Draft version December 22, 2022
Typeset using LaTeX **preprint** style in AASTeX631

Properties of Flare-Imminent versus Flare-Quiet Active Regions from the Chromosphere through the Corona I: Introduction of the AIA Active Region Patches (AARPs)

Karin Dissauer,[1] K.D. Leka,[1,2] and Eric L. Wagner[1]

[1] *NorthWest Research Associates, 3380 Mitchell Lane, Boulder, CO 80301 USA*
[2] *Institute for Space-Earth Environmental Research, Nagoya University, Furo-cho Chikusa-ku, Nagoya, Aichi 464-8601 JAPAN*

(Received; Revised; Accepted)

Submitted to ApJ

ABSTRACT

We begin here a series of papers examining the chromospheric and coronal properties of solar active regions. This first paper describes an extensive dataset of images from the Atmospheric Imaging Assembly on the Solar Dynamics Observatory curated for large-sample analysis of this topic. Based on (and constructed to coordinate with) the "Active Region Patches" as identified by the pipeline data analysis system for the Helioseismic and Magnetic Imager (HMI) on the same mission, the "HARPs"), the "AIA Active Region Patches" (AARPs), described herein, comprise an unbiased multi-wavelength set of FITS files downsampled spatially only by way of HARP-centered patch extractions (full spatial sampling is retained), and downsampled in the temporal domain but still able to describe both short-lived kinematics and longer-term trends. The AARPs database enables physics-informed parametrization and analysis using Nonparametric Discriminant Analysis in Paper II of this series, and is validated for analysis using Differential Emission Measure techniques. The AARP dataset presently covers mid-2010 through December 2018, is ≈9TB in size, and available through the Solar Data Analysis Center (Dissauer et al. 2022b).

Keywords: methods: statistical – Sun: flares – Sun: corona – Sun: chromosphere

1. INTRODUCTION

Coronal magnetic topology, energetics, and dynamics are all believed to play key roles in triggering, powering, enabling energetic events. However, large-sample studies of active regions that

Corresponding author: Karin Dissauer
Corresponding author: K.D. Leka
dissauer@nwra.com; leka@nwra.com

leka@nwra.com

search for clues as to solar energetic event productivity have long focused on characterizing photospheric active region complexity through the analysis of continuum-image and magnetic field data, and the subsequent association of resulting descriptors with flare activity (Sawyer et al. 1986; Zirin & Liggett 1987; McIntosh 1990; Bornmann & Shaw 1994; McAteer et al. 2005; Leka & Barnes 2007; Kontogiannis et al. 2019; Leka et al. 2018; Al-Ghraibah et al. 2015; Korsós et al. 2014). In addition to descriptors based on the morphology of white-light images and the spatial distribution and character of the magnetic fields, such as the Zurich sunspot classification system, photospheric analysis of solar active region flare productivity has also included plasma velocity and helicity patterns (Welsch et al. 2009; Park et al. 2018, 2021), wavelet and fractal analysis of photospheric images (Abramenko 2005; McAteer et al. 2005; Georgoulis 2012; Al-Ghraibah et al. 2015), and inferred subsurface plasma flows (Komm et al. 2011; Braun 2016).

The focus on photospheric magnetic fields and the drivers of their evolution makes physical sense, as these are ultimately the source of energy to power solar energetic events. The physical interpretation of the state of the photosphere for flare-productive active regions includes highly non-potential magnetic fields that indicate stored magnetic energy, strong electric current systems and spatial gradients that provide pathways for magnetic reconnection, and emerging flux episodes that can destabilize the system (e.g. Zirin & Tanaka 1973; Krall et al. 1982; Hagyard et al. 1984; Canfield et al. 1975; Wang et al. 1996; Smith et al. 1996), (see also Leka & Barnes 2003a, 2007, and references therein). These quantitatively interpretable characteristics provide the physics-based insight that can then guide or constrain numerical modeling and further understanding (recent examples include Threlfall et al. 2017). We intentionally cite some of the originating literature to highlight these physics-inspired investigations that led to *e.g.*, the Space-Weather HMI Active Region Patch ("SHARP" Bobra et al. 2014) *parameters* currently published as meta-data physical summaries of solar magnetic complexes as a data product from the Solar Dynamics Observatory (SDO Pesnell et al. 2012) Helioseisemic and Magnetic Imager (HMI Scherrer et al. 2012; Schou et al. 2012; Centeno et al. 2014; Hoeksema et al. 2014).

In this series of papers we begin to do the same, quantitatively, for the chromosphere, transition region, and corona. The upper layers of the atmosphere mirror, reflect, and react to the photospheric drivers. Case-studies of the pre-event chromosphere, transition region and corona have shown evidence of specific pre-event energization and kinematics, starting with enhanced Hydrogen Balmer-series "Hα" emission in flare-imminent active regions (Sawyer et al. 1986; Zirin & Marquette 1991), an increase in chromospheric non-thermal velocities and high blueshifts (Cho et al. 2016; Harra et al. 2013; Woods et al. 2017; Seki et al. 2017), very localized chromospheric heating (Li et al. 2005; Bamba et al. 2014). Coronal brightness and morphological signals include the energization and increased dynamic behavior of EUV structures ("crinkles"; Sterling & Moore 2001; Joshi et al. 2011; Sterling et al. 2011; Imada et al. 2014, and references therein), and coronal dimmings (Imada et al. 2014; Zhang et al. 2017; Qiu & Cheng 2017) in the hours prior to energetic events.

Large-sample studies of the corona that relate its physical state to energetic-event productivity have generally focused on modeling the coronal magnetic field and observationally inferring relevant topological characteristics, often for coronal mass ejections (*e.g.* Barnes 2007; Barnes & Leka 2006; Georgoulis & Rust 2007; Kusano et al. 2020) (although see Aggarwal et al. 2018, regarding filament dynamics and eruptivity).

Thus far, investigations using the large-sample data available from the SDO Atmospheric Imaging Assembly (AIA Lemen et al. 2012) in the context of energetic events have been carried out with a primary goal of forecasting, rather than physical understanding. To achieve this expressed objective, machine-learning tools have been invoked (Nishizuka et al. 2017; Jonas et al. 2018; Alipour et al. 2019), although these routinely subject the AIA data to significant degradation by spatial binning, which hampers the ability to discern any small-scale processes. The outcomes have focused on skill scores rather than physical interpretation, as "interpretable machine-learning" approaches have yet to be widely implemented in this context. In other words, on many levels, information about the physics of the corona has been lost or not pursued.

We focus here on the need to move away from case studies and toward the scientific objective of quantitatively characterizing the behavior of the chromosphere and corona using physically-meaningful analysis. We ask parallel questions about the chromosphere and corona as were posed in previous works that centered on the photosphere (Leka & Barnes 2003a,b; Barnes & Leka 2006; Leka & Barnes 2007): what are the broad physical differences between "flare-imminent" and "not-flare-imminent" active regions? For the photosphere, we focused on descriptors of the vector magnetic field and parametrizations that quantify, for example, the free magnetic energy available for energetic events. For the present objectives, we focus on both high-frequency kinematic analysis indicative of small-scale reconnection activity, and the hours-long trends in the brightness distributions indicative of growth, decay, or changing temperature distribution, or increasing or decreasing levels of small-scale activity. We design the AARP database to address these objectives with a statistically-significant sample. As such, while we are *not* actively focusing on flare "prediction" *per se*, the results of these investigations may inform those activities eventually, we propose that the AARP database is designed such that it is appropriate for other, very different, scientific objectives, as well.

In this first paper of the series, we describe the data preparation for, and full description of, the AIA Active Region Patch extractions (AARPs; Dissauer et al. 2022b, see Section 2). We present global, cycle-scale trends across wavelengths in Section 3, plus some additional analysis aimed to motivate community use of the AARP dataset. We leave the full discussion of physics-informed parametrization and results from Nonparametric Discriminant Analysis to Leka et al. (Paper II; 2022), and of a statistical analysis of the temperature and density of flare-imminent regions using Differential Emission Measure tools to Dissauer et al. (Paper III; 2022a).

2. THE AIA ACTIVE REGION PATCH (AARP) DATABASE

In this section we describe the AIA Active Region Patch (AARP) Database and its construction (see Figure 1). The data cubes were designed to make available a statistically significant sample of solar active region coronal and chromospheric imaging data with a database of tenable size (noting that the full-disk full-cadence AIA dataset for our target period is over 25PB). The resulting cubes are the basis for subsequent analysis (e.g. Leka et al. 2018) and are available to the community (Dissauer et al. 2022b).

In the present incarnation, we specifically select the cadence and duration of the cubes to match those we have produced for photospheric active region investigations using the vector-field data from HMI (Leka et al. 2018); future studies may require different such particulars. The infrastructure to produce the present dataset as well as others with different such options is available (Dissauer et al. 2022b), although it presently relies on specific data access and architecture (see Appendix B, C). Not

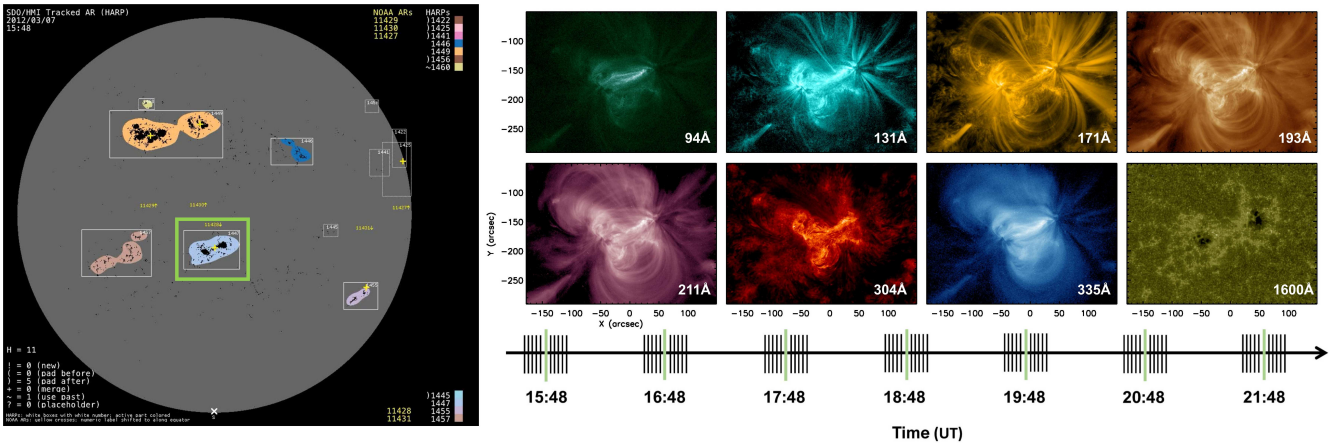


Figure 1. Overview of the AARP extraction and down-sampling in both spatial and temporal domains, for HARP/AARP 1447 on March 07, 2012. The AARP field-of-view is based on the HARP box (highlighted by the green boundary in the left panel; see Section 2.2). In the temporal domain, 11 images (right panel) per hour (see time line at the bottom) over the course of 1/4 of each day are extracted for 8 different (E)UV wavelengths (all but 1700Å; panels in upper/right).

everyone in the community has access to the hardware and software resources required to produce the AARP dataset; we hope it will be of use to the community as a curated, ready-for-analysis resource.

The HMI Active Region Patch ("HARP" Hoeksema et al. 2014; Bobra et al. 2014) definitions are used for the active-region identification system, and we intentionally include all HARPs. We do not down-select at the active-region level, either for size or activity. HARP designations are based on concentrations of magnetic activity, and the low-activity, small concentrations statistically dominate. As such, we avoid imposing a bias for later analysis, although any comparison of (for example) active *vs.* less-active regions must statistically account for the uneven sample sizes.

2.1. *AIA Data*

The AARP database comprises coronal- and chromospheric time-series patches that are extracted from the full-cadence full-size data of the Atmospheric Imaging Assembly (AIA; Lemen et al. 2012) on board NASA's Solar Dynamics Observatory (*SDO*; Pesnell et al. 2012). AIA obtains 4096^2-pixel full-disk coronal and chromospheric images at 1.5″ spatial resolution, sampled at 0.6″ through multiple extreme-ultraviolet filter bands at [94, 131, 171, 193, 211, 304, 335] Å with a 12 s cadence, and full-disk photospheric images at ultraviolet [1600, 1700] Å with a 24 s cadence. We embarked on this project to avoid copying *all* of the data, instead focusing on active-regions specifically for the first "down-selection" operation for the final database. However, that still required handling a significant amount of AIA data.

NWRA employs the Joint Science Operations Center ("JSOC") "NetDRMS" (Network Data Record Management System) / "RemoteSUMS" (Storage Unit Management System) system to be able to essentially mimic the availability of *SDO* (and other) data in a manner directly analogous to the host institution, Stanford University. Large data transfers are handled through the "JSOC Mirroring Daemon" (see Appendix B). The AIA meta-data reside in the `aia.lev1_euv_12s` and `aia.lev1_uv_24s` series, while the images reside in the `aia.lev1` series within the NetDRMS/RemoteSUMS system (the `aia.lev1` series is largely transparent for data transfers through, *e.g.*, the JSOC "LookData"

facility). Details of setting up automatic targeted Java Mirroring Daemon (JMD) transfers and performing the extractions are provided in Appendix B. For small numbers of HARPs or customized extractions, the `SolarSoft ssw_cutout_service.pro` (Freeland & Handy 1998) is recommended to be used, however as described in Appendix C, it was not the correct tool by which to develop the AARP database.

2.2. *Down-Select in the Spatial Domain: Extracting the Active Region Patches*

This dataset is created to address the almost untenable requirements of large-sample coronal studies. The "AIA Active Region patches" (AARPs) are first and foremost a down-selection in the spatial domain, providing sub-area extractions from the full-disk AIA image data that rely on the HMI Active Region Patches (HARPs; Hoeksema et al. 2014) metadata (`hmi.Mharp_720s` series) for target selection, pointing, and field-of-view (FOV) definition.

Because coronal structures extend in height from the solar surface, the extraction box is modified from the original HARP FOV to accommodate this projection effect both when a region is viewed face-on but also as it is viewed away from disk center (see Figure 2). Hence, we expand the original HARP FOV by 20% in both x-, y- directions, limited to a maximum expansion of 25″ to avoid unwieldy increases in already-large HARPs. Additionally, we augment with an expansion towards the East or the West $x_{\rm ex}$ depending on the position of the HARP center with respect to the central meridian in the form of

$$x_{\rm ex} = x_{\rm cen}[^\circ]/90^\circ \cdot 50'' \,, \tag{1}$$

where $x_{\rm cen}$ is the center of the HARP in degrees. This expansion is limited to a maximum extension of 50″ and is done to accommodate the additional extension of loops viewed from the side as a region approaches the solar limb. Figure 2 shows two examples, a disk-center and a near-limb target, comparing the original HARP (cyan box) and resulting AARP fields of view. While loop tops originating from the target active region may still be cut off, the primary science this dataset is designed for concerns relative (vs. absolute) changes in AR behavior, and since all regions are treated consistently, this should impose no statistical concerns. In addition, automatically defining an optimal FOV for each individual active region is challenged by (1) AARPs frequently being in close proximity during periods of high activity, and (2) an optically-thin corona, such that there is arguably no optimal way to automatically disentangle all visible loops, their source active regions, and still ensure their tops are always included and correctly assigned.

All HARPs are included regardless of flaring activity levels, size of active regions, or observing angle. The originating HARP bounding-box information is used for the central-time (the "XX:48") image targets. The extracted higher-cadence 11-image timeseries is co-aligned between the (nearly co-temporal) wavelengths and the central-time image using differential rotation to allow for Differential Emission Measure (DEM) analysis (Figure 3; see also Appendix E) but no further tracking is needed.

2.3. *Down-selecting in the Temporal Domain*

To reduce the data load while preserving the scientific requirements of the research performed in Paper II and further projects, we down-sample the data in the temporal domain. The native cadence for AIA EUV images is 12 s, and for the UV images it is 24 s. The AIA Automatic Exposure Control (AEC) system is an observation mode designed to avoid image saturation during solar flares. When the AEC system is invoked, the AEC-regulated images for the EUV data are inter-leaved with

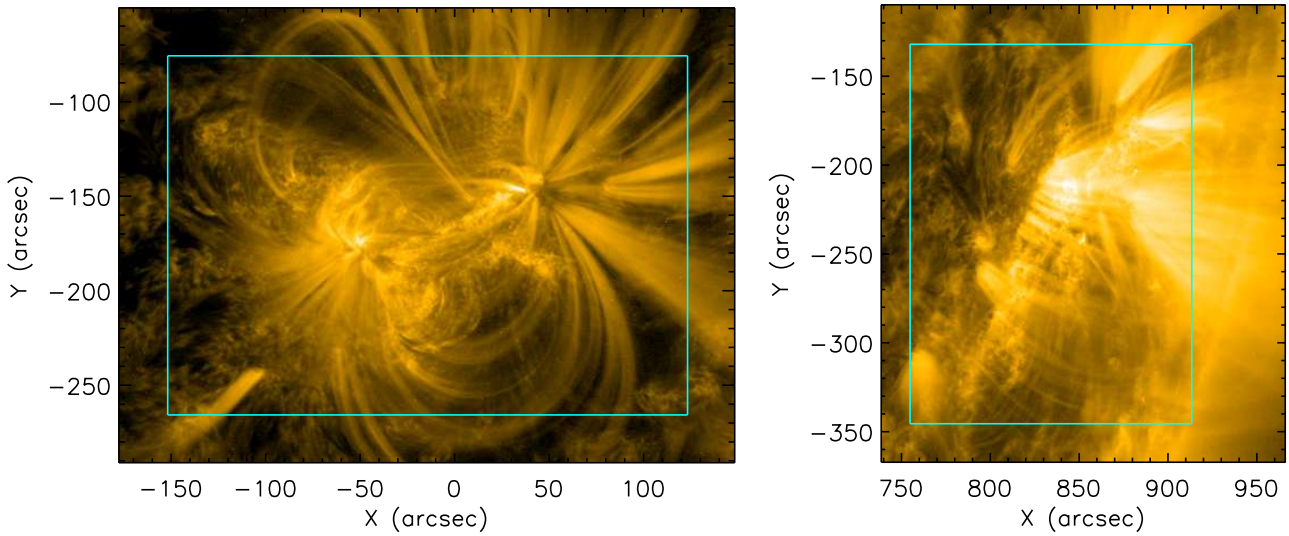


Figure 2. Two examples of the same AARP in 171Å illustrating the field-of-view expansions as a function of location on the disk and HARP dimensions. Symmetrical (left) extensions are used near disk center, and asymmetric extensions to include coronal loops are employed as regions approach the limbs, *e.g.* West (right) limbs. The cyan box marks the original HARP field-of-view. Left: AARP #1447 (NOAA AR 11428) 2012.03.07 15:48 UT; Right: AARP #1447 (NOAA AR 11428), 2012.03.12 15:48 UT.

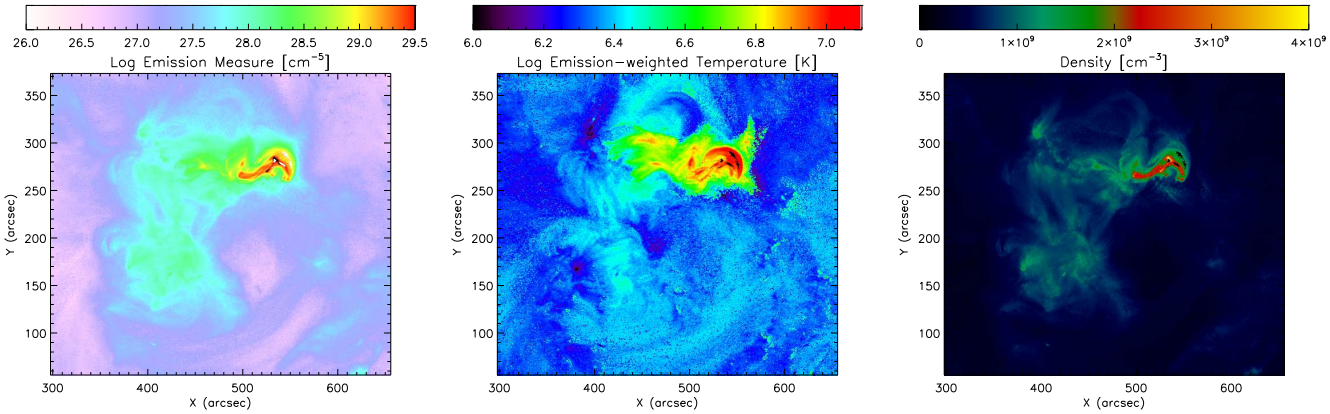


Figure 3. Analysis using Differential Emission Measure (Cheung et al. 2015) for HARP/AARP 3894 (NOAA ARs 12017 & 12018) on March 29, 2014 at 17:42:01 UT. Left: emission measure map (log-scale), Middle: emission-weighted temperature map (log-scale), Right: density map.

normal-exposure (possibly saturated) images, reducing the effective cadence for consistent-exposure EUV data to 24s. Native images in data numbers ("DN") or counts can be normalized by the exposure time to obtain consistent per-pixel count rate-based images. For reasons described below, however, we decided to avoid the AEC-regulated images and sample the AIA images at a 72 s cadence.

The image parametrization for the statistical analysis in Paper II and later work uses moment analysis to describe the coronal and chromospheric behavior. During flare times, even for small flares, the higher moments jump dramatically – as expected in the presence of localized yet high-magnitude brightness changes (Figure 4, left four panels). Our scientific focus is not, generally, the

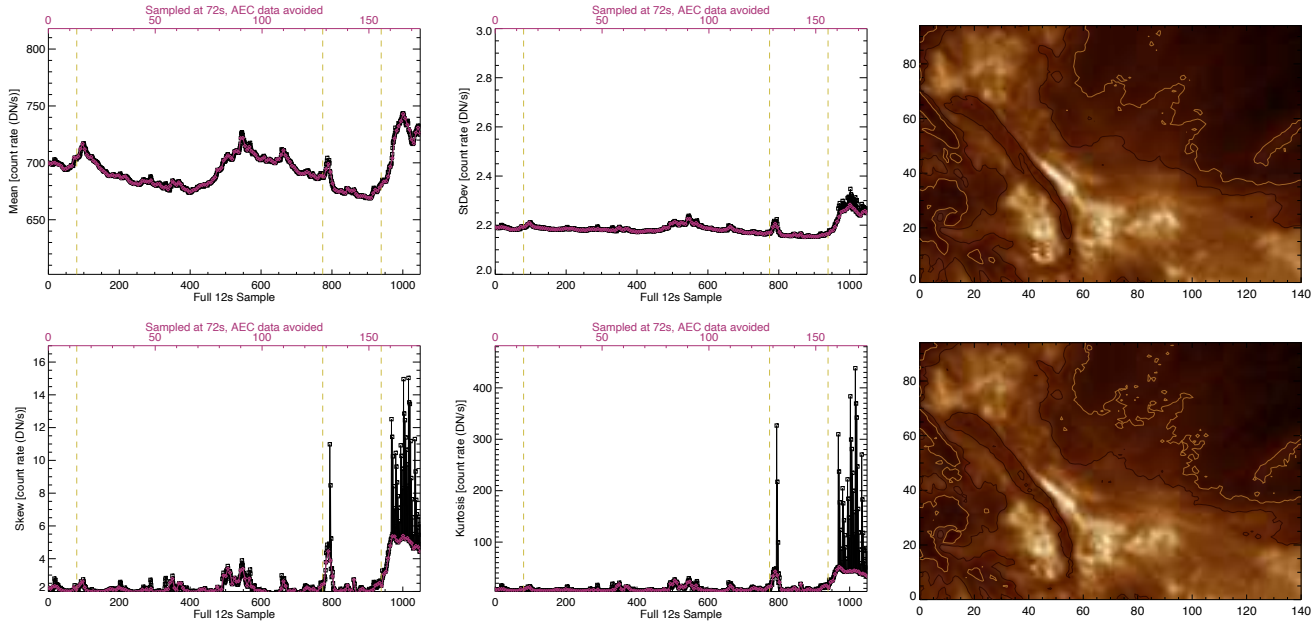


Figure 4. Example parameter curves from AARP #750 (NOAA AR #11261) for 193Å on 2011.08.03 00:00UT – 03:30 UT; the start of a sub-flare at 00:18 UT, a C1.2 at 02:35 UT and the M1.1 flare at 03:08 UT are indicated by dashed vertical lines. The bottom x-axis is an image index for the original 12 s sampling (black), the top x-axis are the indicies for the 72 s sampling (red) used for the AARPs that naturally avoided the AEC-triggered images. All data were exposure-normalized. Left four panels show the four moments of the images, as indicated. Right two panels show a small sub-area of the full AARP field of view that includes a small filament and nearby loop-bottoms (top/right of the images), but avoids the M1.1 flare-emission area, for two consecutive 12 s-sampled images, **Top**: normal 2 s exposure, **Bottom**: 0.43 s exposure, when AEC-mode was invoked. Light/Dark contours indicate the [2, 3] σ signal/noise levels, with the noise level determined by the most-probable value of low-signal parts of the image.

flaring plasma itself but the broad active region behavior. Avoiding the AEC images allows for times of flares to be included with less dramatic impact to the moment analysis. Additionally, the shorter exposures during flaring times lead to lower signal/noise ratios (SNR) for the larger active region areas (Figure 4, right panels). Since these regions are our focus, ensuring a consistent SNR is key. Finally, given the format of the AARP FITS files (see Appendix A), avoiding the AEC data allows a simplified metadata format, as the exposure time is consistent for a particular wavelength.

When working with running-difference images, central to our analysis approaches, the native cadence presents two challenges: (1) for consistent-exposure sequential images the resulting SNR was very low, and (2) when sequential images had varying exposure times, the resulting variation in the images' SNR contributed to even higher noise in the running-difference images. While the running-difference magnitudes increase as Δt, one must also balance this increase in SNR with the ability to evaluate and interpret the resulting images. Experimentation indicated that the full time resolution is unnecessary to capture the short-lived brightenings and kinematics that we wish to examine.

A coarser 72 s cadence for the AARP "high-cadence" bursts of images thus provides a good balance of consistent and acceptable SNR and dataset volume reduction. Selecting a 72 s cadence is also a "common factor" between the EUV and UV native sampling, allowing for the dataset to sample the different atmospheric layers at the same cadence.

The evolution of the corona over multiple hours is of high interest here. An interval covering ≈ 6 hr is motivated by numerical models of pre-event evolution and trigger formation (Ishiguro & Kusano 2017; Inoue et al. 2018), and provides sufficient data from which to quantify overall evolutionary trends within an active region. Still, 72 s sampling over 6 hr is a large data load. Hence we further down-select to ≈ 13 min of images at the 72 s cadence, centered hourly 15:48 – 21:48 UT (seven hourly "bursts" of images, over six hours, inclusive). The choice of timing is driven by an already-developed set of HMI vector-field time-series extracted data-set (Leka et al. 2018). Centering the data on the "XX:48" times was necessary to avoid those HMI data affected by calibration sequences (taken at 00:00 and 18:00 TAI, see Hoeksema et al. (2014)). Those studies also had a scientific requirement to minimize the statistical impacts on time-series analysis that were solely due to the spacecraft orbital velocity-induced artifacts (Hoeksema et al. 2014). Finally, we chose 11 images covering 13 m to effectively match the time period over which the input spectra are averaged for the hmi.B_720s series magnetograms. In this sense, we sample the upper atmosphere on a cadence appropriate to its physics (reconnection, heating, flows) but carefully pair the extractions to their arguably slower-evolving photospheric driver, at least at the HMI spatial resolution.

Thus, in summary, with the AARPs dataset the detailed behavior of the lower solar atmosphere is sampled at a fair cadence (72 s) for 13 m every hour, repeatedly over a quarter of a day for a full 8.5 yr (06/2010 – 12/2018) over which we can query about active region trends using a large-sample approach that enables robust statistical analysis.

2.4. *Final Data-Cube Preparation*

All AIA data are processed through the `SolarSoft` (Freeland & Handy 1998) routine `aia_prep.pro` and a correction to account for time-dependent degradation of the instrument was applied using the following form:

$$\frac{A_{eff}(t_{obs})}{A_{eff}(t_0)}\left(1 + p_1\delta t + p_2\delta t^2 + p_3\delta t^3\right) \ , \tag{2}$$

where $A_{eff}(t_{obs})$ is the effective area calculated at the calibration epoch for t_{obs}, $A_{eff}(t_0)$ is the effective area at the first calibration epoch (i.e. at launch), $p1$, $p2$, $p3$ are the interpolation coefficients for the t_{obs} epoch, and δt is the difference between the start time of the epoch and t_{obs} (Barnes et al. 2020).

All wavelengths and imaging data centered at each hour are coaligned and differentially rotated to the central time step (usually **:48 UT) using `drot_map.pro`. The pre-processed output data are therefore ready for running-difference analysis and Differential Emission Measure (DEM) analysis (see Appendix E). Full FITS headers are generated with wavelength, HARP number, NOAA Active Region (AR) number (from the relevant keywords in the `hmi.MHarp_720s` series). The seven sets of hourly image sets are saved as extensions each with its own header recording times, number of valid images, and detailed pointing information as well as additional keywords imported from the `hmi.MHarp_720s` series such as `LAT_FWT, LON_FWT` and the World Coordinate System (WCS; Thompson 2006) keywords for the target 720s-based mid-time. Appendix A contains examples of these AARP fits headers.

The final data product thus comprises a set of eight FITS files, one for each wavelength (by default: 94, 131, 171, 193, 211, 304, 335, and 1600 Å), for each HARP covering six hours inclusive (seven hourly samples 15:48 – 21:48 UT) of 13 minutes of data sampled at 72 s, resulting in 11 images per hour. The full dataset is summarized in Table 1, its archive size being only ≈ 9.5 TB, in contrast to the size of the full-disk AIA dataset for this period which is ≈25PB. The power of the dataset is

its large, yet tenable sample size, designed to enable unbiased statistical analysis on spatially native-resolution images. We next present some topics for which models or case-studies may now be tested on a statistically significant, cycle-covering sample.

Table 1. AARP Data Set Summary

Date Range	HARP Range	NOAA AR Range	Total Number of Samples	Archive Size
06/2010 – 12/2018	36 – 7331	11073 – 12731	256,976	≈ 9.5 TB

3. ANALYSIS & RESULTS

We present here a brief analysis to highlight the breadth of physics investigations that could be performed using the AARP database. We do not discuss flare- or event-related studies in depth, deferring this topic to Paper II (although see Section 3.4).

3.1. *Active Region Coronal Behavior with Solar Cycle*

With this extensive database we can begin to examine some subtleties of coronal behavior for active regions as a function of solar cycle. Generally speaking the AARPs will not include significant coronal hole areas, and thus will avoid that contamination and source of confusion with regards to interpretation.

In Figure 5 we present density histograms of total emission in AARPs for two AIA bands over most of Solar Cycle 24, as well as the same presentation for the mean emission in the same bands. We focus on 171 Å that is sensitive to plasma at "quiet" coronal temperatures and transition region emission, and thus is often used to identify coronal magnetic structures such as coronal loops, and the 94 Å band that is sensitive to hot plasma (Lemen et al. 2012), but that can contain contributions from cooler lines within the bandpass (O'Dwyer et al. 2010; Warren et al. 2012; Del Zanna 2013).

The "total" emission ($\Sigma(I_*)$) distribution tracks nicely with sunspot number, as expected, since it is an extensive parameter that scales with the size of the box, number of pixels, or more physically, the total amount of magnetic flux present. In both bands we see that the vast majority of AARPs have small total emission, a high density of points that itself tracks the sunspot number generally. In both bands there are also the less-frequent larger-total-emission regions whose distribution tracks the sunspot number even closer. None of this is unexpected.

What is interesting is that while for most "bumps" in total emission, *i.e.* those for which the eye sees a good correlation with a "bump" in sunspot number, there is good correspondence between the two passbands. However, there are apparently some large active regions for which a large total 171 Å is *not* matched with a correspondingly large total in 94 Å (*e.g.* late 2010, late 2017, early 2018) and *vice versa* (*e.g.* mid 2012, early 2015). Thus, while well correlated, the total emission displayed in these two channels may not be uniquely correlated, which prompts the questions "why or why not?" One immediate reason is the multi-thermal nature of both the AIA channels and active regions themselves. Clearly this extensive dataset can begin to address the details of active region coronal emission with, *e.g.*, AR age, magnetic field morphology, and flaring history, and provide constraints for modeling efforts both with regards to intra-filter expected emission and variations with solar

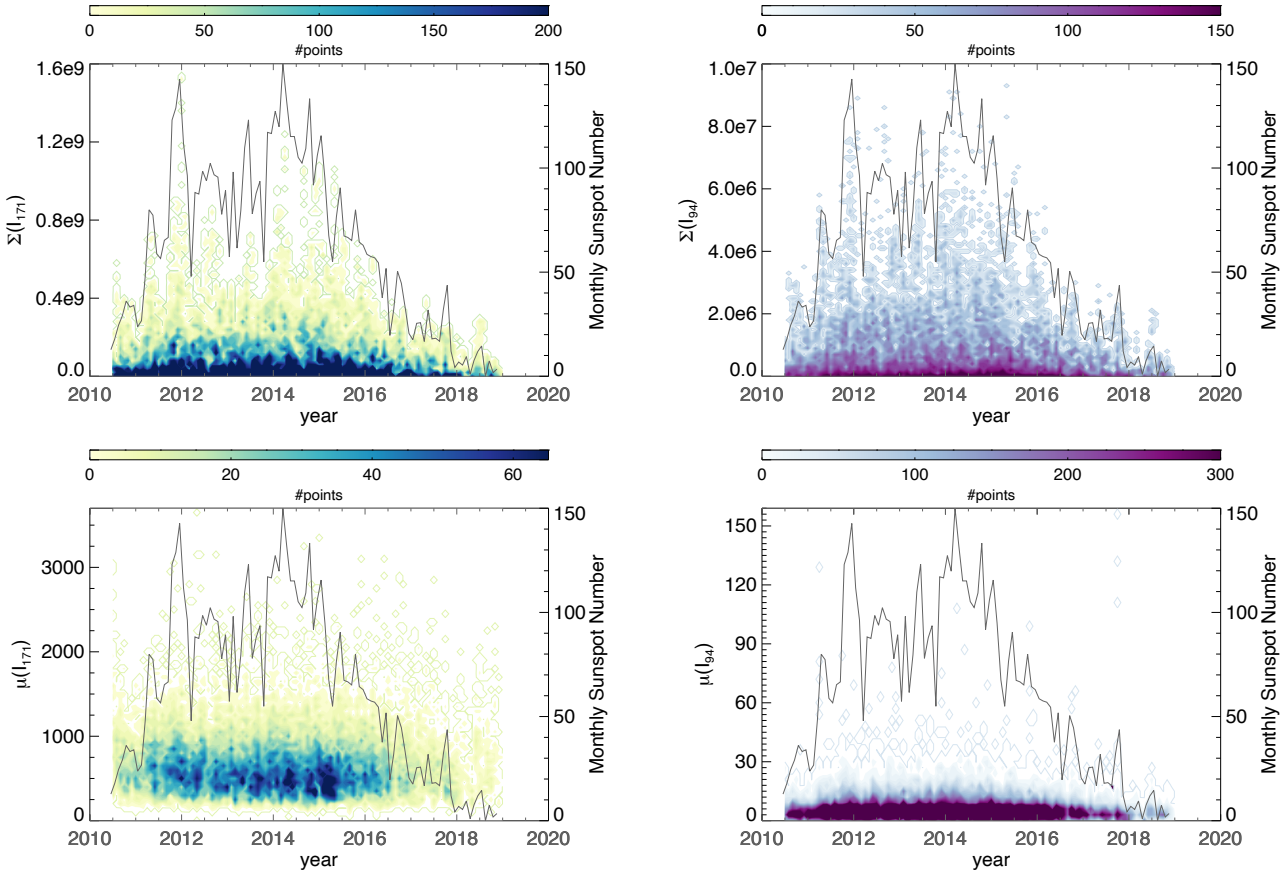


Figure 5. Top Row: Density histograms of the total emission of each AARP for 171Å (left) and 94Å (right) as a function of date, left axes. Also plotted is the monthly mean sunspot number from SILSO World Data Center (2010-2018), right axes. The density histograms do include the 'single-point' values as the lowest contour / lightest color. Bottom Row: Same presentation but for the mean emission (size-normalized) of the two bandpasses.

cycle. This is especially true since, as discussed in sections 2.4, Appendix E - the data are ready for not just hot-plasma isolation (Warren et al. 2012), but full DEM analysis.

The mean emission tells a slightly different story. Figure 5 shows that if there magnetic flux is present (as defined the HARPs and hence the AARPs), there is emission in 171 Å, or $\mu(I_{171}) \gg 0$, essentially. The same is not quite as true for $\mu(I_{94})$ where the majority of AARPs display a low or near-zero mean emission. This is consistent with $\mu(I_{94})$ having sensitivity to hot coronal emission, which not all active regions contain (especially flare-quiet ones). However, $\mu(I_{171})/\mu(I_{94})$ may not be perfectly constant with solar cycle. These simple findings are consistent with Schonfeld et al. (2017)'s sun-as-a-star investigation showing that the cool / quiet-Sun corona varied little over the first half of SC24 while the hot component varies strongly with the rise in solar activity. A thorough investigation using the AARP database could provide further physical constraints especially with respect to the AR-based contributions to the two components and possibly with regards to "terminator" analysis (Leamon et al. 2022).

3.2. *Active Region Coronal Behavior with Observing Angle*

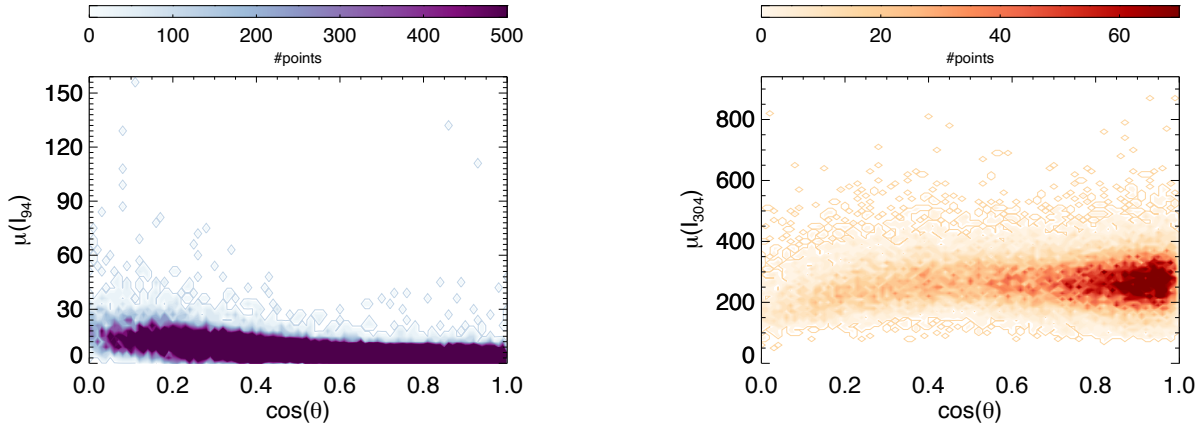


Figure 6. Density histograms of the area-averaged 94 Å emission (left) and 304 Å emission (right) as a function of observing angle. Shown are the results for all AARPs regardless of size or other selection criterion.

The coronal emission lines are generally optically thin. However, a few of the AIA bandpasses are centered on, or include, optically thick emission (Golding et al. 2017). Additionally, when densities become large such as during solar energetic events, some of the normally optically-thin lines in the sampled bandpasses become optically thick (Thiemann et al. 2018). The differences in behavior can be seen when the mean emission levels of different lines are presented according to their location on the disk (Figure 6). The 94 Å channel generally displays a small mean emission that is flat with $\mu = \cos(\theta)$ except at the very limb, where the optically-thin properties afford a propensity to integrate over more emitting structures along the line-of-sight.

The He II 304 Å emission is optically thick but has challenging radiative transfer characteristics (Golding et al. 2017), and thus expected to show a correlation with observing angle $\cos(\theta)$ (decreasing toward the limb). The density distribution of $\mu(I_{304})$ (Figure 6, right) does show a slight turnover beyond $\mu = \cos(\theta) = 0.25$ or $\theta = 75°$ but it is extremely slight. There is no severe or obvious attenuation with observing angle, but the behavior of He II 304 Å emission in active regions may not have yet been studied with large-sample data. There is the possibility of contamination within the AIA filter by optically-thin emission that would mask the He II 304 Å behavior, and it also may be contaminated by the different behavior expected by different magnetic structures within the AARP fields of view (Mango et al. 1978; Worden et al. 1999). The detailed behavior of the corona as inferred from optically-thin dominated *vs.* optically-thick dominated AIA filters for active regions can now, with the AARP database, be studied in detail with large-sample statistical analysis.

3.3. *Coronal Behavior with Active Region Emergence / Decay*

The relationships (in the spatial and temporal domains) between coronal emission and emerging magnetic flux should provide insights into the energy transfer to the upper atmosphere from the photosphere and below. Similarly, the relationships between coronal emission and decaying active regions should elucidate the final transfer of magnetic energy out of the Sun and the dominant dispersion and dissipative mechanisms at play. The AARPs provide the ability to examine these processes by sampling regions at all sizes and stages of evolution with a curated, large-sample dataset.

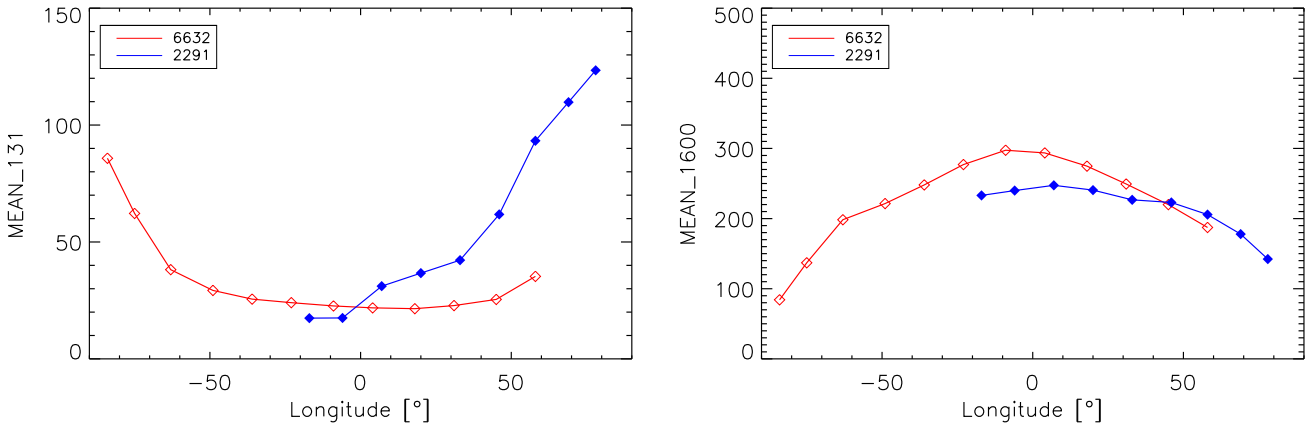


Figure 7. Two HARP/AARP regions and their mean emissions in two AIA filters. AARP 2291 grew rapidly starting near disk-center in 2012.12 to become NOAA AR 11631 and 11632. AARP 6632 (NOAA AR 12556, 2016.06) decays during disk passage. See text for discussion.

In Figure 7 we show the transition region/flaring plasma sensitive $\mu(I_{131})$ and upper-photospheric $\mu(I_{1600})$ for two AARPs that are in significantly different stages of evolution as they traverse the disk. AARP 2291 underwent rapid growth starting near disk center and eventually became NOAA ARs 11631 and 11632. It displays a rapid growth in the mean intensity in the Fe VIII,XXI 131 Å filter emission. AARP 6632 rotated on to the disk as NOAA AR 12556 and decayed to plage by disk-center. Its 131 Å filter emission decays to essentially background. The comparison of the two regions in the $\mu(I_{1600})$, however, indicate that (as expected) this emission is optically thick (decreases systematically toward the limb) and is sensitive to the presence of plage rather than sunspots, with little difference between young and old plage.

Obviously we simply present a comparison here in order to encourage interest. We do not, for example, try to perform a larger-sample re-examination of the relationship between flux transport, flux decay, and active region EUV radiance decay (Ugarte-Urra et al. 2017) but that is one study the AARPs could confirm with larger-sample statistics. Other questions are ready for analysis: Are there significant differences in coronal characteristics between old and new plage? Between emerging regions that will become flare productive and those that will stay quiet?

3.4. *Coronal Behavior with Activity Level*

The AARP dataset was prepared specifically for a large-sample investigation into the coronal behavior as related to flaring activity. We defer our study details and results to Paper II (Leka et al. 2022) but here present a preliminary example of the directions available for study with the AARP dataset.

We note that indeed, the AARPs are not extracted according to the time of flares, meaning that they are *not* designed for super-posed epoch analysis (*e.g.* Reinard et al. 2010; Mason & Hoeksema 2010; Bobra & Couvidat 2015; Jonas et al. 2018). They are, however, extracted for quantitative interpretable analysis and not flare prediction in & of itself, as have been many recent studies using large samples of AIA image data (*e.g.* Nishizuka et al. 2017; Jonas et al. 2018; Alipour et al. 2019).

In Figure 8 we show both the "direct" ("I_{193}") and running-difference ("ΔI_{193}") images, in the 193Å band, for NOAA AR 11261 (HMI HARP 750, see Figure 4) on 2011.08.03. The quieter example is from the images centered at 21:48 UT, and the "active" example is taken from images centered at 19:48 UT during a C8.5 flare. The two "direct" images look fairly similar, even during a flare, with a small localized brightening the only easily discernible difference. The running-difference images look very different. In the quiet example, there is a lot of small-scale fluctuation happening even outside the flare time. In the flare case, it is true that the saturated-emission area will show a lack of any dynamics, but do note that as the edges change, there are strong temporal gradients in this area. The expanding loops are also clearly visible in the running difference image.

The images look similar to the eye, as the flare was not very large, but the images are quantitatively different as summarized by the moments (mean μ, standard deviation σ, skew ς and kurtosis κ) presented in the table associated with Figure 8. In particular, the mean intensity at the two times is similar for both the direct and running-difference images (the latter both being fairly close to zero relative to the mean intensities) but the higher-order moments both for the direct and especially the running-difference are very different. Physically this implies enhanced localized brightening and kinematics on short timescales with little overall brightness enhancement or mean brightness increase as would indicate (for example) significant differences in active-region-scale heating or density between the two time periods.

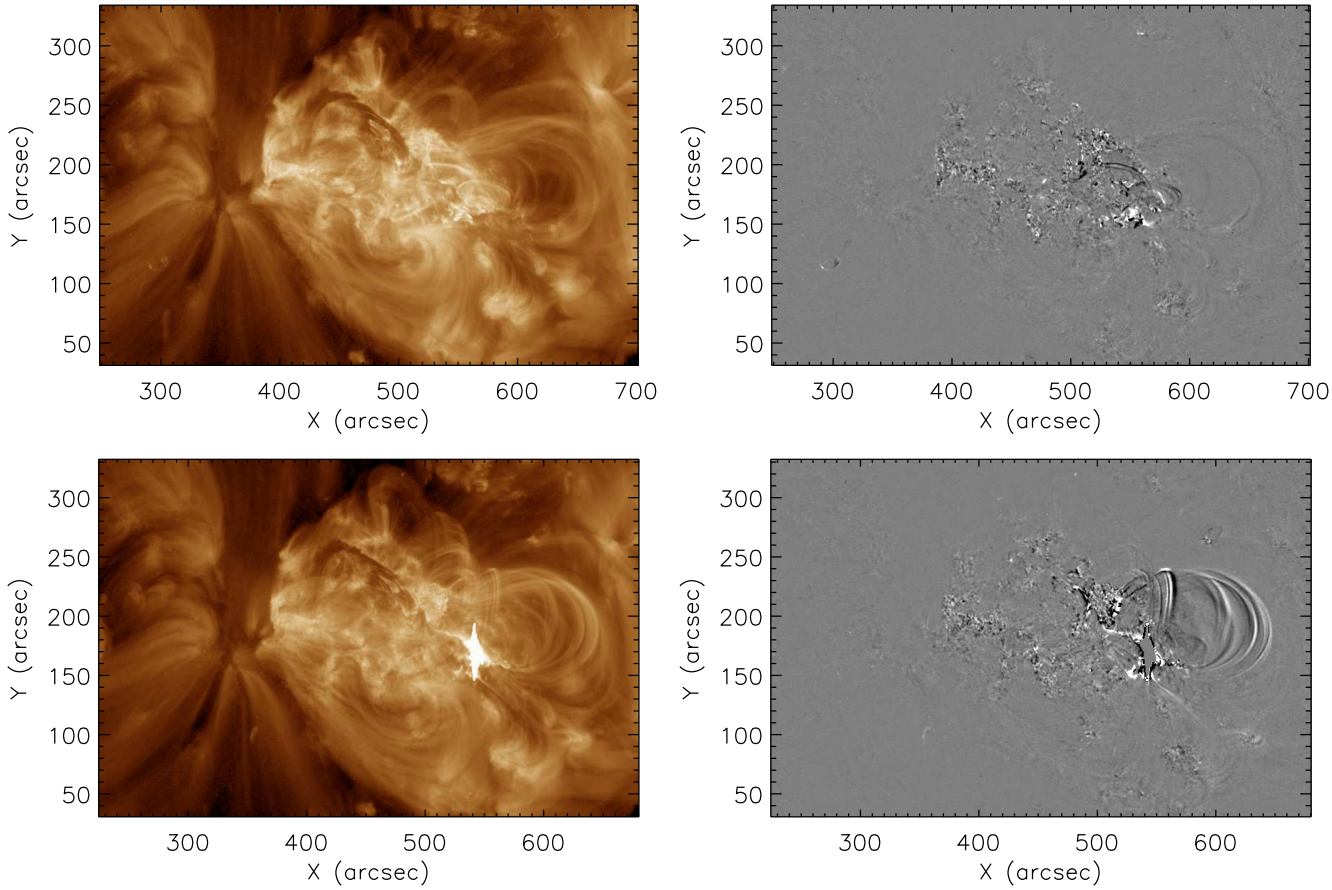
4. SUMMARY AND CONCLUSION

We present here the NWRA "AIA Active Region Patches" (AARP) Database (Dissauer et al. 2022b) that well-samples the temporal evolution of solar active regions as captured in the E/UV with SDO/AIA over the majority of Solar Cycle 24 from 2010/06 – 2018/12. The current version of the database includes daily, 7 hr samples of 13 minutes of images (centered hourly on "*:48 UT" from 15:48 – 21:48 UT) targeting all magnetic patches identified by SDO/HMI, resulting in a total sample size of 256,976 FITS files containing in total almost 20 Million individual SDO/AIA images. We show a teaser of the analysis that could be performed using this dataset, including coronal behavior as related to the solar cycle, to the emergence and decay of solar active regions, as well as center-to-limb variations.

Crucially, the dataset is prepared with attention to quantitative analysis methodology. The AARPs dataset preserves the native spatial resolution of SDO/AIA (i.e. 1.5″ sampled at 0.6″), in contrast to other studies that downsample the full-disk images to *e.g.*, 512×512 (e.g. Galvez et al. 2019) or use solely an active region's total intensity or maximum intensity in a particular channel (e.g. Nishizuka et al. 2017). Keeping the full spatial resolution allows us to capture small-scale dynamics of the studied active regions (see e.g. Figure 8), which are otherwise lost due to binning.

However, we note that the dataset is (currently) limited to a selected time range (to match the NWRA database of HARP-based photospheric vector field timeseries data (Leka et al. 2018)) although with an extended field-of-view from the original defining HARP bounding box to capture the projections of structures that extend in height. For our initial purpose of investigating unique coronal and chromospheric characteristics of flare-imminent active regions, this is a valid approach since both the short-term changes (over the course of 13 minutes every hour at a cadence of 72 s) as well as longer-term trends (7 hr of evolution per day per HARP) can be studied.

We stress that the AARP dataset is ready for machine-learning applications or any other large-sample analysis. Importantly, it has been validated for Differential Emission Measure analysis (see



	Direct				Running-Difference			
	μ	σ	ς	κ	μ	σ	ς	κ
Quiet	965.9	837.8	2.38	7.22	0.58	77.0	1.78	342.5
Flare	1022.2	1192.6	6.87	77.8	-0.87	225.2	-21.2	1665.5

Figure 8. "Direct" (left) and running-difference (right) images of NOAA AR 11261 (HMI HARP 750, see Figure 4) on 2011.08.03 showing the differences between a quiet time (top, 21:46:50-21:45:38 UT) and during a small flare (bottom, 19:46:38-19:45:26 UT). Of note, the pairs of images are scaled the same, and while the flaring time does show a small patch of saturation since we avoid the AEC images, the pixels around the saturation area have extreme signals in the running-difference images as the saturated area changes with time.

Appendix E), which is a data product to be released in the near future. We note that this dataset could be easily expanded to include additional times sampled over each day, even at a different cadence, such that both forecast-mode and superposed epoch mode analysis can be performed on every single flare that occurred since the start of the SDO mission. Although currently beyond the scope of this paper and not needed for the analysis presented in Paper II, such an expansion is quite feasible given the infrastructure NWRA has now developed.

The authors thank the referee for a thorough reading and insightful feedback that helped improve the paper. This work was made possible by funding from AFRL SBIR Phase-I contract FA9453-14-M-0170, (initial exploration), but primarily by NASA/GI Grant 80NSSC19K0285 with some additional support from NASA/GI Grant 80NSSC21K0738 and NSF/AGS-ST Grant 2154653. We extend our sincere thanks to Sam Freeland for his patience and help with the `ssw_cutout_service.pro` challenges.

APPENDIX

A. THE AARP FITS FILE FORMAT

AARP FITS files contain in total eight extensions. The first extension, i.e. extension 0 is the primary header and contains no data. A sample primary header is given in Table 2. The primary header is be followed by seven image extensions that contain 11 images per extension, from 15:48–21:48 UT. An example for an image extension header is given in Table 3.

The FITS keyword values are either extracted from the original AIA keywords[1] or computed from the WCS coordinates upon field-of-view determination and extraction.

SIMPLE	INT	1	
BITPIX	LONG	16	
NAXIS	LONG	0	NUMBER OF DIMENSIONS
EXTEND	INT	1	FITS FILE CONTAINS EXTENSIONS
INSTRUME	STRING	'AIA'	Instrument name
TELESCOP	STRING	'SDO/AIA'	
WAVELNTH	LONG	211	[angstrom] Wavelength
HARPNUM	LONG	377	HMI Active Region Patch number
NOAA_AR	LONG	11158	NOAA AR number that best matches this HARP
NOAA_NUM	LONG	1	Number of NOAA ARs matching this HARP (0 allowed)
NOAA_ARS	STRING	'11158'	List of NOAA ARs matching this HARP
NSAMPS	LONG	11	Number of images in each extension
NTIMES	LONG	7	Total number of times in time series
T_START	STRING	'2011.02.15_15:48:00_TAI'	Time of first observation in timeseries
T_STOP	STRING	'2011.02.15_21:48:00_TAI'	Time of last observation in time series
TREC_00	STRING	'2011.02.15_15:48:00_TAI'	Extension #1 time
TREC_01	STRING	'2011.02.15_16:48:00_TAI'	Extension #2 time
TREC_02	STRING	'2011.02.15_17:48:00_TAI'	Extension #3 time
TREC_03	STRING	'2011.02.15_18:48:00_TAI'	Extension #4 time
TREC_04	STRING	'2011.02.15_19:48:00_TAI'	Extension #5 time
TREC_05	STRING	'2011.02.15_20:48:00_TAI'	Extension #6 time
TREC_06	STRING	'2011.02.15_21:48:00_TAI'	Extension #7 time
NXMAX	LONG	692	Max NAXIS1 over all extensions
NYMAX	LONG	425	Max NAXIS2 over all extensions
NXMIN	LONG	688	Min NAXIS1 over all extensions
NYMIN	LONG	421	Min NAXIS2 over all extensions
HISTORY	STRING	'aia_patches completed...'	
END			

Table 2. Sample FITS header of the primary extension (extension 0) and its keywords. The information contained includes the wavelength, the HARP number, the associated NOAA active region number, the central time of each image extension as well as the maximal dimensions of the image data cubes.

[1] http://jsoc.stanford.edu/~jsoc/keywords/AIA/AIA02840_K_AIA-SDO_FITS_Keyword_Document.pdf

XTENSION	STRING	'IMAGE'	AIA DATA IMAGE EXTENSION
BITPIX	LONG	-32	
NAXIS	LONG	3	NUMBER OF DIMENSIONS
NAXIS1	LONG	692	X Pixels
NAXIS2	LONG	421	Y Pixels
NAXIS3	LONG	11	Samples in time
T_REC	STRING	'2011.02.15_15:48:00_TAI'	[TAI time]
T_IDX	LONG	0	Hours since timeseries start
VALID	LONG	1	Extension contains valid (1) or no (0) data
NIMVALID	LONG	11	Valid images out of 11
T_IMG00	STRING	'2011-02-15T15:42:02Z'	[ISO8601] time for image 00
T_IMG01	STRING	'2011-02-15T15:43:14Z'	[ISO8601] time for image 01
T_IMG02	STRING	'2011-02-15T15:44:26Z'	[ISO8601] time for image 02
T_IMG03	STRING	'2011-02-15T15:45:38Z'	[ISO8601] time for image 03
T_IMG04	STRING	'2011-02-15T15:46:50Z'	[ISO8601] time for image 04
T_IMG05	STRING	'2011-02-15T15:48:02Z'	[ISO8601] time for image 05
T_IMG06	STRING	'2011-02-15T15:49:14Z'	[ISO8601] time for image 06
T_IMG07	STRING	'2011-02-15T15:50:26Z'	[ISO8601] time for image 07
T_IMG08	STRING	'2011-02-15T15:51:38Z'	[ISO8601] time for image 08
T_IMG09	STRING	'2011-02-15T15:52:50Z'	[ISO8601] time for image 09
T_IMG10	STRING	'2011-02-15T15:54:02Z'	[ISO8601] time for image 10
REFHELIO	STRING	'W21S21'	Heliographic pointing string
HARPNUM	LONG	377	HMI Active Patch number
NOAA_AR	LONG	11158	NOAA AR number that best matches this HARP
NOAA_NUM	LONG	1	Number of NOAA ARs matching this HARP (0 allowed)
NOAA_ARS	STRING	'11158'	List of NOAA ARs matching this HARP
LAT_FWT	DOUBLE	-20.157000	[deg] Stonyhurst LAT of flux-weighted cntr pix
LON_FWT	DOUBLE	20.295200	[deg] Stonyhurst LON of flux-weighted cntr pix
WAVELNTH	LONG	211	[angstrom] Wavelength
WAVEUNIT	STRING	'angstrom'	Wavelength unit: angstrom
T_REF	STRING	'2011.02.15_15:48:00_TAI'	[time] Pointing reference time
DATE_OBS	STRING	'2011-02-15T15:48:00.62'	[ISO8601] Date when observation started
T_OBS	STRING	'2011-02-15T15:48:02.07Z'	[ISO8601] Observation time
EXPTIME	DOUBLE	2.9012420	[sec] Exposure duration: mean shutter open time
EXPSDEV	DOUBLE	0.00020700000	[sec] Exposure standard deviation
INT_TIME	DOUBLE	3.1562500	[sec] CCD integration duration
PIXLUNIT	STRING	'DN'	Pixel intensity unit
DN_GAIN	DOUBLE	18.300000	[DN/electron]
AECDELAY	LONG	1536	AIA_IMG_AEC_DELAY
AECTYPE	LONG	0	AIA_IMG_AEC_TYPE
AECMODE	STRING	'ON'	AIA_IMG_AEC_MODE
AIAWVLEN	LONG	2	AIA_IMG_WAVELENGTH
CDELT1	DOUBLE	0.60000000	X Pixel size in CUNIT1 units
CDELT2	DOUBLE	0.60000000	Y Pixel size in CUNIT1 units
CUNIT1	STRING	'arcsec'	X pixel units
CUNIT2	STRING	'arcsec'	Y pixel units
CRPIX1	DOUBLE	-201.50000	[pix] X location of sun center in CCD
CRPIX2	DOUBLE	622.50000	[pix] Y location of sun center in CCD
CRVAL1	DOUBLE	0.0000000	[arcsec] X origin - center of the solar disk
CRVAL2	DOUBLE	0.0000000	[arcsec] y origin - center of the solar disk
CROTA2	DOUBLE	0.0000000	[deg] Angle between satellite N and solar N
DSUN_OBS	DOUBLE	1.4773886e+11	[m] Distance from SDO to Sun center
RSUN_OBS	DOUBLE	971.72022	[arcsec] of Sun = arcsin(RSUN_REF/DSUN_OBS)
RSUN_REF	DOUBLE	6.9600000e+08	[m] Reference radius of Sun
INST_ROT	DOUBLE	0.056433000	[deg] Master pointing CCD rotation wrt SDO Z axis
IMSCL_MP	DOUBLE	0.60075802	[arcsec/pixel] Master pointing image scale
OBS_VR	DOUBLE	-1972.2185	[m/s] Speed of observer in radial direction
OBS_VW	DOUBLE	28105.801	[m/s] Speed of observer in solar-W direction
OBS_VN	DOUBLE	-1553.5497	[m/s] Speed of observer in solar-N direction
XCEN	DOUBLE	328.80002	[arcsec] Ref pixel pointing, arc sec W from sun center
YCEN	DOUBLE	-246.90001	[arcsec] Ref pixel pointing, arc sec N from sun center
DEGRAD	STRING	degradation performed using calibration 'version 10'	Degradation correction
PCOUNT	LONG	0 No Group Parameters	
GCOUNT	LONG	1 One Data Group	
CTYPE1	STRING	'HPLN-TAN'	Type of X image coordinate axis
CTYPE2	STRING	'HPLT-TAN'	Type of Y image coordinate axis
END			

Table 3. Sample FITS header of an image extension.

B. DETAILS ON DATA ACQUISITION FOR REMOTE-DRMS SITES

In order to produce the AARP dataset, NWRA made extensive use of being a remote-SUMS/DRMS (Storage Unit Management System/Data Record Management System) site for the SDO mission. As such, NWRA can query, access, and, in some aspects, "mirror" the SDO databases in a manner directly analogous to the host institution (Stanford University) through the Stanford Joint Science Operations Center (JSOC). Many institutions have invested in this capability. To produce this curated dataset required numerous steps and considerations that we describe here, many of which are unique to the SUMS/DRMS system, but some were unexpected, hence worth providing to the community.

We did find this approach most tenable (*vs.* using the JSOC on-line interfaces including the `ssw_cutout_service.pro`, see Appendix C), in part because we could simultaneously process multiple HARP-based AARPs that were often present on the disk, reducing the I/O load and processing time significantly.

Thus, NWRA subscribed to three AIA data series: `aia.lev1` (the "data series"), `aia.lev1_euv_12s`, and `aia.lev1_uv_24s` (the "header series"). The first queries the actual image data, the latter two query the metadata or header information for EUV (94, 131, 171, 193, 211, 304, 335Å) and UV (1600, 1700Å) wavelengths respectively; the metadata alone, which comes with the subscriptions, is ≈100GB. In order to download and then access a complete image including its header information within the SUMS/DRMS system, both the data and header series must be queried; data within JSOC series are referred to by "prime keys" (often, but not exclusively, a reference time such as `T_REC`) then data are selected further by refining keyword searches. From servers hosting a remote-SUMS/DRMS system, querying the SUMS/DRMS database can be done using the command-line `show_info` (http://jsoc.stanford.edu/doxygen_html/group_show_info.html), which is callable from within other codes (*e.g.* `bash`, `IDL`, or `Python`) and the output used accordingly.

To construct an AARP data set, the needed image data are downloaded first then additional queries are constructed in order to access the correct entries in the SUMS/DRMS for analysis. Image data were batch-transferred using the Java Mirroring Daemon (JMD), written and implemented by the National Solar Observatories (http://docs.virtualsolar.org/wiki/jmd; http://vso.tuc.noao.edu/VSO/w/index.php/Main_Page). Using an example of the 13 m interval centered at 15:48:00 UT on 2011.01.01 for 94Å, the following steps are taken:

1. For 13 m of data at 72 s cadence (see Section 2.3), the "slot time" or first target image is 15:42:00. This somewhat generic timestamp is not accepted by `aia.lev1` series, so the prime keys are obtained by querying the `aia.lev1_euv_12` series:

```
$ show_info key="T_REC,T_OBS" "aia.lev1_euv_12s[2011-01-01T15:42:00/13m@72s][?
QUALITY=0 ?][?  EXPTIME>1.8 and EXPTIME<3.0 ?][?  WAVELNTH=94 ?]"
> T_REC T_OBS
> 2011-01-01T15:42:02Z 2011-01-01T15:42:03.57Z
> 2011-01-01T15:43:14Z 2011-01-01T15:43:15.57Z
> 2011-01-01T15:44:26Z 2011-01-01T15:44:27.57Z
...
```

   as `T_REC` is required for the header series (`aia.lev1_euv_12`) whereas `T_OBS` is needed for the data series (`aia.lev1`). Note that we specify not just wavelength and exposure time (to avoid AEC data), but the data-quality requirement as well. In point of fact, we would query *e.g.*

`2011.01.01_15:42:00/13m@72s` initially and try *e.g.* `2011.01.01_15:41:48/13m@72s` if fewer than the expected number of records were present, or if the difference between `T_REC, T_OBS` was greater than 18 s. Note that the wavelength can be used in the header series as a second prime key, but not in the data series, so we generally adopted the practice of invoking the keyword request for all.

2. Populate the local `aia.lev1` series with the requested full-disk data. To set up the JMD data-transfer request requires the `sunum` (storage unit unique identifier that is associated with a given data series query) and `recnum` (a record number corresponding to each record in the storage unit) available from the data series using the prime keys from above:

```
$ show_info -rS key="FSN,T_REC,T_OBS" "aia.lev1[2011-01-01T15:42:03.57Z/13m@72s][?
QUALITY=0 ?][?EXPTIME>1.8 and EXPTIME<3.0?][?WAVELNTH=94?]"
> recnum sunum FSN T_REC T_OBS
> 67456313 190140361 18250728 2011-01-01T15:42:04Z 2011-01-01T15:42:03.57Z
> 67456361 190140436 18250776 2011-01-01T15:43:16Z 2011-01-01T15:43:15.57Z
> 67456409 190140506 18250824 2011-01-01T15:44:28Z 2011-01-01T15:44:27.57Z
...
```

   Note here: there can be more than one record per `sunum` (depending on the series), the returned `T_REC` differs from that returned from the header series, and `sunum` information is not available from the header series.

   The Filtergram Sequence Number ("FSN") integers become important later, but can be queried and saved at this step. For each imaging instrument of SDO (*e.g. HMI or AIA*) the FSN is a *unique* number for each original image produced by the instrument; we use them to validate the correspondence between the two series. The `aia.lev1` series is not designed to accept multiple-entry requests ("`13m@72s`"), and when it fails, the queries must be handled individually.

3. We query for the metadata entries with the new `T_REC` to confirm the FSN number correspondence:

```
$ show_info key="FSN, T_REC,T_OBS" "aia.lev1_euv_12s[2011.01.01T15:42:04Z/13m@72s][?
QUALITY=0 ?][?EXPTIME>1.8 and EXPTIME<3.0?][?WAVELNTH=94?]"
> FSN T_REC T_OBS
> 18250728 2011-03-01T15:42:02Z 2011-03-01T15:42:03.57Z
> 18250776 2011-03-01T15:43:14Z 2011-03-01T15:43:15.57Z
> 18250824 2011-03-01T15:44:26Z 2011-03-01T15:44:27.57Z
...
```

4. Upon data transfer, we query the data series again, for the location of the files using the FSN numbers returned by the `aia_lev1_euv_12s`. Every remote-SUMS/DRMS system will have its own upper-level structure, but the `sunum` is consistent between them and is integral to the location:

```
$ show_info -qP 'aia.lev1[][18250728, 18250776, 18250824, ...  ]' seg=image_lev1
> /nwra/SUMS/SUM0/D190140361/S00000/image_lev1.fits
> /nwra/SUMS/SUM0/D190140436/S00000/image_lev1.fits
> /nwra/SUMS/SUM0/D190140506/S00000/image_lev1.fits
...
```

5. We proceed to prepare the data by specifying wavelength, etc. keywords and the file location (*e.g.*, `/nwra/SUMS/SUM0/D190140361/S00000`). Invoking `drot_map.pro`, degradation, field of view determination and extraction, coalignment, *etc.*, requires numerous additional keyword-queries through `show_info` to the `hmi.Mharp_720s` series. Along the processing, intermediate steps are saved as temporary `IDL` "idl.sav" files for subsequent assembly into the AARP extraction FITS files.

The data transfers and referencing were indeed time consuming to automate, as these series are constructed with different structure than, *e.g.* the HMI series. However, having the full-disk data locally was key while we formulated the spatial extraction details (see Section 2.2) and the sensitivity corrections (see Section 2.1), because we were not required to re-request full-disk data repeatedly as the respective approaches were refined.

C. NOTES ON USING THE `SOLARSOFTWARE CUTOUT SERVICE`

Our initial attempt, over 2014-2015, to construct the AARP timeseries HARP-congruent dataset relied on JSOC and its on-line interfaces including through the *Solar SoftWare* IDL package. We used the (SSW Freeland & Handy 1998) tool `ssw_cutout_service.pro` written by Sam Freeland at LMSAL. This approach is advantageous for many studies, as the AARP-required data can be downloaded and assembled without the effort and resource allocation necessary to subscribe to the voluminous DRMS series, instead utilizing LMSAL's own subscription. Further, the LMSAL-generated SSW routines are equipped to deal with the subtleties of AIA data processing, such as how to set the cadence to avoid AEC frames.

We automated various steps required to implement the cutout service. Drivers were written specifically to:

1. Get HMI metadata, using `show_info` to select desired times, and patch size/location coordinates:
```
SSWIDL> fovxa = abs(crsize1*cos(crota2)-crsize2*sin(crota2))*cdelt1
SSWIDL> fovya = abs(crsize2*cos(crota2)+crsize1*sin(crota2))*cdelt1
```

2. Parse the AIA metadata, and select start/end times to avoid AEC images, getting specified URLs:
```
SSWIDL> ssw_jsoc_time2data, t0, t1, index, urls, /urls_only, wave=wave, ds='aia.lev1_euv_12s'
```

3. Queue AIA cutout exports corresponding to HMI Active Region Patches (HARPs), for the given day using `ssw_cutout_service`, invoking the option of an email when the job is complete:
```
SSWIDL> ssw_cutout_service, t0n, t1n, query, stat, ref_time=tref, fovx=fovxa,
fovy=fovya, wave=wavelengths, ref_helio=ref_helio, instrument='aia',aec=0, cadence=cadencestr
description=descr, max_frames=10000, email=email, /RICE
```

4. Wait for the exports to complete, and move exported FITS files into place

5. Read in the FITS files for each HARP/wavelength/time

6. Call `aia_prep` to perform alignment/adjustments on each data cube:
```
SSWIDL> aia_prep, indexIn, dataIn, index, data, /cutout, index_ref=indexIn[refidx]
```

7. Save the final processed data cubes to FITS (and IDL .sav) files (one per HARP per day)

All of the above steps were automated, initiated by specifying a date. The 94, 131, 171, 193, 211, 304, and 335 Å filter data can be processed together, but the different cadence and data series means the 1600 (UV) must be processed separately.

To produce the full AARP database, however, we came to a few situations that nudged us to the approach in Appendix B. The main failure point stemmed from the (lack of) robustness of data retrieval and needed information not being available to easily detect and validate failures. Data transfer throttling – which can occur on either end of the transfer, and it is not always easy to assess where it occurs – plus "standard-issue network intermittency" often created problems accessing records at the query step. When records then came up absent, it was not straightforward to determine exactly which were missing.

Additionally, the cutout service proved to be slower than expected, primarily due to the need to request data by individual HARPs and individual days, which necessitated repeated full-disk I/O at the host institution when, for example, there were multiple HARPs on the same day and time step. We found significant speed up by internally accessing the full-disk data and performing cutouts *once* for all targets per day and time step.

The `ssw_cutout_service.pro` works very well for requests that focus on limited time periods or a small number of targets. It allows for fully customized fields-of-view, significant flexibility on cadence and filter selection, and provides up-to-date pre-processing. Some challenges that were faced during this exploration were beyond the purview of `ssw_cutout_service.pro` (*e.g.* the too-generic JSOC email content). Thus, while the exploration into using the `ssw_cutout_service.pro` involved significant investment, it also provided good education and preparation for the switch to relying upon NWRA's remote-DRMS/SUMS subscription (see Appendix B).

NWRA will share `ssw_cutout_service.pro`-related driver codes with the community upon request.

D. PATHOLOGY

The AIA Active Region Patch (AARP) Database (Dissauer et al. 2022b) follows the HMI Active Region Patch (HARP) definitions (e.g. Hoeksema et al. 2014). By default, all HARPs that were identified between 2010/06 – 2018/12 are included. However, the HARPs include "active pixel" maps within the bounding boxes that are unique to each HARP, even if two (or more) HARP bounding boxes overlap. We do not include the masks (such as are in the `hmi.Mharp_720s bitmap.fits` segment) as they would only refer, roughly, to the footpoints of the coronal structures. Hence, some AARPs may spatially overlap, or be fully included in one another. We do not remove or try to disentangle these situations as all, but do note that generally the "subsuming" situation is rare and generally for which one of the AARPs is very small.

However, we did find pathological HARPs that needed to be excluded, resulting primarily from periods when the HMI magnetograms were corrupted in such a way as to flag large sections of limb as "active pixels" which then propagated to a HARP definition even though there were otherwise no magnetic concentrations. Of note: the following HARP numbers were removed from the AARP Database:

`4276, 4280, 6712, 6713, 6849, 6851, 7207`

In addition, HARP `4225` was also initially defined with a corrupted bounding box but in fact there were two NOAA active regions within, although they took up only a small fraction of the original

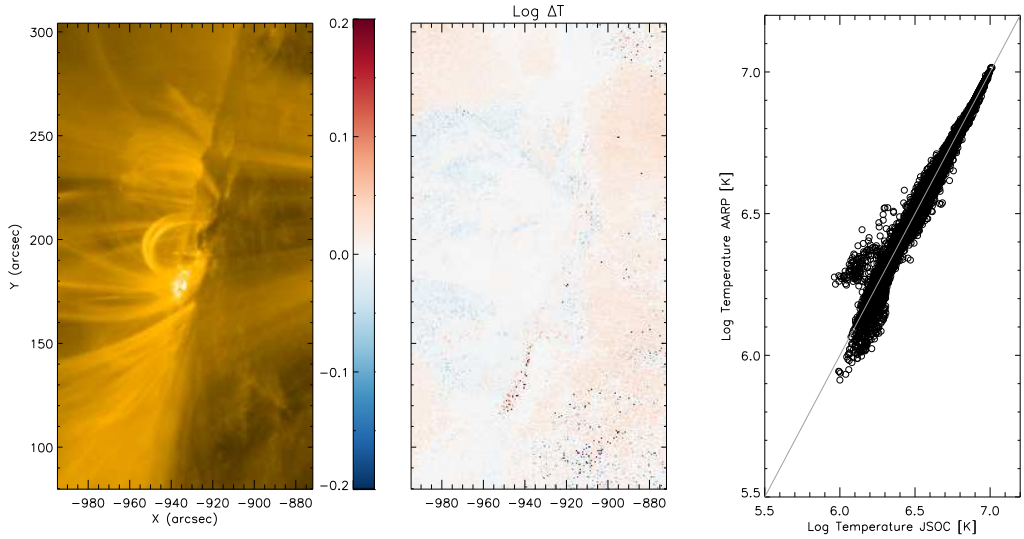


Figure 9. Demonstrating uncertainties in the DEM results using the degradation-corrected AARP dataset to reconstruct temperature, as compared to using the un-corrected JSOC downloaded image files ("ground truth"). Left: 171 Å image for context; Middle: $\log(T_{\rm AARP}/T_{\rm JSOC})$ map showing spatial differences between the two; Right: Correlation plot between $T_{\rm AARP}$ and $T_{\rm JSOC}$ in log scale. The gray line shows the 1:1 correspondence line.

bounding box. We redefined a HARP-appropriate bounding box according to the active pixel masks (`hmi.Mharp_720s bitmap.fits` segment, see Hoeksema et al. (2014)), and then extended that FOV as per described in Section 2.2 for inclusion in the AARP database.

E. DIFFERENTIAL EMISSION MEASURE VALIDATION

The AARP data set is corrected for instrument degradation. In general, we use the DEM code by Cheung et al. (2015) to calculate emission measure, temperature and density maps (Dissauer et al. 2022a). By definition, this particular DEM code handles instrument degradation by correcting the temperature response functions returned from `aia_get_response.pro` using the image observation times.

Here, we demonstrate that the DEM calculation using the AARP dataset as input, and turning "off" the degradation correction, provides similar results as using the original JSOC output where the degradation has not yet been performed and allowing the DEM code to make these corrections (Figure 9). The differences are displayed as both spatial maps and a correlation plot, and do not exceed the [-0.1,0.1] range in a pixel-by-pixel comparison. This error range is in agreement with uncertainties for the sparse solution of Cheung et al. (2015) when compared to a "ground truth" (*c.f*, their Figures 2 & 3).

Facilities: SDO (HMI and AIA); GOES (XRS)

Software: SolarSoft (Freeland & Handy 1998)

REFERENCES

Abramenko, V. I. 2005, ApJ, 629, 1141, doi: 10.1086/431732

Aggarwal, A., Schanche, N., Reeves, K. K., Kempton, D., & Angryk, R. 2018, ApJS, 236, 15, doi: 10.3847/1538-4365/aab77f

Al-Ghraibah, A., Boucheron, L. E., & McAteer, R. T. J. 2015, A&A, 579, A64, doi: 10.1051/0004-6361/201525978
Alipour, N., Mohammadi, F., & Safari, H. 2019, ApJS, 243, 20, doi: 10.3847/1538-4365/ab289b
Bamba, Y., Kusano, K., Imada, S., & Iida, Y. 2014, PASJ, 66, S16, doi: 10.1093/pasj/psu091
Barnes, G. 2007, ApJL, 670, L53, doi: 10.1086/524107
Barnes, G., & Leka, K. D. 2006, ApJ, 646, 1303, doi: 10.1086/504960
Barnes, W. T., Cheung, M. C. M., Bobra, M. G., et al. 2020, Journal of Open Source Software, 5, 2801, doi: 10.21105/joss.02801
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135, doi: 10.1088/0004-637X/798/2/135
Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, SoPh, 289, 3549, doi: 10.1007/s11207-014-0529-3
Bornmann, P. L., & Shaw, D. 1994, SoPh, 150, 127, doi: 10.1007/BF00712882
Braun, D. C. 2016, ApJ, 819, 106, doi: 10.3847/0004-637X/819/2/106
Canfield, R. C., Priest, E. R., & Rust, D. M. 1975, NASA STI/Recon Technical Report N, 76, 361
Centeno, R., Schou, J., Hayashi, K., et al. 2014, SoPh, 289, 3531, doi: 10.1007/s11207-014-0497-7
Cheung, M. C. M., Boerner, P., Schrijver, C. J., et al. 2015, ApJ, 807, 143, doi: 10.1088/0004-637X/807/2/143
Cho, K., Lee, J., Chae, J., et al. 2016, SoPh, 291, 2391, doi: 10.1007/s11207-016-0963-5
Del Zanna, G. 2013, A&A, 558, A73, doi: 10.1051/0004-6361/201321653
Dissauer, K., Leka, K. D., Barnes, G., & Wagner, E. L. 2022a, ApJ, in preparation
Dissauer, K., Leka, K. D., & Wagner, E. L. 2022b, NorthWest Research Associates (NWRA) Atmospheric Imaging Assembly (AIA) Active Region Patches (AARP) Database, V1, doi: 10.48322/50TD-NZ33
Freeland, S. L., & Handy, B. N. 1998, SoPh, 182, 497, doi: 10.1023/A:1005038224881
Galvez, R., Fouhey, D. F., Jin, M., et al. 2019, ApJS, 242, 7, doi: 10.3847/1538-4365/ab1005
Georgoulis, M. K. 2012, SoPh, 276, 161, doi: 10.1007/s11207-010-9705-2
Georgoulis, M. K., & Rust, D. M. 2007, ApJL, 661, L109, doi: 10.1086/518718
Golding, T. P., Leenaarts, J., & Carlsson, M. 2017, A&A, 597, A102, doi: 10.1051/0004-6361/201629462
Hagyard, M. J., Smith, J. B. J., Teuber, D., & West, E. A. 1984, SoPh, 91, 115
Harra, L. K., Matthews, S., Culhane, J. L., et al. 2013, ApJ, 774, 122, doi: 10.1088/0004-637X/774/2/122
Hoeksema, J. T., Liu, Y., Hayashi, K., et al. 2014, SoPh, 289, 3483, doi: 10.1007/s11207-014-0516-8
Imada, S., Bamba, Y., & Kusano, K. 2014, PASJ, 66, S17, doi: 10.1093/pasj/psu092
Inoue, S., Kusano, K., Büchner, J., & Skála, J. 2018, Nature Communications, 9, 174, doi: 10.1038/s41467-017-02616-8
Ishiguro, N., & Kusano, K. 2017, ApJ, 843, 101, doi: 10.3847/1538-4357/aa799b
Jonas, E., Bobra, M., Shankar, V., Todd Hoeksema, J., & Recht, B. 2018, SoPh, 293, #48, doi: 10.1007/s11207-018-1258-9
Joshi, B., Veronig, A. M., Lee, J., et al. 2011, ApJ, 743, 195, doi: 10.1088/0004-637X/743/2/195
Komm, R., Ferguson, R., Hill, F., Barnes, G., & Leka, K. D. 2011, SoPh, 268, 389, doi: 10.1007/s11207-010-9552-1
Kontogiannis, I., Georgoulis, M. K., Guerra, J. A., Park, S.-H., & Bloomfield, D. S. 2019, SoPh, 294, 130, doi: 10.1007/s11207-019-1523-6
Korsós, M. B., Baranyi, T., & Ludmány, A. 2014, ApJ, 789, 107, doi: 10.1088/0004-637X/789/2/107
Krall, K. R., Smith, Jr., J. B., Hagyard, M. J., West, E. A., & Cummings, N. P. 1982, SoPh, 79, 59, doi: 10.1007/BF00146973
Kusano, K., Iju, T., Bamba, Y., & Inoue, S. 2020, Science, 369, 587, doi: 10.1126/science.aaz2511
Leamon, R. J., McIntosh, S. W., & Title, A. M. 2022, Frontiers in Astronomy and Space Sciences, 9, 886670, doi: 10.3389/fspas.2022.886670
Leka, K. D., & Barnes, G. 2003a, ApJ, 595, 1277
—. 2003b, ApJ, 595, 1296
—. 2007, ApJ, 656, 1173, doi: 10.1086/510282
Leka, K. D., Barnes, G., & Wagner, E. L. 2018, Journal of Space Weather and Space Climate, 8, A25, doi: 10.1051/swsc/2018004
Leka, K. D., Dissauer, K., Barnes, G., & Wagner, E. L. 2022, ApJ, accepted

Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, SoPh, 275, 17, doi: 10.1007/s11207-011-9776-8
Li, J., Mickey, D. L., & LaBonte, B. J. 2005, ApJ, 620, 1092, doi: 10.1086/427205
Mango, S. A., Bohlin, J. D., Glackin, D. L., & Linsky, J. L. 1978, ApJ, 220, 683, doi: 10.1086/155952
Mason, J. P., & Hoeksema, J. T. 2010, ApJ, 723, 634, doi: 10.1088/0004-637X/723/1/634
McAteer, R. T. J., Gallagher, P. T., & Ireland, J. 2005, ApJ, 631, 628, doi: 10.1086/432412
McIntosh, P. S. 1990, SoPh, 125, 251
Nishizuka, N., Sugiura, K., Kubo, Y., et al. 2017, ApJ, 835, 156, doi: 10.3847/1538-4357/835/2/156
O'Dwyer, B., Del Zanna, G., Mason, H. E., Weber, M. A., & Tripathi, D. 2010, A&A, 521, A21, doi: 10.1051/0004-6361/201014872
Park, S.-H., Guerra, J. A., Gallagher, P. T., Georgoulis, M. K., & Bloomfield, D. S. 2018, SoPh, 293, 114, doi: 10.1007/s11207-018-1336-z
Park, S.-H., Leka, K. D., & Kusano, K. 2021, ApJ, 911 press, 79, doi: 10.3847/1538-4357/abea13
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3, doi: 10.1007/s11207-011-9841-3
Qiu, J., & Cheng, J. 2017, ApJL, 838, L6, doi: 10.3847/2041-8213/aa6798
Reinard, A. A., Henthorn, J., Komm, R., & Hill, F. 2010, ApJL, 710, L121, doi: 10.1088/2041-8205
Sawyer, C., Warwick, J. W., & Dennett, J. T. 1986, Solar Flare Prediction (Boulder, CO: Colorado Assoc. Univ. Press)
Scherrer, P. H., Schou, J., Bush, R. I., et al. 2012, SoPh, 275, 207, doi: 10.1007/s11207-011-9834-2
Schonfeld, S. J., White, S. M., Hock-Mysliwiec, R. A., & McAteer, R. T. J. 2017, ApJ, 844, 163, doi: 10.3847/1538-4357/aa7b35
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229, doi: 10.1007/s11207-011-9842-2
Seki, D., Otsuji, K., Isobe, H., et al. 2017, ApJL, 843, L24, doi: 10.3847/2041-8213/aa7559
SILSO World Data Center. 2010-2018, International Sunspot Number Monthly Bulletin and online catalogue
Smith, J. B. J., Neidig, D. F., Wiborg, P. H., et al. 1996, in ASP Conference Ser., Vol. 95, Solar Drivers of Interplanetary and Terrestrial Disturbance, 54–65
Sterling, A. C., & Moore, R. L. 2001, ApJ, 560, 1045, doi: 10.1086/322241
Sterling, A. C., Moore, R. L., & Freeland, S. L. 2011, ApJL, 731, L3, doi: 10.1088/2041-8205/731/1/L3
Thiemann, E. M. B., Chamberlin, P. C., Eparvier, F. G., & Epp, L. 2018, SoPh, 293, 19, doi: 10.1007/s11207-018-1244-2
Thompson, W. T. 2006, A&A, 449, 791, doi: 10.1051/0004-6361:20054262
Threlfall, J., Neukirch, T., & Parnell, C. E. 2017, SoPh, 292, 45, doi: 10.1007/s11207-017-1060-0
Ugarte-Urra, I., Warren, H. P., Upton, L. A., & Young, P. R. 2017, ApJ, 846, 165, doi: 10.3847/1538-4357/aa8597
Wang, J., Shi, Z., Wang, H., & Lü, Y. 1996, ApJ, 456, 861
Warren, H. P., Winebarger, A. R., & Brooks, D. H. 2012, ApJ, 759, 141, doi: 10.1088/0004-637X/759/2/141
Welsch, B. T., Li, Y., Schuck, P. W., & Fisher, G. H. 2009, ApJ, 705, 821, doi: 10.1088/0004-637X/705/1/821
Woods, M. M., Harra, L. K., Matthews, S. A., et al. 2017, SoPh, 292, 38, doi: 10.1007/s11207-017-1064-9
Worden, J., Woods, T. N., Neupert, W. M., & Delaboudinière, J.-P. 1999, ApJ, 511, 965, doi: 10.1086/306693
Zhang, Q. M., Su, Y. N., & Ji, H. S. 2017, A&A, 598, A3, doi: 10.1051/0004-6361/201629477
Zirin, H., & Liggett, M. 1987, SoPh, 113, 267
Zirin, H., & Marquette, W. 1991, SoPh, 131, 149, doi: 10.1007/BF00151751
Zirin, H., & Tanaka, K. 1973, SoPh, 32, 173