files and then automatically extracted from the QHEK-retrieved JSON files. We also manually check events periodically over time to ensure they still have the exact same set of attributes present. A succinct overview of these attributes can be found in the event table SQL files provided with the dataset, as well as the data summary tables in the Appendix.

3.5.1. Null values

The first step is to take note of any empty (or null) values found in a report. We find that except for 21 consecutive SS events which supply no event-specific attributes, only the intensity attributes for AR and CH events are sporadically null-valued. Specifically, of the 65,562 AR events, only 4,602 (7.01%) are missing these values, and of the 47,960 CH events, 17,357 (36.19%) are missing these values. Note that all of these events are missing all of the various intensity attributes, so further investigation could be made as to the cause of this. No other optional attribute is ever missing from any other event report over the entire time period.

3.5.2. Real values

We ignore all null values from our statistical calculations, and again include the 5-statistic data summaries of each attribute in Table A2. We group attributes together that are common across multiple event types and present box plots of these attributes for easier comparison, examples of which can be seen in Figure 12. This beneficial cross-comparison can be seen quite clearly in Figure 12a, where we show the mean intensity ("intensmean") for AR and CH events, which by the nature of the bright and dark event types, respectively, should be as different as the data appears to indicate. In Figure 12b, we show the normalized event areas calculated at disk center ("area_atdiskcenter"), affirming the findings of our prior bounding box area estimations (Fig. 6b). Given the lengthy list of possible event attributes, all plots in this section use general-purpose labels and axes for automated creation and first-look analysis.

We can also visualize attribute-specific histograms segmented into subsets of the entire time period to more easily see any trends that may exist within the overall distributions. Here we arbitrarily segment the data by calender year (five subsets in total) and show several interesting examples in Figure 13. Notice how easily we can see significant distribution differences over each time period for the mean intensity of AR and CH events. Also included are the axis orientation of EF events and tilt angle of FI events, which both show much more consistent distributions over time.

[6] http://www.lmsal.com/hek/api.html

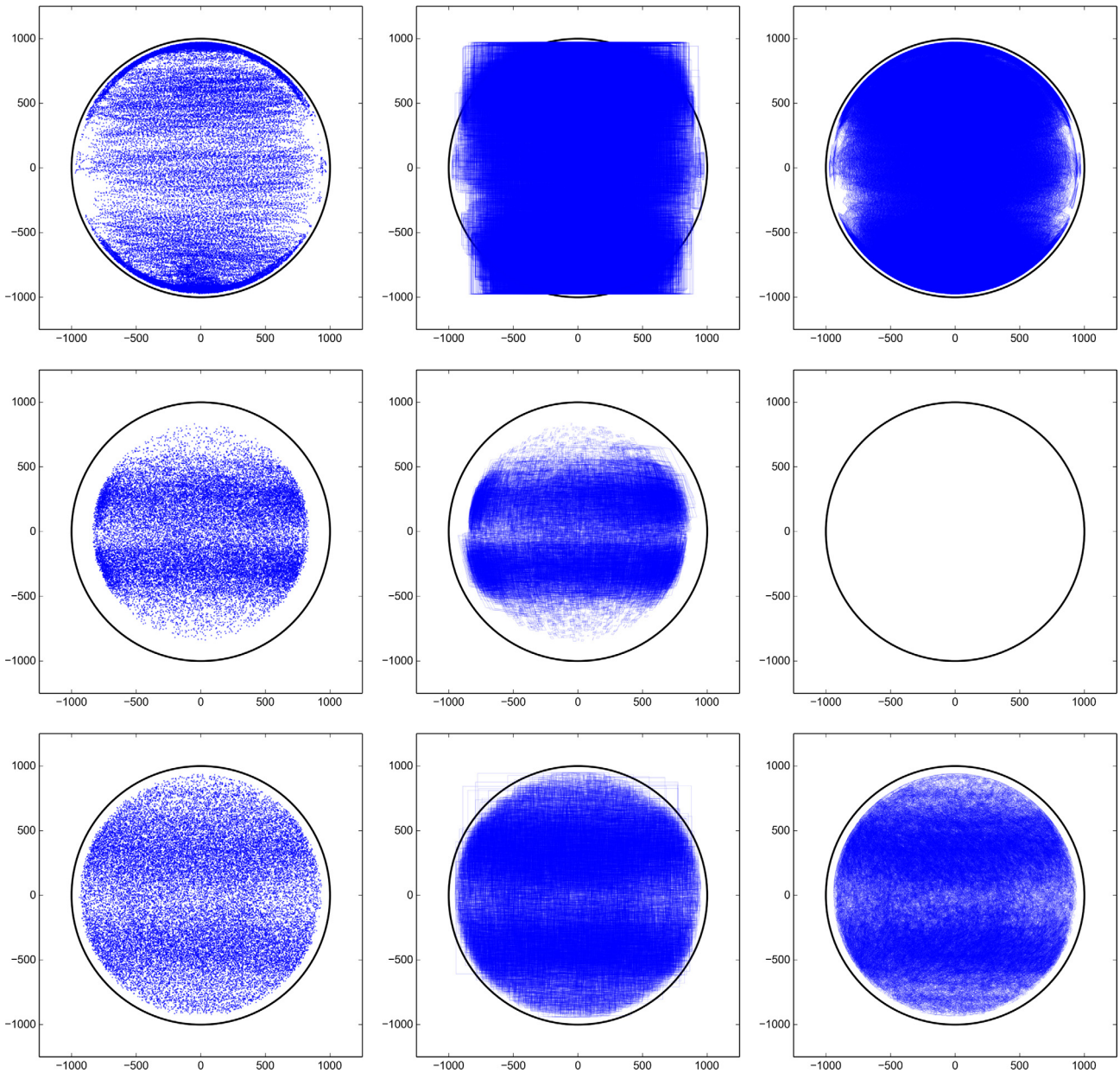

Fig. 10. The spatial locations (from top-to-bottom) of CH, EF, and FI events based on their center, bounding box, and chain code attributes (left-to-right) in the HPC coordinate system using arcseconds.

3.5.3. String attributes

We briefly mention the existence of several string-based attributes for each event type as well. In Table A3, we present the list of attributes and the string value observed for each one. As we can see, these attributes are used as unit descriptors for corresponding numerical attributes, and almost all records contain the same string value across all types of events. The only major exception to this is the FI "event_pixelunit" attribute that provides the exact numerical arcsec/pix value for each event report, which is unlike all other event types that have this same attribute. We also found that a small subset (less than 1%) of FI events report the length attribute ("fi_lengthunit") in arcsec instead of cm units. While we retain these events for completeness, they are omitted from the calculation of the FI length summary statistics. Another minor discrepancy we discovered is that while the EF axis orientation ("ef_axisorientationunit") is listed as units in degrees, the data values are actually provided in radians.

3.5.4. Event sub-types

Two event types already contain attributes for explicit sub-class differentiation performed by the automated FFT detection algorithms. Seen in Table 5, these are the FI chirality ("fi_chirality"") and SG shape ("sg_shape") attributes. Note that we discretized the SG shape attribute from the two string values ("Inverse-S Sigmoid" and "Forward-S Sigmoid") to binary integer values (−1, 1) for easier use as labels. The chirality label values are already integers, corresponding to sinistral (−1), neutral (0), and dextral (1) orientation. We can see here that the majority of FI events have neutral chirality, and the other FL and SG event labels are much more evenly split.

In Table A4, we list other general-purpose attributes that belong to more than one event type. These attributes can aid in intra-class (event sub-type) distinction as well as overall event classification efforts. For example, by simply looking at the total area and/or location of an event, we can make

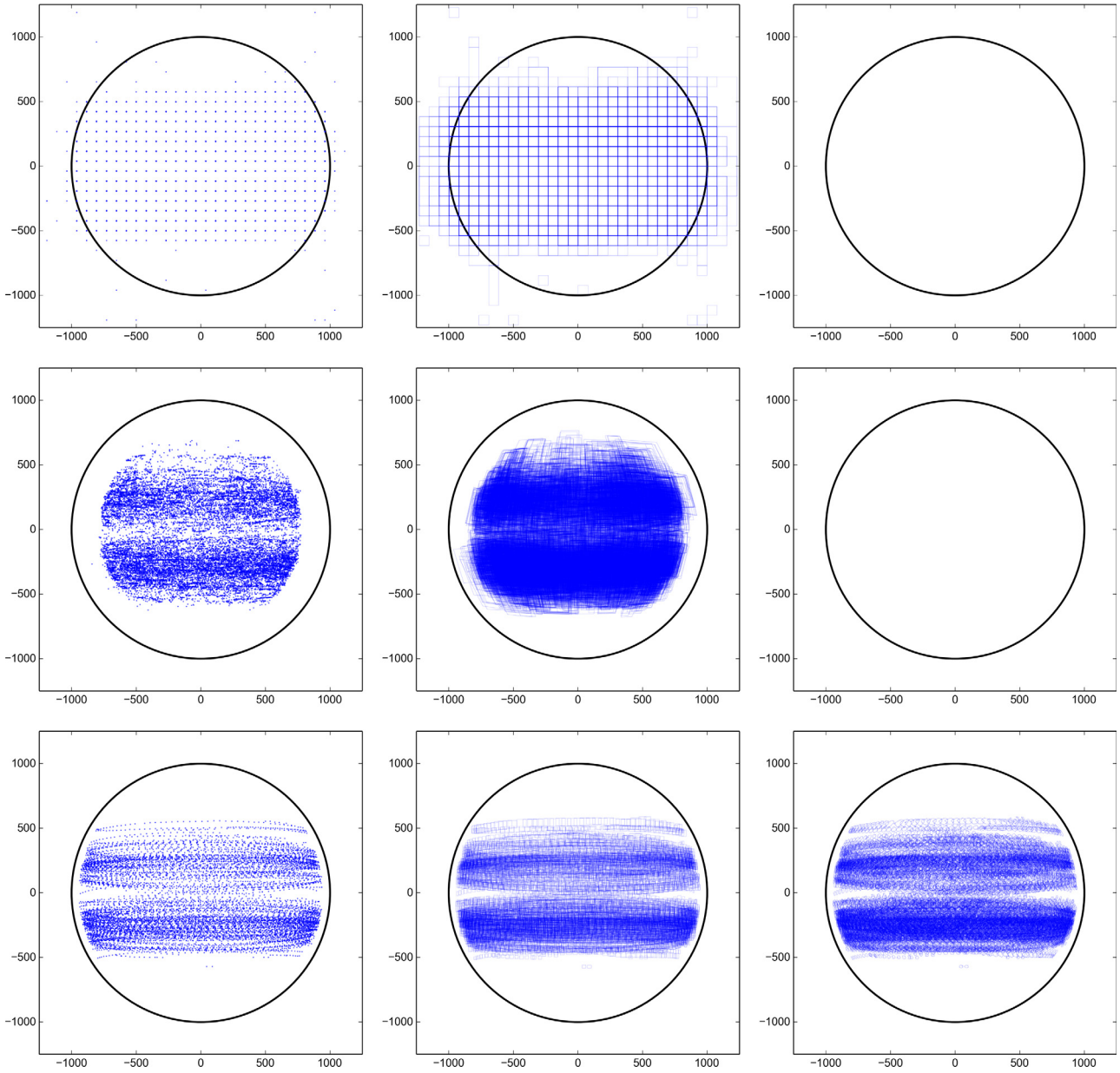

Fig. 11. The spatial locations (from top-to-bottom) of FL, SG, and SS events based on their center, bounding box, and chain code attributes (left-to-right) in the HPC coordinate system using arcseconds.

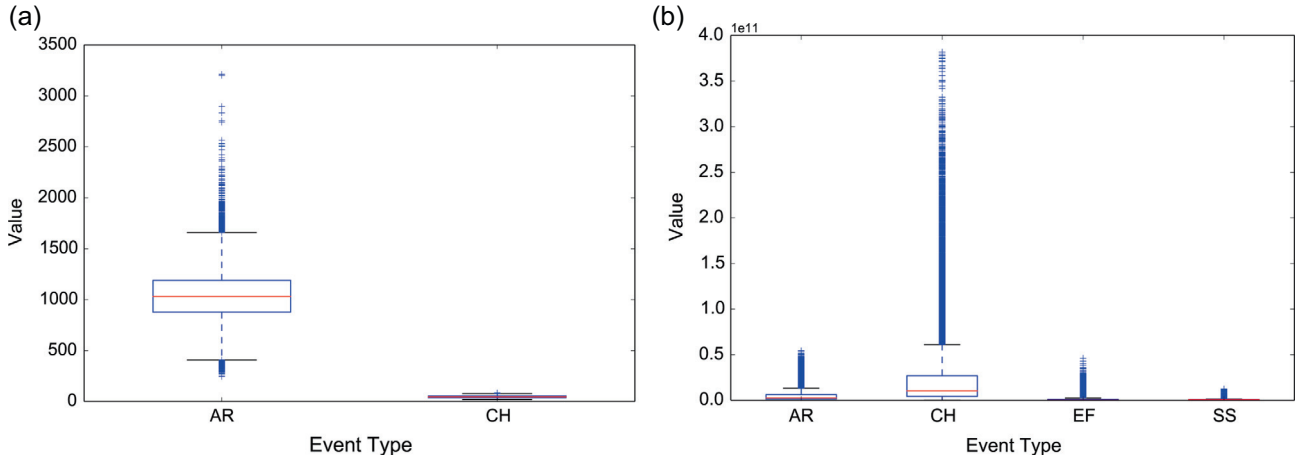

Fig. 12. An example of box plots grouped by event type for mutual event-specific attributes. (a) Intensity mean, (b) area at disk center.

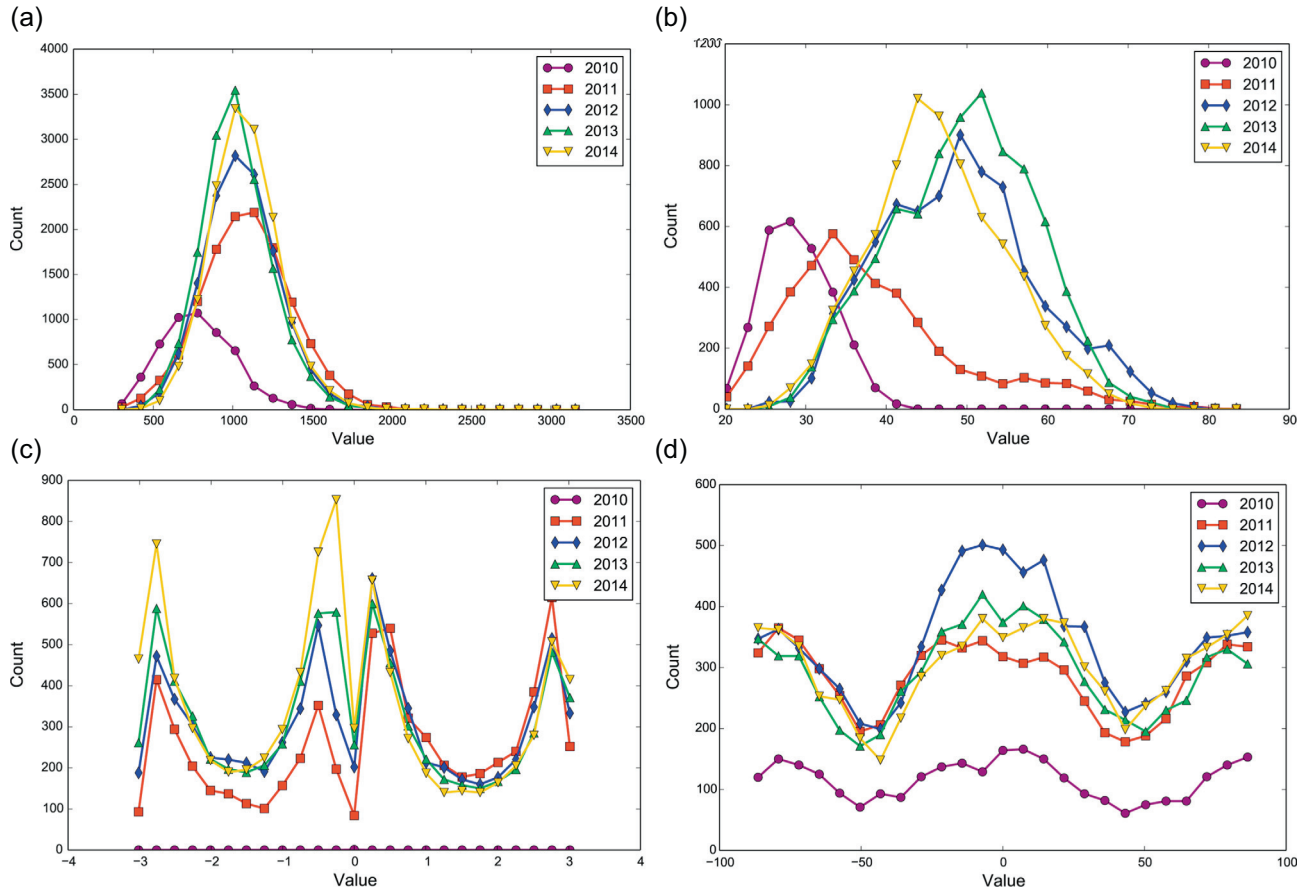

Fig. 13. An example of time-segmented histograms for event-specific attributes. (a) AR intensity mean, (b) CH intensity mean, (c) EF axis orientation, (d) FI tilt angle.

Table 5. A summary of class-valued attributes.

Event	Attribute	Sub-type	Reports	Percent (%)
FI	Chirality	Sinistral	6,100	18.24
		Neutral	21,719	64.94
		Dextral	5,624	16.82
SG	Shape	Inverse	11,018	53.19
		Forward	9,698	46.81

reasonable assumptions on what type of event it is. In the future, we could perform machine learning with clustering and classification algorithms on these event attributes to try to characterize well-separated sub-classes within each event type. An important example of this is the flare peak flux value ("fl_peakflux"), which should correlate to existing defined classes (energy magnitudes) of FL events. However, this correlation requires a cross-module calibration beyond the scope of this current work.

3.6. Dataset dissemination

As discussed, the point of this paper is to explore all available SDO FFT module data available through the HEK. To remain as transparent and reproducible as possible, we also provide the entire dataset that was used and analyzed in this work on our website[7] and through the Zenodo service, which provides persistent third-party URLs and DOIs for reference. This supporting dataset (Schuh et al. 2016) is available in raw text-file format, as well as the database tables and records described above. It is our intention that this dataset be considered the de-facto collection of all current SDO FFT module data available through the HEK. This provides the community a ready-to-use dataset for research that will be far more verifiable and reproducible by others.

Through these preliminary investigations, several event reports were cleaned or removed from analysis, but we note that this represents an extremely small fraction of the total event reports, and therefore, indicates a reasonably clean dataset overall. While future investigations may find more nuanced reporting errors or questionable data anomalies, these initial analyses and data validations are crucial first steps toward proper dataset curation and application.

4. Data-driven analysis

Finally, in this last section we investigate several more advanced data analyses. This offers an initial look into the possibilities of using broad large-scale multitype event reports to push data-driven knowledge discovery and strengthen theoretical hypotheses.

One important concern is the reporting of events from multiple sources of data using the same FFT module, as is the case for FI, FL, and SG events. For example, in Figure 14 we plot the total event reports per unique reporting time for each separate source of FI events. Here we can confirm that both sources of data are reported in an interleaved fashion (which is known

[7] http://dmlab.cs.montana.edu/solar/

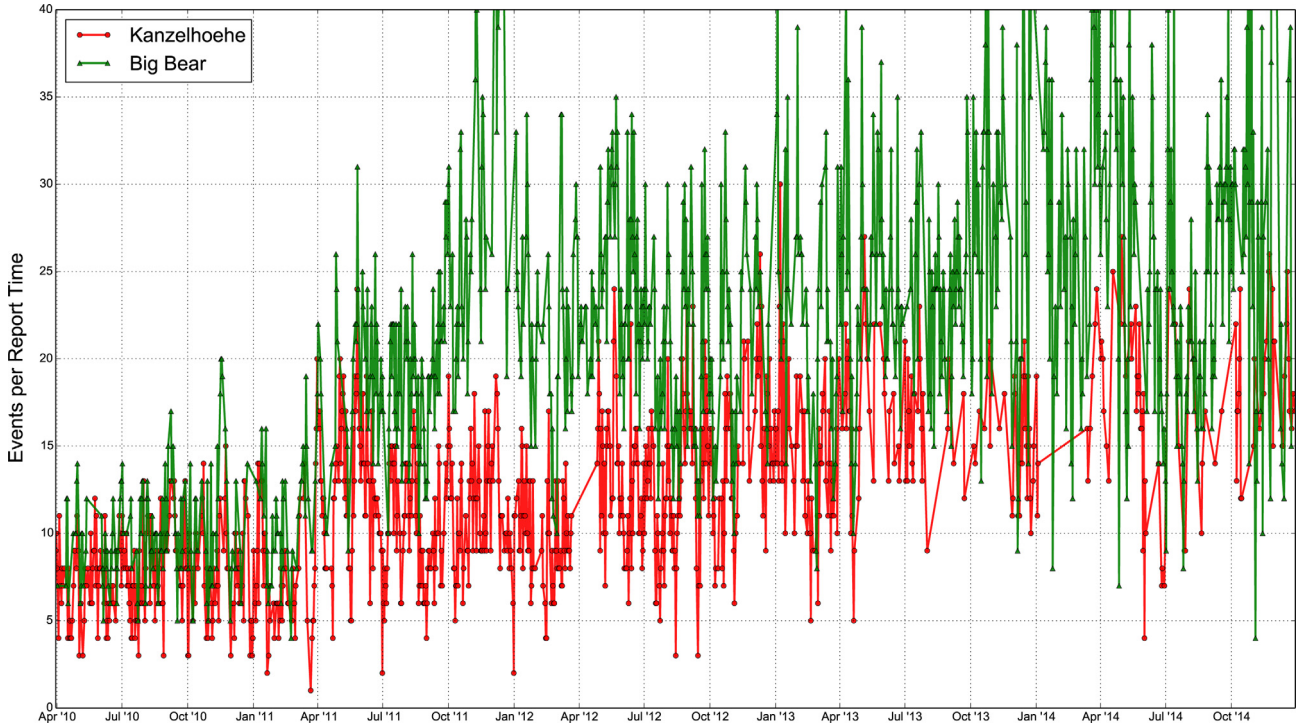

Fig. 14. The report counts of FI events for each distinct observatory.

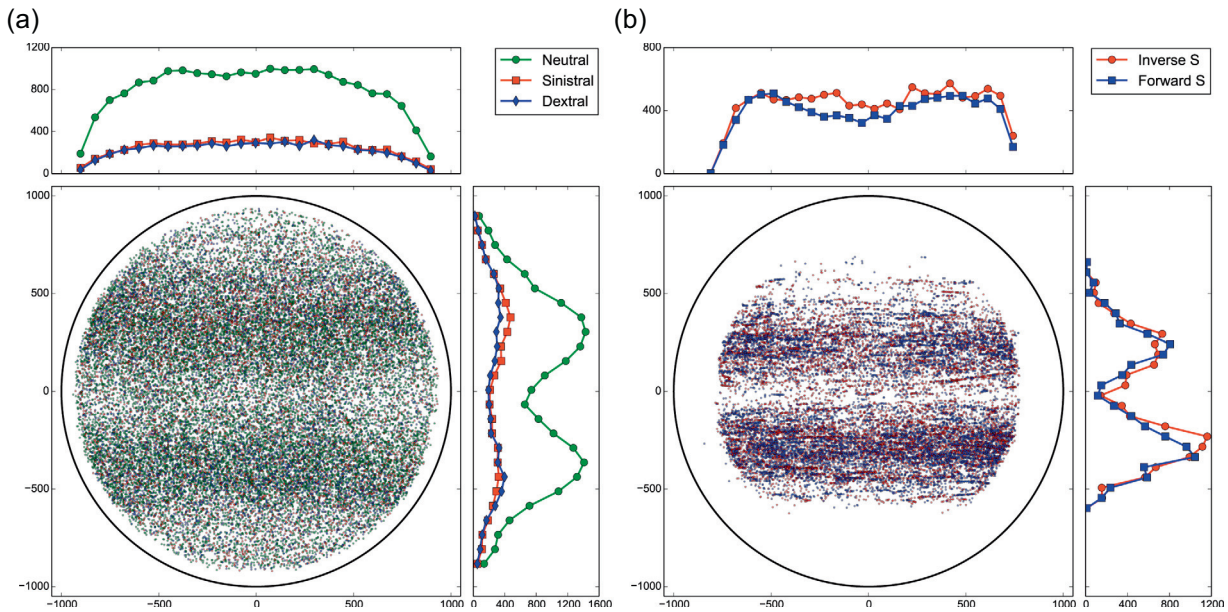

Fig. 15. The classwise breakdown of spatial center locations in HPC arcseconds for (a) FI events and (b) SG events with accompanying single-dimensional histograms.

based on the observatory data availability), but we can also see a general trend difference in the overall frequency of FI events reported at each independent source. While this requires deeper investigation, it could be a result of instrumentation quality and the effects on the detection algorithm. Note this is the only event type reported on two independent observational sources, and therefore follow-up analyses may benefit from comparing other FI detection modules as well. Similar charts are available for FL and SG events, but they are far less informative. As we noted earlier, FL and SG events are reported on multiple AIA channels at almost always unique times. So while these reports are not exact duplicates (in the sense of data records in the dataset), many could likely be reporting on the same phenomenon in a very similar space and time.

As a final and most interesting application of this data, we again plot the spatial locations of events, but this time we categorize the events by sub-class attributes described in the previous section. In Figure 15, we present the color-coded center locations for (a) FI events and (b) SG events in HPC coordinates, again with an artificial circle radius of 1,000 arcsec. Here we also include the single-dimensional histograms split by class value on each axis of both scatter plots to better convey the frequency of each event class. Although not overly prominent, we do see a slight hemispheric inclination – especially in

filament chirality, which has been well known to the community (Pevtsov et al. 2003). It will be of great interest to see these trends extended over larger spans of time and more complex correlation analyses.

5. Conclusions

This work presented a comprehensive overview of all SDO FFT event module reports publicly available through the HEK. We provided background information about the volume and velocity of the various data sources, as well as a detailed outline of the data collection, cleaning, and analysis processes. This work serves as the foundation for using and trusting this source of large-scale solar event data. By providing ready-to-use datasets to the public, we hope to interest more researchers from various backgrounds (computer vision, statistics, machine learning, data mining, etc.) in the domain of solar physics, further bridging the gap between many interdisciplinary and mutually-beneficial research domains. In the future, we will maintain extended and up-to-date datasets and statistics online and refer to this work as reference.

Acknowledgements. This work was supported in part by two NASA Grant Awards (No. NNX11AM13A and No. NNX15AF39G) and one NSF Grant Award (No. AC1443061). The NSF Grant Award 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. The editor thanks David Pérez-Suárez and an anonymous referee for their assistance in evaluating this paper.

References

Aydin, B., D. Kempton, V. Akkineni, R. Angryk, and K. Pillai. Mining spatiotemporal co-occurrence patterns in solar datasets. *Astron. Comput.*, **13**, 136–144, 2015, DOI: 10.1016/j.ascom.2015.10.003.

Banda, J.M., R. Angryk, et al. Selection of image parameters as the first step towards creating a CBIR system for the solar dynamics observatory. In *Digital Image Computing: Techniques and Applications (DICTA) 2010 International Conference on*, IEEE, 528–534, 2010, DOI: 10.1109/DICTA.2010.94.

Banda, J.M., M.A. Schuh, R.A. Angryk, K.G. Pillai, and P. McInerney. Big data new frontiers: mining, search and management of massive repositories of solar image data and solar events. In *New Trends in Databases and Information Systems*, Springer International Publishing, Cham, 151–158, 2014, DOI: 10.1007/978-3-319-01863-8_17.

Bernasconi, P.N., D.M. Rust, and D. Hakim. Advanced automated solar filament detection and characterization code: description, performance, and results. *Sol. Phys.*, **228 (1–2)**, 97–117, 2005, DOI: 10.1007/s11207-005-2766-y.

Council. *N. R. Severe Space Weather Events – understanding societal and economic impacts: a workshop report*. The National Academies Press, 2008, DOI: 10.17226/12507.

Hurlburt, N., M. Cheung, C. Schrijver, L. Chang, S. Freeland, et al. Heliophysics event knowledgebase for the Solar Dynamics Observatory (SDO) and beyond. In *The Solar Dynamics Observatory*, Springer, 67–78, 2012, DOI: 10.1007/978-1-4614-3673-7_5.

Kempton, D., and R. Angryk. Tracking solar events through iterative refinement. *Astron. Comput.*, **13**, 124–135, 2015, DOI: 10.1016/j.ascom.2015.10.005.3.1.

Kempton, D., K. Pillai, and R. Angryk. Iterative refinement of multiple targets tracking of solar events. In *2014 IEEE International Conference on Big Data (Big Data)*, 36–44, 2014, DOI: 10.1109/BigData.2014.7004402.

Lemen, J., A. Title, D. Akin, P. Boerner, C. Chou, et al. The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Sol. Phys.*, **275**, 17–40, 2012, DOI: 10.1007/s11207-011-9776-8.

Martens, P., G. Attrill, A. Davey, A. Engell, S. Farid, et al. Computer Vision for the Solar Dynamics Observatory (SDO). In: P. Chamberlin, W.D. Pesnell, and B. Thompson, Editors. *The Solar Dynamics Observatory*, Springer, 79–113, ISBN: 978-1-4614-3672-0, 2012, DOI: 10.1007/978-1-4614-3673-7_6.

Pesnell, W., B. Thompson, and P. Chamberlin. The Solar Dynamics Observatory (SDO). *Sol. Phys.*, **275**, 3–15, 2012, DOI: 10.1007/s11207-011-9841-3.

Pevtsov, A.A., K. Balasubramaniam, and J.W. Rogers. Chirality of chromospheric filaments. *Astrophys. J.*, **595 (1)**, 500, 2003, DOI: 10.1086/377339.

Pillai, K.G., R.A. Angryk, J.M. Banda, T. Wylie, and M.A. Schuh. Spatiotemporal co-occurrence rules. In *New Trends in Databases and Information Systems*, Springer International Publishing, Cham, 27–35, 2014, DOI: 10.1007/978-3-319-01863-8_3.

Scherrer, P., J. Schou, R. Bush, A. Kosovichev, R. Bogart, et al. The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). *Sol. Phys.*, **275**, 207–227, 2012, DOI: 10.1007/s11207-011-9834-2.

Schuh, M., R. Angryk, K.G. Pillai, J. Banda, and P. Martens. A large scale solar image dataset with labeled event regions. In *Proc. International Conference on Image Processing (ICIP)*, 4349–4353, 2013, DOI: 10.1109/ICIP.2013.6738896.

Schuh, M., J. Banda, P. Bernasconi, R. Angryk, and P. Martens. A comparative evaluation of automated solar filament detection. *Sol. Phys.*, **289 (7)**, 2503–2524, 2014, DOI: 10.1007/s11207-014-0495-9.

Schuh, M., J. Banda, T. Wylie, P. McInerney, K.G. Pillai, and R. Angryk. On visualization techniques for solar data mining. *Astron. Comput.*, **10**, 32–42, 2015, DOI: 10.1016/j.ascom.2014.12.003.

Schuh, M.A., and R.A. Angryk. Massive labeled solar image data benchmarks for automated feature recognition. In *2014 IEEE International Conference on Big Data (Big Data)*, IEEE, 53–60, 2014, DOI: 10.1109/BigData.2014.7004404.

Schuh, M.A., R.A. Angryk, and P.C. Martens. Supporting data: a large-scale dataset of solar event reports from automated feature recognition modules, 2016, DOI: 10.5281/zenodo.48187.

Thompson, W. Coordinate systems for solar image data. *A&A*, **449 (2)**, 791–803, 2006, DOI: 10.1051/0004-6361:20054262.3.

Verbeeck, C., V. Delouille, B. Mampaey, and R. De Visscher. The SPoCA-suite Software for extraction, characterization, and tracking of active regions and coronal holes on EUV images. *A&A*, **561**, A29, 2014, DOI: 10.1051/0004-6361/201321243.

Withbroe, G.L. Living With a Star. In *AAS/Solar Physics Division Meeting #31*, vol. 32 of Bulletin of the American Astronomical Society, 839, 2000, DOI: 10.1029/GM125p0045.

Zharkov, S., V.V. Zharkova, and S.S. Ipson. Statistical properties of sunspots in 1996–2004: I. Detection, North-South asymmetry and area distribution. *Sol. Phys.*, **228 (1)**, 377–397, 2005, DOI: 10.1007/s11207-005-5005-7.

Cite this article as: Schuh MA, Angryk RA & Martens PC. A large-scale dataset of solar event reports from automated feature recognition modules. *J. Space Weather Space Clim.*, **6**, A22, 2016, DOI: 10.1051/swsc/2016015.

Appendix

Here we include numerous tables detailing the raw statistics of all available event attributes. Note that event types are indicated by their labels presented in Table 1, and all available attribute units are described in Table A3.

Table A1. A summary of spatial attributes by event type, where *cenx* and *ceny* are the event center value locations in HPC arcseconds, *bboxarea* is the derived bounding box area in square arcseconds, and *ccpts* is the number of coordinate points in the event chain codes.

Event type	Attribute	Min	Mean	Median	Max	STD
AR	cenx	−968.062	−4.2582	−11.2734	970.17	654.4383
	ceny	−757.54	−25.1748	−53.004	766.076	305.4122
	bboxarea	129.6005	14,114.5961	7,797.6	211,639.68	17,179.5351
	ccpts	0	9.854	8	54	4.8074
CH	cenx	−969.977	3.5871	7.916	969.313	371.6026
	ceny	−970.799	−18.9467	9.1931	972.988	689.4635
	bboxarea	153	71,871.311	24,865.92	1,379,838.6	117,912.4412
	ccpts	0	14.3212	10	101	11.7254
EF	cenx	−834.576	21.0606	39.6695	834.54	456.419
	ceny	−837.084	−13.2525	−22.168	832.596	312.5886
	bboxarea	0.7943	5,159.4355	622.6166	306,826.8855	12,710.3634
	ccpts	n/a	n/a	n/a	n/a	n/a
FI	cenx	−939.594	−17.3933	−15.6858	933.778	463.6827
	ceny	−919.736	5.1539	45.0938	934.724	427.5714
	bboxarea	0	8,011.5614	3,764.2581	383,718.0908	13,634.0821
	ccpts	14	65.5248	46	877	57.0134
FL	cenx	−1,190.4	−21.1982	−38.4	1,113.6	619.6332
	ceny	−1190.4	−36.7188	−115.2	1,190.4	274.6577
	bboxarea	5,898.2323	9,588.3826	5,898.24	424,673.28	11,927.4292
	ccpts	n/a	n/a	n/a	n/a	n/a
SG	cenx	−842.688	11.0563	23.0722	774.33	433.033
	ceny	−624.42	−68.8113	−158.1399	687.582	283.9277
	bboxarea	3,466.469	16,417.615	11,885.1877	231,581.6319	13,677.0289
	ccpts	n/a	n/a	n/a	n/a	n/a
SS	cenx	−939.51	−10.1182	−14.721	936.61	563.2124
	ceny	−570.728	−68.118	−162.951	557.365	255.8895
	bboxarea	116.64	1,874.3158	1,280.79	47,151.45	2,571.1772
	ccpts	34	711.8052	587	8,728	515.3151

Table A2. A statistical summary of event-specific attribute values.

Event type	Attribute	Min	Mean	Median	Max	STD
AR	intensmin	0.00E+00	4.15E+02	3.91E+02	1.46E+03	1.55E+02
	intensmax	3.95E+02	2.94E+03	2.61E+03	9.57E+03	1.45E+03
	intensmean	2.44E+02	1.04E+03	1.03E+03	3.21E+03	2.47E+02
	intensmedian	2.33E+02	9.63E+02	9.57E+02	2.46E+03	2.33E+02
	intensvar	0.00E+00	1.49E+05	1.18E+05	7.12E+06	1.69E+05
	intensskew	−2.15E+00	1.13E+00	1.02E+00	1.22E+01	8.34E-01
	intenskurt	−2.00E+00	2.88E+00	1.03E+00	2.61E+02	8.11E+00
	intenstotal	0.00E+00	1.26E+07	7.65E+06	1.51E+08	1.45E+07
CH	intensmin	3.00E+00	2.39E+01	2.30E+01	7.80E+01	1.05E+01
	intensmax	3.20E+01	6.16E+01	6.30E+01	9.20E+01	1.14E+01
	intensmean	1.89E+01	4.52E+01	4.55E+01	8.48E+01	1.09E+01
	intensmedian	1.72E+01	4.51E+01	4.55E+01	8.35E+01	1.12E+01
	intensvar	0.00E+00	6.46E+01	5.23E+01	3.46E+02	4.33E+01
	intensskew	−1.34E+00	−2.30E-03	−2.00E-02	2.05E+00	2.48E-01
	intenskurt	−2.00E+00	−6.47E-01	−6.53E-01	4.08E+00	2.46E-01
	intenstotal	0.00E+00	2.17E+06	6.59E+05	3.86E+07	3.69E+06
EF	ef_pospeakfluxonsetrate	−1.13E+02	4.36E+02	3.80E+01	1.15E+05	2.73E+03
	ef_negpeakfluxonsetrate	−5.13E+05	−4.22E+02	−4.05E+01	8.99E+01	3.46E+03
	ef_sumpossignedflux	−7.10E+03	1.12E+03	2.54E+01	6.17E+05	1.26E+04
	ef_sumnegsignedflux	−8.60E+05	−1.03E+03	−2.60E+01	1.31E+04	1.11E+04
	ef_axisorientation	−3.14E+00	−3.15E-02	−7.74E-02	3.14E+00	1.82E+00
	ef_axislength	1.18E-01	2.58E+00	1.75E+00	2.13E+01	2.26E+00
	ef_posequivradius	2.06E-01	1.44E+00	7.45E-01	2.13E+01	1.80E+00
	ef_negequivradius	2.06E-01	1.53E+00	7.77E-01	1.90E+01	1.85E+00
	ef_aspectratio	2.66E-02	2.00E+00	1.87E+00	7.44E+00	8.50E-01
	ef_proximityratio	2.30E-03	3.67E-01	3.59E-01	1.63E+00	2.09E-01
	maxmagfieldstrength	0.00E+00	4.42E+02	2.15E+02	3.59E+03	5.18E+02
FI	fi_length	1.44E+09	6.88E+09	4.66E+09	1.18E+11	6.18E+09
	fi_tilt	−9.00E+01	−5.95E-02	−2.64E-01	9.00E+01	5.17E+01
	fi_barbstot	0.00E+00	2.22E+00	2.00E+00	3.40E+01	2.25E+00
	fi_barbsr	0.00E+00	1.01E+00	1.00E+00	1.80E+01	1.26E+00
	fi_barbsl	0.00E+00	9.45E-01	1.00E+00	1.70E+01	1.20E+00
	fi_chirality	−1.00E+00	−1.42E-02	0.00E+00	1.00E+00	5.92E-01
FL	fl_peakflux	1.07E+00	4.50E+02	1.42E+02	7.65E+03	5.88E+02
SG	sg_aspectratio	6.00E+00	1.97E+01	1.02E+01	4.29E+03	5.42E+01
	sg_shape	−1.00E+00	−6.37E-02	−1.00E+00	1.00E+00	9.98E-01

Table A3. A description of string attributes provided for each event type.

Attribute	Event types	Value
intensunit	AR, CH	DN/s
area_unit	AR, CH, EF, FI, SS	km^2
event_pixelunit	AR, CH	DN/s
event_pixelunit	EF	HMI pixels
event_pixelunit	FI	0.98873197 arcsec/pix (for example)
ef_onsetrateunit	EF	emx/hr
ef_fluxunit	EF	emx
ef_axisorientationunit	EF	degrees CCW from + pole west in local solar frame
ef_lengthunit	EF	cm
maxmagfieldstrengthunit	EF	gauss
fi_lengthunit	FI	cm

Table A4. A statistical summary of attribute values for attributes shared by multiple event types.

Attribute	Event type	Min	Mean	Median	Max	STD
area_atdiskcenter	AR	0.00E+00	4.96E+09	2.64E+09	5.45E+10	5.81E+09
	CH	7.15E+07	2.46E+10	1.01E+10	3.82E+11	3.83E+10
	EF	8.61E+06	1.11E+09	1.52E+08	4.63E+10	2.36E+09
	SS	3.00E+08	7.26E+08	5.36E+08	1.26E+10	7.56E+08
area_atdiskcenteruncert	AR	0.00E+00	3.77E+08	2.37E+08	3.86E+09	3.74E+08
	CH	7.15E+07	1.34E+09	7.62E+08	1.47E+10	1.59E+09
	EF	4.69E+06	1.37E+09	2.14E+08	7.48E+10	3.18E+09
area_raw	AR	2.85E+07	2.57E+09	1.57E+09	2.48E+10	2.73E+09
	CH	5.61E+07	1.17E+10	4.00E+09	1.67E+11	1.93E+10
	EF	8.40E+06	8.66E+08	1.19E+08	3.92E+10	1.86E+09
	FI	6.79E+07	1.16E+09	6.30E+08	2.51E+10	1.60E+09
area_uncert	AR	1.22E+07	2.01E+08	1.36E+08	1.86E+09	1.85E+08
	CH	1.98E+07	6.75E+08	3.33E+08	9.91E+09	8.79E+08
	EF	4.58E+06	1.08E+09	1.62E+08	7.17E+10	2.55E+09
event_npixels	AR	1.54E+02	1.36E+04	8.35E+03	1.28E+05	1.44E+04
	CH	3.05E+02	6.19E+04	2.11E+04	8.97E+05	1.02E+05
	EF	6.40E+01	6.50E+03	8.96E+02	3.00E+05	1.40E+04
	FI	6.60E+01	1.13E+03	6.12E+02	2.44E+04	1.55E+03
	SS	2.72E+02	2.98E+03	2.17E+03	6.39E+04	3.49E+03