Preemergence Signatures of Horizontal Divergent Flows in Solar Active Regions

T. Rees-Crockford1, C. J. Nelson1,2, and M. Mathioudakis1

1 Astrophysics Research Centre, School of Mathematics and Physics, Queen’s University Belfast, Belfast BT7 INN, UK; t.reescrockford@qub.ac.uk
2 European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands

Received 2022 April 13; revised 2022 August 11; accepted 2022 October 9; published 2022 November 25

Abstract

Solar active regions (ARs) play a fundamental role in driving many of the geoeffective eruptions, which propagate into the solar system. However, we are still unable to consistently predict where and when ARs will occur across the solar disk by identifying preemergence signatures in observables such as the Doppler velocity (without using helioseismic methods). Here we aim to determine the earliest time at which preemergence signatures, the horizontal divergent flow (HDF) in particular, can be confidently detected using data from the Solar Dynamics Observatory’s Helioseismic and Magnetic Imager. Initially, we follow previous studies using the thresholding method, which searches for significant increases in the number of pixels that display a specific line-of-sight velocity. We expand this method to more velocity windows and conduct a basic parameter study investigating the effect of cadence on the inferred results. Our findings agree with previous studies with 37.5% of ARs displaying an HDF, with average lead times between the HDF and flux emergence of 58 minutes. We present a new potential signature of flux emergence, which manifests as cadence-independent transient disruptions to the amplitudes of multiple velocity windows and recover potential preemergence signatures for 10 of the 16 ARs studied, with lead times of 60–156 minutes. Several effects can influence both the estimated times of both HDF and flux emergence suggesting that one may need to combine Doppler and magnetic field data to get a reliable indicator of continued flux emergence.

Unified Astronomy Thesaurus concepts: Solar magnetic flux emergence (2000); Solar active regions (1974); Solar active region magnetic fields (1975); Solar active region velocity fields (1976)

Supporting material: figure sets

1. Introduction

Active regions (ARs) are locations where large amounts of magnetic flux break through the solar surface from the solar interior and are thought to occur due to buoyancy acting upon flux (Parker 1955; Acheson 1979) that originates within the deep convection zone. Once sufficient flux breaks through the solar surface, it typically forms into a bipolar and a surrounding plage, with the regions of negative and positive fields aligning with Hale’s and Joy’s laws (Hale et al. 1919; in reality a huge variety of different permutations exist in terms of AR structure). In addition to the detection of flux in magnetograms, velocity fields accompanying the newly emerging magnetic field have also been identified (Kawaguchi & Kitai 1976, and references therein), corresponding to the bulk movement of plasma both contained within and around the magnetic flux. Helioseismological techniques have been used to determine that velocity fields could also be recovered preemergence (Braun 1995; Zharkov & Thompson 2008; Ilonidis et al. 2011; Birch et al. 2013) indicating that it may be possible to predict AR emergence, though the recovered velocities and time lead varied with the AR and technique. More recently, Toriumi et al. (2012, 2014) identified signatures of preemergence (referred to as horizontal divergent flows, or HDFs) in simple Doppler measurements sampled by the Solar Dynamics Observatory’s Helioseismic and Magnetic Imager (SDO/HMI; Scherrer et al. 2012); however, their statistical results were limited to HDFs detected in a single velocity window between [1 km s\(^{-1}\), 1.5 km s\(^{-1}\)]. In this work, we extend the results of Toriumi et al. (2012) to a larger range of velocity windows and discuss the effects of cadence on the methods used.

Much of our understanding of AR emergence has been developed through analysis of realistic numerical simulations over the past few decades. It is now thought that emergence could take place in two steps, with the first being the flux tube rising to just below the solar surface, and the second being when the flux tube actually emerges into the atmosphere (Archontis et al. 2004). As part of this second step, the flux tube would expand laterally as the plasma above it would impede its upward motion. Flows associated with this lateral expansion have been seen in simulations (Cheung et al. 2007, 2010), where the lateral expansion of the flux scaled with the magnetic field and the plasma density such that \(B \sim \rho^{1/2}\). It was noted by Toriumi & Yokoyama (2010) that deceleration of the emerging flux in the chromosphere, rather than the upper convection zone (i.e., beneath the surface), may occur due to the structure of the emerging flux, thereby allowing faster draining of the plasma and thus a faster emergence. These authors considered the emergence of said flux sheet from a depth of 20 Mm below the surface and suggested a threshold for two-step emergence (over failed emergence) at \(10^{21}–10^{22}\) Mx.

Further 3D work by Toriumi & Yokoyama (2013) studied the effect that varying the field strength, twist, and radius of curvature of a flux tube has on flux emergence. They found that the rise speed of the flux was strongly dependent on the magnetic field strength, but only weakly on the twist and radius of the tube. In their strong field cases, they found that the increased field strength led to faster emergence, and thus faster but shorter HDFs as the plasma has less time to drain before emergence. No correlation was found between the duration of the HDF and the twist, but they did note that higher twists did also have a positive correlation...
with the maximum speed of the HDF. They suggested that this was due to the twist allowing the flux tube to remain more intact during the emergence, thereby allowing more of it to emerge at once. This work was expanded upon by Syntelis et al. (2019), who also investigated the length of the emerging portion of the flux tube, and the scaling of $B \sim \rho^{a}$. They found that there were also geometric parameters that could define whether flux would successfully emerge, with a larger tube radius causing more rapid emergence. Furthermore, tubes with shorter buoyant sections emerged more slowly than those with longer buoyant sections, with emergence not being possible below a certain length (scaled to other parameters; see their Section 3.3.3).

Observational evidence of the flows predicted by these numerical simulations has also been acquired over recent years using simple Dopplergrams. For instance, Grigor’ev et al. (2007) used Solar and Heliospheric Observatory Michelson Doppler Imager data to determine the line-of-sight (LOS) velocity field of an emerging AR. Using these observations, they found upward velocities with a mean of $-0.23 \text{ km s}^{-1}$ and a maximum of $\sim 2 \text{ km s}^{-1}$ over the first 2 hr of the emergence. Although they only studied data preceding the emergence by 3 hr, they also noted the appearance of velocities around $-0.4 \text{ km s}^{-1}$ just before the emergence of the first magnetic flux. They attributed these motions to the emergence of the flux tube through the photosphere. Early observational evidence of two-step flux emergence have been reported by Otsuji et al. (2011), who found the apparent deceleration and spreading of flux as it emerged into the chromosphere. The first observational evidence of the lateral flows reported by Cheung et al. (2010) was presented by Toriumi et al. (2012). Those authors studied an AR close to the solar limb, reasoning that, at a sufficient angle, one should be able to see the horizontal component of these motions in the wings of LOS Doppler profiles before flux emergence. Following their analysis, Toriumi et al. (2012) found significant increases in the number of pixels, which displayed horizontal velocities around $0.6–1.5 \text{ km s}^{-1}$ preceding the emergence by 100 minutes. Additional evidence of these flows was found in the follow-up statistical work of Toriumi et al. (2014), who found HDFs in 13 of 23 events. They found an average maximum HDF velocity (see their Section 3.2) of $\sim 3.1 \text{ km s}^{-1}$, and an average time difference of 61 minutes between the onset of these motions and the emergence of the flux. Observationally, we must of course be careful to distinguish what different authors mean by emergence. In some articles, emergence is used to refer to the initial breach of flux into the photosphere, whereas other authors use it to describe the entire time period before the AR reaches its maximum flux. Here we define emergence as the initial breach of the flux above a threshold value in order to provide comparable results with Toriumi et al. (2012, 2014).

In this article, we build on the work of Toriumi et al. (2012, 2014) in order to investigate whether preemergence HDF signatures can be detected in a wider variety of velocity bins. Additionally, we study the effect of cadence on our ability to detect a clear HDF, thereby investigating whether this method is scalable between different instruments. Our work is structured as follows: In Section 2 we detail our AR selection and data processing methods; in Section 3 we present our results; in Section 4 we discuss our results in the context of the literature; we draw our conclusions in Section 5.

### Table 1

Properties of the 16 P0 Active Regions Studied

| NOAA | HARP | Schunker | Lon. | Lon. | Lat. |
|------|------|----------|------|------|------|
| #    | T_{FIRST} & $T_{F}$ | $T_{S}$ | $T_{G}$ | $T_{G}$ |
| 11066 | 2010.05.02, 23:12 | 23:48 | -30.9 | -17.4 | 26.7 |
| 11072 | 2010.05.20, 16:24 | 17:12 | -50.4 | -36.7 | 15.1 |
| 11076 | 2010.05.31, 04:12 | 06:24 | -30.2 | -17.4 | 19.4 |
| 11088 | 2010.07.11, 07:48 | 08:36 | -64.0 | -50.4 | 20.1 |
| 11114 | 2010.10.14, 04:12 | 04:12 | 18.3 | 23.2 | 20.8 |
| 11116 | 2010.10.16, 19:24 | 22:48 | -15.2 | -03.2 | 22.3 |
| 11122 | 2010.11.06, 00:48 | 01:12 | -25.5 | -11.3 | 13.8 |
| 11130 | 2010.11.27, 15:12 | 18:12 | -32.4 | -19.9 | 13.6 |
| 11136 | 2010.12.24, 07:48 | 08:24 | 20.4 | 34.0 | 21.4 |
| 11142 | 2010.12.31, 09:00 | 09:24 | -71.5 | -57.4 | 13.8 |
| 11148 | 2011.01.17, 02:00 | 02:24 | 08.0 | 21.5 | 27.6 |
| 11294 | 2011.09.11, 01:24 | 04:12 | -65.3 | -52.8 | 16.4 |
| 11400 | 2012.01.14, 02:00 | 02:00 | -24.5 | -10.3 | 13.9 |
| 11414 | 2012.02.04, 09:24 | 09:24 | -03.5 | 10.8 | -05.4 |
| 11472 | 2012.04.29, 04:12 | 05:24 | -66.1 | -53.1 | 28.3 |
| 11523 | 2012.07.11, 23:12 | 23:24 | -58.4 | -44.5 | 27.5 |

**Notes.**

- $^a$ Time of initial detection in HARP.
- $^b$ HARP LON_FWT differentially rotated to start of data set.
- $^c$ LON_FWT at Schunker start time.
- $^d$ As b, for latitude.

### 2. Methods

#### 2.1. Data Selection and Processing

In order to study the LOS magnetic field strengths and Doppler velocities of ARs in the solar photosphere, we use data obtained by the SDO/HMI instrument (using the 6173 Å line). The data are provided in the form of full-disk magnetograms (hmi.M_45s) and Dopplergrams (hmi.V_45s) with a spatial resolution of 1’ and a 45 s cadence. These parameters are sufficient for studying large-scale, long-lived solar phenomena such as ARs. We use data keywords provided by the Space-weather HMI Active Region Patches (SHARP) hmi.sharp_720s series (Bogart 2007) in order to identify the spatial locations of ARs, used for derotating the data to the relevant times. For our subsequent analysis, we choose the ARs defined as “P0” (“emergence into a very quiet region”) by Schunker et al. (2016) as our targets. Of the 21 ARs defined as P0 in that work, 16 had data available at the time of our analysis. In Table 1, we provide the basic information for these ARs, including timings and relevant spatial coordinates. Note that all times presented here are in International Atomic Time (BIPM 2019). Although Schunker et al. defined their emergence time as “the time when the absolute flux, corrected for line-of-sight projection, reaches 10% of its maximum value over a 36 hr interval following the first appearance of the sunspot (or group) in the NOAA record,” we, as previously stated, define the emergence time as the first point at which the number of pixels within a specific magnetic field strength window increases over a threshold value of 1σ in order to better align our work with that of Toriumi et al. (2012). In Figure 1 we show the tracked location of each AR studied here.

In this article, we track the number of pixels with specific magnetic field strengths and Doppler velocities within a given region of interest (ROI). The ROI is defined by a 140″ × 140″ box centered on the [1:] HARP keywords of LON_FWT and LAT_FWT (in Stonyhurst coordinates), which are defined as the flux-weighted center of the AR at the time of the initial
detection as recorded [1:] in the HARP keyword T_FRST1 [1:] (defined as the first recorded time of this HARP number). These coordinates are then differentially rotated back to the appropriate position at each sampled time. The data used here cover the 24 hr period preceding that of the Schunker emergence time for each event. We use either a 45 s or a 12 minute cadence depending on the section, with the studied cadence being noted explicitly for each result. [2:] We use the 45 s cadence on the three ARs common to both this analysis and Toriumi et al. (2012) to allow for comparison between them. This also allows us to identify if any differences between the methods could cause significant differences in the results. This is detailed further in Section 3.1. The lower 12 minute cadence is chosen to both conduct a basic parameter study about the effect of the cadence on the method and due to reasons of data storage and processing speed. Our data extraction method, based on that of Toriumi et al. (2012) with some differences, is performed using Python (Numpy: Harris et al. 2020; Sunpy: The SunPy Community et al. 2020; Astropy: Astropy Collaboration et al. 2013, 2018). The basic outline of the method is included in Figure 2, with this figure being referenced throughout the following method:

1. We begin by retrieving the HARP coordinate keywords for the relevant AR (E2).

2. We then perform a running average from the Dopplergram data sampled over 48 minutes (E4). Unlike Toriumi et al. (2012) who do a running average over 30 minutes, we average over 48 minutes in order to maintain consistency between the higher (45 s) and lower (12 minute) cadences considered in our analysis. [3:] The coordinates for each time step are then differentially rotated to the appropriate time (E5), and the 140° × 140° submap of the ROI is created (E6). Unlike in Toriumi et al. (2012, 2014), here we do not perform a Postel projection. This is detailed further below.

4. A histogram is taken of the data for each time step (E7). In the case of the Dopplergram, this histogram is shifted to have its peak at zero (E8). This is done by subtracting the radial velocity of the satellite ([3:] as defined in the FITS header keyword OBS_VR), and then the mean velocity of the histogram to mitigate part of the effect of the satellite motion. This deviates from the method used by Toriumi et al. (2014), and will be discussed below. For magnetograms, we take the histogram of the absolute value of the signed flux.

5. A reference profile histogram is created using the average of the first 3 hr worth of data (E10). The standard deviation, σ, in each histogram bin across that time is then calculated (E10).
6. The reference profile is subtracted from each histogram across the entire time range to create residual profiles, which are the profiles presented in this work (E11).

As mentioned in Step 3, we do not perform a Postel projection. This is primarily because we wish to analyze the data while minimizing any potential artifacts introduced by processing. Additionally, as mentioned in Step 4, we deviate slightly from the reduction method used by Toriumi et al. (2014), as outlined in their Section 2.2, as again we wish to minimize the number of post-processing steps applied to the data. In that sense, we believe that it is more appropriate to simply subtract the mean as is performed in Toriumi et al. (2012). There is also another issue in Step 4 that we must address. While we attempt to account for the orbital motion of the satellite, we cannot fully remove it. This is primarily because we wish to analyze the magnetic field strengths within a specific window that display magnetic field strengths within a specific window that exceed the averaged standard deviation of that set of bins calculated from the time period used to create the reference profile. For our magnetic profiles we create bins with widths of 100 G (i.e., over the interval [lower bound, lower bound + 100] G), over the range 100–800 G (5:1 using the absolute value of the magnetic field). We exclude the [0, 100] G set due to the inherent noise; the latter variance and standard deviation. Once $T_{DE}$ has been determined from the relevant magnetic profiles, we then begin searching the Doppler profiles for an HDF, which has a time denoted by $T_{DE}$, using one of two methods (M4, D4). This smoothing has been applied to avoid biasing the method against cases where the profile dips below the threshold for only one frame.  

2.2. Creating Data Products for Analysis

The purpose of our analysis is to determine the earliest time at which we can detect HDF signatures in Dopplergrams. In order to do this we must first determine the time of magnetic flux emergence, $T_{FE}$ (see left side of Figure 3). As previously discussed in Toriumi et al. (2012), the time of emergence is defined as the time at which the number of pixels that display magnetic field strengths within a specific window exceeds the averaged standard deviation of that set of bins calculated from the time period used to create the reference profile. For our magnetic profiles we create bins with widths of 100 G (i.e., over the interval [lower bound, lower bound + 100] G), over the range 100–800 G (5:1 using the absolute value of the magnetic field). We exclude the [0, 100] G set due to the inherent noise; the latter variance and standard deviation. Once $T_{DE}$ has been determined from the relevant magnetic profiles, we then begin searching the Doppler profiles for an HDF, which has a time denoted by $T_{DE}$, using one of two methods (M4, D4). This smoothing has been applied to avoid biasing the method against cases where the profile dips below the threshold for only one frame.  

As mentioned in Step 3, we do not perform a Postel projection. This is primarily because we wish to analyze the data while minimizing any potential artifacts introduced by the underlying systemic reasons (see Section 3.1 in Schuck et al. 2016) and the lack of tested and verified libraries within sunpy. Furthermore, the Sun’s rotation creates an asymmetry between the east and west limbs, as one moves toward and away from the observer, proportional to the differential rotation at any given latitude. As this trend reverses across the disk, we must also be aware of the effect when we take the reference profile. Furthermore, the nonradial components of the motion of the satellite and additional instrumental effects (Schuck et al. 2016) will further modify the underlying distribution, and thus how the subsequent residual profiles will evolve. For this work, these effects manifest as follows: first, there is an underlying oscillatory behavior to the Doppler profiles; second, as the longitude of the studied AR changes relative to its reference profile, it will experience inherent changes to the residual profile. Ultimately, however, the underlying trend will provide the backdrop against which we are looking, but it will not dictate whether we are capable of seeing the transient features we are looking for. This is because the satellite movement will vary smoothly over the course of the observation (24 hr) and will have only minor effects over the time ranges considered for the transient changes studied here (0–3 hr). In addition, we believe that the removal of the radial component of the satellite motion and the zero-shift (see point 4 in the method outline) of the histogram results in velocity profiles are sufficient for this analysis. In addition, these profiles have been smoothed using the `numpy.convolve` function of width 5 in “valid” mode (M4, D4). This smoothing has been applied to avoid biasing the method against cases where the profile dips below the threshold for only one frame. Note that this smoothing is applied to the resultant profiles, not the data themselves.

![Figure 2. Flowchart of the data extraction method for both the magnetic field strength and Doppler velocity data sets.](image)
magnetogram data by first summing the number of pixels within velocity bins, which cover the interval [lower bound, lower bound+0.1) km s$^{-1}$. Velocity windows are then created, as in Toriumi et al. (2014), mirrored around 0 km s$^{-1}$, each with six bins. This differs slightly from the work of Toriumi et al. (2014) who used eight bins per window. The specific velocity windows studied at any time are indicated in all relevant figures (see, e.g., the legend of Figure 4). We chose six bins instead of eight as this increases the number of windows we can compare against each other and gives us more information at lower Doppler velocities. [12:] In this work we do not explicitly distinguish between the vertical and horizontal velocity components, we nonetheless deal with them implicitly. This is done through the consideration of all velocity windows, and the assumption that either there will be a recoverable component of the emerging flux at any longitude, or that the emergence will cause sufficient disruption to the underlying distribution to be measurable.

It should be noted that the flowcharts presented in Figure 3 assume a “successful” determination of $T_{DE}$ in at least one Doppler velocity window. In the case that this is not possible, no times are recorded for that AR. This is not of concern here as there is a sufficient number of windows in both the magnetograms and Dopplergrams to allow the determination of $T_{DE}$ for each AR, and as such it will not be considered further.

In order to better outline our method for the reader, and to minimize the introduction of any processing artifacts. Second, while we do not explicitly distinguish between the vertical and horizontal velocity components, we nonetheless deal with them implicitly. This is done through the consideration of all velocity windows, and the assumption that either there will be a recoverable component of the emerging flux at any longitude, or that the emergence will cause sufficient disruption to the underlying distribution to be measurable.
provide a more direct comparison to the work of Toriumi et al. (2012, 2014), we include an example of this method applied to full-cadence (45 s) data in Section 3.1.

3. Results

3.1. Method Comparison

We initially apply our method to three ARs using full-cadence (45 s) magnetic field and Dopplergram data. The three ARs chosen (ARs 11066, 11072, and 11076) were also studied by Toriumi et al. (2014) meaning we are able to identify whether the slight differences in our methodology introduce any major changes to the results. For these three ARs, Toriumi et al. (2014) found $T_{FE}$ values of 20:30 UT, 14:51 UT, and 16:31 UT, respectively, for the magnetic field strength window spanning [200, 300] G (assuming these are “typical” ARs). Through our analysis, we find values of 22:35 UT, 14:06 UT, and 18:53 UT, respectively, from the same magnetic field strength window. We do expect to report slightly different emergence times due to the minor variations in our methodology, and we consider these differences of around 2 hr to be acceptable (given the timescales of the processes studied here). For $T_{DET}$, Toriumi et al. (2014) reported values of 19:25 UT, 13:05 UT and 17:08 UT (note this is after $T_{FE}$ and, as such, no HDF was reported) for a velocity window spanning ±[1, 1.5] km s$^{-1}$. From our analysis, and using a velocity window spanning ±[0.9, 1.2] km s$^{-1}$ we find values of 19:16 UT and 11:56 UT for ARs 11066 and 11072, but find no HDF for AR 11076, agreeing with the results of Toriumi et al. (2014). Overall, we are confident that our method is providing accurate results [6] that are comparable to the previous literature despite the slight changes to the methodology, which will allow us to study the presence of HDF signatures in a wider range of velocity windows.

We plot the evolution of the reported magnetic and velocity windows at full cadence in Figure 4 for each of the three ARs. The red lines denote the standard deviation normalized number of pixels within the appropriate magnetic field strength window, while the dashed blue lines plot the same for the velocity window. The horizontal dashed lines denote the 1σ level used to define the

Figure 4. Evolution of the relevant magnetic field (red lines) and Doppler velocity (dashed blue lines) windows for the three ARs compared to Toriumi et al. (2014; as discussed in Section 3.1). The counts have been normalized against the standard deviations calculated from the initial 3 hr background time. The flux emergence above 1σ is clear in each event (denoted by the vertical solid black lines), with the presence of the HDF also being apparent in the Doppler data for ARs 11066 and 11072 (denoted by the vertical dotted-dashed black lines). No HDF is detected for AR 11076.
threshold values in this study. For AR 11066 (top panel), the HDF is clearly present (indicated by the vertical dotted–dashed black line) as the large rise in the dashed blue line several hours before the emergence of the flux (which takes place right at the end of the studied time period; denoted by the solid vertical black line). For AR 11072 (middle panel), the HDF begins only a few minutes before the emergence of the flux with the dashed blue and red lines seemingly increasing around the same time. For AR 11076 (bottom panel), the flux clearly begins to emerge around half way through the studied time period, but no associated HDF is detected.

Given the inherent errors in the definition of the flux emergence time, we also investigated the \( T_{FE} \) and \( T_{DET} \) values calculated from lower-cadence data in order to understand whether HDF signatures could be detected. This acts as a parameter study aimed at establishing whether this method is scalable to lower-cadence data. In Figure 5, we plot the equivalent to Figure 4 but for the lower 12 minute cadence. Similar behavior is observed between the high-cadence and low-cadence plots for each of these three ARs. For \( T_{FE} \), we now find values of 20:48 UT, 12:12 UT, and 18:48 UT for ARs 11066, 11072, and 11076, respectively. These values are also within the same “error” range as the full-cadence data studied previously. For \( T_{DET} \), we find 19:36 UT, 11:48 UT, and 16:48 UT for the three ARs. The main difference between these results and the high-cadence results is the return of a potential HDF signature for AR 11076. We note that this is only a small bump in the velocity bin window, which falls back below the 1σ level after three frames, indicating it may not be a true HDF.

An initial comparison may suggest that the differences in the results for \( T_{FE} \) between the high- and low-cadence are quite
large (several hours); however, close examination of image sequences shows that both are returning results close to the early (ill-defined) onset of emergence. There are also several unrelated processes occurring within the FOVs, which may explain the small differences in times reported. In the image sequences of AR 11066, we see the fragmentation and cancellation of a decaying bipole; however, this occurs at the edge of the FOV leaving some flux just outside the FOV during the three hours used for referencing. Thus, an increase in the weakest field strength bin and a co-temporal decrease in the other bins is perhaps to be expected as the flux decays; however, this will effect the standard deviations returned differently depending on the frames sampled. This flux does not appear to be related to the emerging flux; however, its presence in the FOV could explain small (several hour) differences in the emergence time reported here. For AR 11072, the background time contains some already emerging weak flux, which appears to separate and grow during this time. This means that we are potentially seeing emergence signatures in our background at the 1σ level.

Overall, it is clear that numerous effects will influence the values returned for \( T_{FE} \) and \( T_{DET} \), including but not limited to preprocessing steps applied to the data, the time period selected as the background, and the properties (e.g., cadence) of the data studied. Given these issues, it is unclear whether these increases in the number of pixels within specific velocity windows, known as the HDF, can be used as a predictive tool on its own for determining the number of pixels within specified velocity windows. Nonetheless, we see general agreement between these increases in the HDF and the properties of each AR studied. In these plots, solid lines denote that

### 3.2. Statistical Analysis

#### 3.2.1. Magnetic Field Data

In order to further progress toward the aims of this work, we now conduct a statistical analysis of the thresholding method applied to all 16 ARs using the ±[0.9, 1.2, 1.2] km s\(^{-1}\) velocity window using the 12 minute cadence data. Initially, we study the magnetogram profiles to determine the time of magnetic flux emergence, \( T_{FE} \), for each AR. The results of this analysis are presented in the second column of Table 2. For a better comparison with Toriumi et al. (2012), we report the [200, 300] G magnetic field strength window unless otherwise specified. In the case that no threshold (1σ) was calculable due to the counts in any specific magnetic field strength window being zero across the reference time range, we instead find the last time the profile is at zero counts. This was usually only necessary for sets above 500 G and, as such, has no significant influence on the results presented in Table 2.

In Figure 6, we present the time profiles of sets of absolute magnetic flux for three example ARs (namely, NOAA ARs 11066, 11114, and 11130) for each of the seven magnetic field strength windows. These plots help detail how we select an appropriate \( T_{FE} \) for each AR. In these plots, solid lines denote that

| NOAA # | \( \dot{B} \) | \( V_{DET} \) | \( \Delta T \) | \( V_{DET} \) | \( \Delta T \) |
|--------|--------|---------|--------|---------|--------|
| 11066  | 20:48  | 19:36   | 72     | 18:12   | 156    |
| 11072  | 12:12  | 11:48   | 24     | 11:12   | 60     |
| 11076  | 18:48  | 16:48   | 120    | 17:36\(^e\) | 72     |
| 11088  | 04:48\(^a\) | 04:36   | 12     | 02:48\(^a\) | 120    |
| 11114  | 19:12\(^a\) | 18:12   | 60     | 17:24   | 108    |
| 11116  | 11:12  | ...     | ...    | 10:36\(^e\) | 96     |
| 11122  | 11:00  | 03:48   | 432\(^f\) | ...    | ...    |
| 11130  | 06:48  | 23:48   | 420\(^f\) | 05:24   | 84     |
| 11136  | 18:00  | ...     | ...    | ...     | ...    |
| 11142  | 03:24  | 14:24   | 780\(^c\) | 01:48   | 96     |
| 11148  | 13:12  | 06:00   | 432\(^f\) | ...     | ...    |
| 11294  | 08:24  | 07:24   | 60     | ...     | ...    |
| 11400  | 16:12\(^e\) | ...     | ...    | 14:36\(^e\) | 96     |
| 11414  | 01:00\(^a\) | ...     | ...    | ...     | ...    |
| 11472  | 19:48  | ...     | ...    | ...     | ...    |
| 11523  | 13:36  | 05:24   | 492\(^f\) | 11:12   | 120    |

**Notes.**

\(^a\) Thresholding method applied to the [0.9, 1.2] km s\(^{-1}\) window using 12 minute cadence data (see Section 3.2.2).

\(^b\) Infection point method using 12 minute cadence data (see Section 3.3).

\(^c\) Taken from [300, 400 G).

\(^d\) Taken from [100, 200 G).

\(^e\) Taken from [600, 700 G).

\(^f\) Not deemed confident HDF.

\(^g\) Tentative signature.

The magnetic field window was normalized against \( \sigma \), and dotted lines denote that this was not possible. The sets have been color coded from dark blue ([100, 200 G]), through green ([400, 500 G]), to red ([700, 800 G]), with the specific colors noted in the legend. The colored vertical lines denote the time at which a set either exceeded the threshold, or stopped being zero. If the set did neither, no vertical line is drawn for that set. Note that a vertical line may not seem to appear if another is drawn over it. In the middle panel of Figure 6, we show the time profiles of AR 11066. Here, the time of emergence is visually clear, and starts around 20:48 UT in the [200, 300] G (9 \[9\] see mid blue vertical line with arrow “1” above) set, with the [300, 400 G] and [400, 500 G] sets [9] see light green vertical line with arrow “2” above) both finding a time of 21:48 UT. We can also see that a few profiles (>600 G) did not exceed the threshold over the course of the analyzed time period, though the [500, 600 G] did exceed 1σ briefly before returning. In the middle panel of Figure 6, we present the magnetic profiles of AR 11114. Here, we see emergence begins at 22:12 UT in the [200, 300] G set [9] (see arrow “3”), with the majority of remaining profiles crossing the threshold over the following hours. This event is of interest here as, when the image sequences of the event are studied, it becomes evident that the [100, 200] G profile [9] (see arrow “4”) in this case is more appropriate. This is due to the emergence having evolved significantly by 22:12 UT [9] (arrow “3”) when the higher magnetic field strength window begins to display emergence signatures. Finally, in the bottom panel of Figure 6, we plot the time profiles of AR 11130. Here, we immediately see that there is little to no activity over 400 G preemergence, with the majority of activity located in the lowest three magnetic field strength windows. Nonetheless, we see general agreement between these...
three windows and we, therefore, consider the [200, 300) G window giving 06:48 UT [9:] as the time of emergence. Interestingly, we see greater difference between the higher and lower profiles than in the other ARs in this panel, though the reasons are not investigated here. Comparing the profiles of these three ARs we see the common trend of a strong, possibly nonlinear increase [8:] (as would be expected from flux emergence, though we do not test nonlinearity here), across multiple sets of magnetic field strengths over the course of several hours, supporting our assertion that differences of several hours in $T_{FE}$ are acceptable.

3.2.2. The [0.9, 1.2) km s$^{-1}$ Doppler Velocity Window

In Figure 7, we plot the equivalent to Figure 6 for these three ARs except for the Doppler velocity bins. Likewise, these bins have been color coded from dark blue ($\pm[1.2, 1.5]$ km s$^{-1}$), through green ($\pm[0.5, 0.8]$ km s$^{-1}$), to red ($\pm[0.0, 0.3]$ km s$^{-1}$). Furthermore, each set has been given a unique combination of line style and marker to help distinguish them. We begin our analysis by studying the $\pm[0.9, 1.2]$ km s$^{-1}$ window in order to complement Section 3.1. The solid black vertical line represents $T_{FE}$ and the dotted-dashed vertical line represents the HDF signature, $T_{DET}$, from the $\pm[0.9, 1.2]$ km s$^{-1}$ window. For AR 11066 (top panel),
the clear rise in the Doppler windows sampling higher velocities is immediately clear at 19:36 UT [9] (see black dotted–dashed line) representing the HDF. For AR 11114 (middle panel), the Doppler velocity windows are much more complex, with large increases in the number of pixels detected in numerous windows rising from around 08:00 UT [9] (see below arrow “1”), before falling below the 1σ value again at around 18:00 UT [9] (see between black dashed and dotted–dashed vertical lines). This is followed by a subsequent rise in the [0.9, 1.2] km s\(^{-1}\) velocity window at 18:12 UT [9] (see black dotted–dashed vertical line). This time is, therefore, recorded as the HDF. For AR 11130 (bottom panel), we visually represent one of the potential problems with the thresholding method. This is that increases well before the time of emergence (more than 6 hr) can be picked up and returned by an automated code. These can be both transient such as for the higher Doppler velocity windows, which peak around 00:00 UT [9] (see Figure 7. Same as for Figure 6 but for the corresponding smoothed Dopplergram profiles. Each panel plots 13 different Doppler velocity windows, each covering six velocity bins of width 0.1 km s\(^{-1}\) over the range \([-1.5, 1.5]\) km s\(^{-1}\). Each Doppler velocity window is individually coded by color, line style, and marker as labeled in the legend. The solid black vertical line denotes the time of magnetic emergence, the vertical dotted–dashed black line indicates the time of preemergence identified using the thresholding method, and the dashed black vertical line locates the determined time of preemergence estimated using the inflection point method. Arrows are ordered by place of reference within text. The complete figure set (16 images) is available in the online journal.

(The complete figure set (16 images) is available.)

10
black dotted–dashed vertical line), or more sustained such as for the [0.0, 0.3] km s$^{-1}$ Doppler velocity window [9]; see red dotted profile [topmost] with asterisks, below arrow “2”). Either way, it is difficult to return a clear $T_{DEP}$ for this AR.

In the third column of Table 2, we present the results of the thresholding method for the [0.9, 1.2] km s$^{-1}$ Doppler velocity window. We find confident HDF signatures for 6 of the 16 ARs studied here. The thresholding technique returns results for five further ARs; however, the HDF signature is found to precede the flux emergence by more than 7 hr in each case (indicated by an “F” in the fourth column of Table 2). We, therefore, ignore those ARs in this analysis. From the six ARs that we return confident HDF signatures from, we find lead times, $\Delta T$, of between 12 and 120 minutes (see the fourth column of Table 2), with an average of 58 minutes. Both the percentage (37.5%) of ARs displaying HDF signatures and the lead times returned by the thresholding method (58 minutes) match well with the results of Toriumi et al. (2012, 2014, 2016) and 61 minutes, respectively, giving us confidence that the lower-cadence data studied here does not significantly influence the results. Equivalent plots to Figures 6 and 7 are included for each analyzed AR in the supplementary material for completeness.

### 3.2.3. Other Doppler Velocity Windows

We now attempt to calculate $T_{DEP}$ using the thresholding method for a larger number of Doppler velocity windows. The results from this analysis are presented in Table 3. Each result is split into one of four categories depending on whether/how the profile increases over 1$\sigma$. We do not differentiate between increases over $1\sigma$ and decreases below $1\sigma$ meaning these categories can be either positive (increases over $1\sigma$) or negative (decreases below $1\sigma$). Category 0 indicates that the absolute value of the Doppler velocity window never increased above $1\sigma$. Category 1 indicates that the absolute value of the profile increased over $1\sigma$ at some point during the analyzed time, but that it had decreased back below the $1\sigma$ value before $T_{FE}$. Category 2 indicates that the window increased over $1\sigma$ within one frame of $T_{FE}$. Finally, category 3 indicates that the Doppler velocity window increased over $1\sigma$ and stayed above that value until after $T_{FE}$.

It is immediately evident that different Doppler velocity windows return different results, with some appearing more effective in detecting HDFs than others. Notably, the $\pm[1.2, 1.5]$ km s$^{-1}$ and $\pm[0.1, 0.4]$ km s$^{-1}$ Doppler velocity windows return very few potential results. The superscript “a” in Table 3 indicates whether a profile crosses the $1\sigma$ threshold within two hours of $T_{FE}$ for any given AR. We consider two hours to be an appropriate upper limit for our discussion as it is double the average HDF lead time reported in the previous subsection and in Toriumi et al. (2014). Given that the method of Toriumi et al. (2014) is limited to increases over a positive threshold value, we will initially focus only on positive potential HDF signatures within two hours of $T_{FE}$. In this sense, the $\pm[0.9, 1.2]$ km s$^{-1}$ and $\pm[0.8, 1.1]$ km s$^{-1}$ Doppler velocity windows return the most results, with both finding six potential HDF signatures. Several other Doppler velocity windows return four or five potential HDF signatures within two hours of $T_{FE}$, with only the $\pm[0.1, 0.4]$ km s$^{-1}$ window returning no results. Considering the values with negative significance (i.e., the profiles that drop below $-1\sigma$ within two hours of $T_{FE}$), we find a different picture with the $\pm[0.0, 0.3]$ km s$^{-1}$ Doppler velocity window identifying the most potential HDF signatures with four. No negative potential HDF signatures (at any time before $T_{FE}$) were identified in the Doppler velocity windows included in the first four columns of Table 3, suggesting analysis of negative values may only provide results at lower Doppler velocities.

Our analysis has shown that if we consider a wider array of Doppler velocity windows, we are able to increase the number of potential HDF signatures within two hours of $T_{FE}$. From the ARs studied, we are able to move from 6 (out of 16) ARs from which HDFs can be identified from the $\pm[0.9, 1.2]$ km s$^{-1}$ Doppler velocity window to 13 ARs from which HDFs can be detected when considering all Doppler velocity windows. One notable example of an AR from which a potential signal can now be detected is AR 11122, which displays positive significance only in the $\pm[0.0, 0.3]$ km s$^{-1}$ Doppler velocity window. This AR emerged close to disk center (at a longitude of around $-25^\circ$) potentially explaining why the clear HDF signatures detected in higher Doppler velocity windows at larger $\mu$ angles may not be visible. Instead the rising flux may inhibit local convection leading to an increase in pixels that display very small Doppler velocities, as may be seen in Cheung et al. (2007).

### 3.3. Collective Group Behavior

Following on from our statistical analysis, we now try to develop a technique for detecting preemergence that does not rely on one single Doppler velocity window, but instead makes use of the collective group behavior across multiple windows. It is here that our proposed method of determining the time of preemergence diverges significantly from that of Toriumi et al. (2012); compare Figure 3 M5-8 to D5-8. In this subsection, we instead look for transient disruptions across a group of Doppler velocity bin profiles, marked by collective changes in relative amplitudes. We specifically look for times when multiple Doppler velocity windows peak (or trough) identified by “inflection points.” This allows us to search profiles that have higher thresholds throughout their evolution (such as AR 11130) and profiles that do not match the implicit “below-until-event” assumption made in Toriumi et al. (2012). This presents the problem of what feature within the profile to choose, when each is inherently noisy and oscillatory. Due to this, we choose to mark our proposed time of preemergence as the final inflection point before magnetic emergence, as this is unambiguous and replicable. Currently, the selection of which group of inflection points to consider is done manually in conjunction with careful analysis of image sequences, with automated methods in development.

We will now discuss the inflection point method by using the three ARs plotted in Figure 7 as examples. The times of the potential preemergence signatures for this method, $T_{DEP}$, are denoted by the vertical dashed black lines. For AR 11066 (top panel), at around 16:24 [9]; see arrow “3”) the Doppler profiles clearly start to show a sudden and rapid change to the amplitude of the signal, reaching a peak (or inflection point) at 18:12 UT [9] (see black dashed vertical line). This is both 156 minutes prior to the magnetic emergence at 20:48 UT [9] (black solid vertical line) and 70 minutes prior to the HDF time reported for this AR by Toriumi et al. (2014). We see the inflection point in two forms. The first is the decrease and convergence of a number of higher velocity sets ($\pm[0.9]$ km s$^{-1}$), and the second is the increase and convergence of a number of lower velocity sets, ($\leq0.8$ km s$^{-1}$) [9] (see arrow “3”). For AR 11114 (middle panel), numerous profiles increase
Table 3

Time of HDF Detection, $T_{\text{DET}}$, for Each Studied Velocity Window for the 16 ARs Analyzed Here

| AR # | [1.2–1.5] | [1.1–1.4] | [1.0–1.3] | (0.9–1.2) | (0.8–1.1) | (0.7–1.0) | (0.6–0.9) | (0.5–0.8) | (0.4–0.7) | (0.3–0.6) | (0.2–0.5) | (0.1–0.4) | (0.0–0.3) |
|------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 11066 | 0–2 | 1: 19:48* | 3: 19:36* | 3: 19:24* | 3: 19:24* | 1: 19:36* | 1: 19:48* | 1: 20:00* | –1: 14:00 | –1: 17:24 | 0–2 | 1: 17:00 |
| 11072 | 0–3 | 2: 12:12* | 2: 12:00* | 3: 11:48* | 1: 08:00 | 1: 08:24 | 1: 08:36 | –1: 11:24* | –3: 09:36 | –2: 12:12* | 0–2 | 1: 06:48 |
| 11076 | 0–3 | 1: 16:36 | 1: 16:48* | –2: 18:36* | –3: 18:00* | –3: 17:24* | –3: 16:36* | –3: 12:12 | –3: 12:00 | –3: 16:24 | 0–3 | 1: 16:24 |
| 11088 | 1: 23:36 | 2: 04:36* | 2: 04:36* | 2: 04:36* | 1: 18:24 | 1: 18:24 | 1: 18:12 | 1: 17:48 | 1: 18:12 | 1: 21:36 | –1: 19:24 | –1: 02:48* |
| 11114 | 0–1 | 1: 15:00 | 1: 11:00 | 1: 18:12* | 3: 10:12 | 3: 09:24 | 3: 09:12 | 3: 09:00 | 3: 09:00 | 3: 10:24 | 1: 17:00 | 0–2 | 1: 06:48 |
| 11116 | 0–3 | 0–0 | 0–1 | 1: 10:24* | 0–1 | 1: 09:24 | –1: 11:24* | –1: 11:24* | –1: 09:00 | –1: 09:12 | 1: 12:00 | 0–2 | 1: 06:48 |
| 11122 | 0–3 | 0–1 | 1: 03:48 | 1: 03:48 | 0–1 | 3: 09:48* | –3: 09:48* | –3: 09:48* | –3: 09:00* | –3: 08:12 | –3: 07:36 | –3: 09:00* | 0–2 | 1: 06:48 |
| 11130 | 0–3 | 2: 06:36* | 1: 23:48 | –1: 04:12 | –1: 02:36 | –2: 06:36* | –3: 02:24 | –3: 01:00 | –3: 01:48 | –3: 01:48 | 0–3 | 2: 06:48 |
| 11136 | 0–3 | 0–1 | 1: 15:00 | 0–0 | 0–1 | 1: 17:24* | 3: 17:12* | 3: 17:00* | –1: 14:12 | –1: 14:12 | 1: 14:24 | 0–0 | 1: 06:48 |
| 11142 | 0–3 | 1: 00:24 | 1: 14:36 | 1: 14:24 | 1: 00:36 | 1: 00:36 | 2: 03:12* | 3: 02:48* | 3: 02:36* | 3: 02:48* | 3: 03:00* | –1: 02:48 | –3: 02:48* |
| 11148 | 0–3 | 0–0 | 1: 10:36 | 1: 00:60 | 1: 06:12 | 0–0 | 0–0 | 0–0 | 0–0 | 0–0 | 0–0 | 0–0 | 1: 10:36 |
| 11294 | 0–3 | 3: 07:24* | 3: 07:24* | 3: 07:24* | 3: 07:24* | 3: 07:24* | 3: 07:24* | 3: 08:00* | 0–0 | 0–0 | 0–0 | 0–2 | 0: 28:48* |
| 11400 | 0–3 | 0–0 | 0–0 | 0–2 | 1: 16:00* | –3: 15:48* | –3: 13:12 | –3: 10:00 | –3: 10:00 | –3: 10:00 | –3: 10:12 | 0–3 | 1: 13:00 |
| 11414 | 0–3 | 0–0 | 0–0 | 0–0 | 0–0 | 0–1 | 1: 16:12 | –2: 00:48* | –3: 00:24* | –3: 00:24* | 0–0 | 0–0 | 1: 16:12 |
| 11472 | 0–3 | 1: 19:36* | 0–0 | 0–0 | 1: 17:48* | 1: 19:00* | 1: 19:24* | 1: 18:36 | –1: 09:24 | –1: 14:36 | 0–0 | 0–0 | 1: 19:36* |
| 11523 | 2: 13:12* | 3: 09:12 | 3: 08:24 | 3: 05:24 | 3: 05:12 | 3: 05:36 | 3: 05:48 | 3: 06:00 | 3: 07:00 | 3: 09:12 | 3: 12:42* | –3: 10:12 | –3: 07:24 |

Notes. The column labels indicate the specific velocity window in km s$^{-1}$ (±). The results are split into four classifications, namely: “0” where the absolute value of the profile never increases above 1σ; “1” where the absolute value of the profile goes above 1σ but returns below this level before $T_{\text{FE}}$; “2” where $T_{\text{FE}}$ and $T_{\text{DET}}$ are either separated by one time step or are equal; and “3” where the absolute value of the profile increases above 1σ and remains above that value until after $T_{\text{FE}}$. Positive and negative values of 1, 2, and 3 indicate whether the velocity window goes above 1σ or below −1σ, respectively.

*a* Within two hours of $T_{\text{FE}}$.
above (decrease below) the $1\sigma$ ($-1\sigma$) value at around 11:00 UT [9\(^\circ\)] (see arrow “4”), eight hours before $T_{FE}$. A number of the profiles remain above $|\sigma|$ for the entire time; however, a clear example of collective group behavior is apparent starting at around 16:30 UT [9\(^\circ\)] (see arrow “5”). This results in an inflection point at 17:24 UT, 108 minutes prior to $T_{FE}$ [9\(^\circ\)] (see dashed black vertical line). For AR 11130 (bottom panel), an increase above the $1\sigma$ threshold occurs at around 01:36 UT [9\(^\circ\)] (see arrow “2”) for the $[0.0, 0.3]$ km s\(^{-1}\) Doppler velocity window, which is too early to be associated with an HDF. The inflection point becomes very useful for estimating preemergence in this case, identifying a preemergence time of 05:24 UT [9\(^\circ\)] (see dashed black vertical line), which is 84 minutes before the magnetic emergence at 06:48 UT [9\(^\circ\)] (see solid black vertical line). Here the change in amplitude is seen relative to already diverging profiles, making it more difficult to judge the precise time of change. We suggest that this inflection point could be due to the subphotospheric flux disrupting the turbulent motions at the surface, as may also be seen in AR 11114.

In the fifth and sixth columns of Table 2, we present the time of the inflection point and the difference between this time and $T_{FE}$, denoted by $\Delta T_I$. We find group collective behavior resulting in an inflection point for 10 (out of 16) ARs, finding an average $\Delta T_I$ of 100.8 minutes, nearly double the 58 minutes found using the thresholding method. Of these ARs, four display only potential signatures of inflection points, denoted by the superscript “g” in Table 2; however, we consider these to be tentative here and include them in our sample. While there is a recoverable inflection point for most of the studied ARs, there are other features within the Doppler velocity profiles of the other ARs that make it harder to identify. For the six ARs from which no inflection point is recoverable, this is typically because there is no clear change in the amplitude across the profiles beyond the underlying oscillation. For AR 11294, it is simply due to the magnetic emergence occurring very near to the start of the data set, making judgments about changes in amplitude impossible. It is clear that further work must be conducted to optimize this method, but we do find it encouraging for the future.

As a final piece of analysis, we study the relationship between longitude and the Doppler velocity windows, which are positive during the inflection point. Here, we define an AR as a “center” if profiles below the $[\pm0.3, 0.6]$ km s\(^{-1}\) Doppler velocity window show a positive number of counts over the majority of the profile before the time of magnetic emergence. A profile is deemed to be a “wing” profile when we see the opposite. To demonstrate this longitudinal dependence upon our results, in Figure 8 we show whether each AR shows an increase in the center or the wings plotted against longitude. As can be seen in Figure 8 there is clear correlation between the longitude and the resultant trend in the data, with center trends being found for events with $|\Delta\lambda| \lesssim [30]^{\circ}$, and wing trends found for $|\Delta\lambda| \gtrsim [20]^{\circ}$. We believe that a small transition region between center and wing trends will become apparent with a larger sample size due to variations between data sets. This analysis once again highlights the importance of studying multiple Doppler velocity windows.

### 4. Discussion

In this article, we have used two methods to analyze 16 ARs of the 21 denoted as “P0” in Schunker et al. (2016; detailed in both Figure 1 and Table 1) in order to determine potential preemergence signatures using two methods. The first method was the thresholding method (as previously used by Toriumi et al. 2012, 2014), and the second method was through the identification of inflection points. Both methods were applied to Doppler velocity and magnetic field windows normalized against their own standard deviations, to allow direct comparison across multiple Doppler velocity windows and multiple ARs. The thresholding method was applied in order to detect signatures of...
Differences as the HDF signature from the ± consistency in our results.

We repeated our analysis on the same three ARs using the ±[0.9, 1.2] km s\(^{-1}\) Doppler velocity window and the lower-cadence 12 minute data to investigate whether HDFs could still be detected. Our results (plotted in Figure 5) match with the results found in our higher-cadence data and in Toriumi et al. (2014) with only minor differences between the ARs being returned. One of the larger differences was the detection of a HDF for AR 11076 when no HDF was detected in the higher-cadence data. However, the increase above the 1\(\sigma\) level returned as the HDF signature from the ±[0.9, 1.2] km s\(^{-1}\) Doppler velocity window was transient, lasting only three frames and suggesting that this may not have been a true HDF. Given the similarities in results, we decided to use the lower-cadence 12 minute data for a larger statistical analysis here. This analysis, using the ±[0.9, 1.2] km s\(^{-1}\) Doppler velocity window, returned HDFs for 6 (37.5\%) of ARs with an average [11\:] time difference (using the threshold time, \(\Delta T_D\), of 58 minutes. These results match well with Toriumi et al. (2014) who found clear HDFs for 25\% of ARs with a \(\Delta T_D\) of 61 minutes. The results from this analysis can be found in Table 2.

Finally regarding the thresholding method, we conducted an analysis of a wider range of Doppler velocity windows spanning from ±[1.2, 1.5] km s\(^{-1}\) to ±[0.0, 0.3] km s\(^{-1}\) to investigate whether HDFs could be detected more widely. This analysis allowed us to increase the number of potential HDF detections (within two hours of \(T_D\)) to 13 (out of the 16 ARs investigated). We did not calculate a \(\Delta T_D\) value from this analysis as our two-hour limit, selected to retrieve general results, would have biased our calculations. Limiting our analysis to positive results initially (i.e., where the Doppler velocity window increased over the \(\sigma\) level), it was found that the ±[0.9, 1.2] km s\(^{-1}\) and ±[0.8, 1.1] km s\(^{-1}\) were the most effective Doppler velocity windows for detecting HDFs with both finding six potential HDFs. Of particular interest was AR 11122, which only displayed evidence of a potential positive HDF signature in the [0.0, 0.3] km s\(^{-1}\) Doppler velocity window, highlighting the benefit of studying a wider range of Doppler velocity windows. Considering negative results across all ARs (i.e., where the Doppler velocity window dropped below the \(-1\sigma\) threshold), the ±[0.0, 0.3] km s\(^{-1}\) window appeared to display the most (four) potential signatures of the HDF (within two hours of \(T_D\)). The inclusion of negative results could be important when the positive HDF signal is blurred across multiple Doppler velocity windows (e.g., closer to the disk center), potentially meaning that it does not increase above the \(1\sigma\) value in any.

Given that potential HDF signatures were detected in multiple Doppler velocity windows for some ARs (see for example AR 11066 in Table 3), we then attempted to define a method that could make use of the collective behavior of multiple Doppler velocity windows when searching for preemergence signatures. Our initial method is based around the identification of concurrent inflection points identified across multiple Doppler velocity windows. Such inflection points were found in 10 ARs across a wide range of longitudes (−71°5.5°−8°) and latitudes (−28°3–22°3). The mean time between the inflection points and \(T_D\), defined as \(\Delta T_D\), was 100.8 minutes, nearly double the lead time from the thresholding method. We also studied whether there was a dependence between the longitude at which an AR emerged and the Doppler velocity windows that were positive during \(T_D\). In Figure 8, we plot the results of this analysis, finding ARs that emerged at less than 30° displayed center trends (Doppler velocity windows <±[0.3, 0.6] km s\(^{-1}\)) while ARs that emerged at higher angles displayed wing trends.

While we have been able to use inflection points for determination here, we do note several downfalls in this method that other authors should be aware of. In a live system, for example, we would have to wait for sufficient data to be collected in order to form the inflection point meaning our \(\Delta T_D\) values are upper limits. Although this is an advantage over the original threshold method, the underlying trends within the data make it a necessary compromise. Other techniques such as cross correlation or Bayesian online change-point detection (Prescott Adams & MacKay 2007) may be able to overcome this, but at the expense of greater computational cost. Future work would require these different techniques to be benchmarked. As this method is currently applied manually, we provide a brief overview of how we interpret different signals for completeness.

I. Strong signals. The first group of ARs we discuss are those that contain strong inflection points. This includes ARs 11066 and 11114, which have been previously discussed in Section 3.3. As previously mentioned for AR 11066 we see a strong increase in multiple velocity bins peaking at an inflection point at 18:12 UT. For AR 11114, we consider the co-temporal change across almost all sets peaking at 17:24 UT to be a strong inflection point.

II. Diverging Signals. The second group is that of divergent sets within a profile, which are not quite as clear as those of group I. Here we include ARs 11072, 11076, 11400, and 11523. For AR 11072 we see the profiles beginning to diverge at the time selected for \(T_D\). Before this, we see a double-peaked burst of activity starting just after 05:24 UT, and ending around 09:24 UT. For AR 11076, we consider this a divergent signal due to the strong activity seen around 16:24 UT across the majority of sets; however, this is tentative. For AR 11400, when viewing both the magnetogram profiles and image sequences of the event, we notice the slow accumulation of flux that creates divergence in the Doppler velocity windows at around 10:00 UT making the use of the threshold method less than ideal in this case. We also notice that just before the flux emergence there exists a small and diffuse bipole that is rapidly separating at the site from which the main flux emerges. As there is a clear large-scale flux emergence later, we take the closest time at which a magnetic profile exceeds the 1\(\sigma\) threshold to the larger scale emergence as \(T_D\). As the relationship between the activity of the smaller preexisting flux and the larger newly emerging flux is unclear, we tentative take \(T_D\) as the inflection point, which occurs just prior to the latter at 14:36 UT. Finally, for AR 11523 we can see a great deal of activity within the Doppler profile after an initial “quiet” period. While we must take care interpreting these profiles due to the AR \(\mu\) angle (see Table 1), we do not see a comparative level of
activity within the magnetic profiles until around $T_{\text{DEI}}$ at 11:12 UT making it hard to determine the source of these velocities.

III. Converging Signals. The next group is that of converging signals. Here we include ARs 11088, 11130, and 11142. For AR 11088, we see a tentative convergence across a majority of sets starting around the same time of 02:48 UT, identified here as $T_{\text{DEI}}$. This convergence, rather than the apparent slower decay seen in the peaks just before flux emergence, suggests the disruption of the process that caused the activity. For AR 11130, we see a prolonged divergence converging at 05:24 UT just before emergence. Finally, for AR 11142 we once more see the convergence in the Doppler profile at around 01:48 UT before emergence. However, this convergence is significantly earlier than our other examples, and is further confounded by some minor flux emergence that occurs just before emergence.

IV. Tentative signals. We consider AR 11116 a tentative signal rather than a null due to a number of minor differences between the behavior sets around $T_{\text{DEI}}$. First, we point to the tight convergence of the sets at the $T_{\text{DEI}}$-excluding the $\pm [0.8, 1.1] \text{ km s}^{-1}$ set, which increases from inactivity. Careful examination of the other bursts shows a lack of such convergence afterward, in addition to the less co-temporal activity across sets. We include this result here in case future analysis can provide greater insight into this event.

V. Null signals. We have six null results, for which no inflection point could be determined in the Doppler profiles. Null results are denoted by dashes in the $\Delta T_{\text{f}}$ column of Table 1.

[12:] As mentioned previously we do not distinguish between the upward velocity of the emerging flux, $V_{\text{c}}$, and the horizontal velocity of the divergent flow, $V_{\text{h}}$, to use the nomenclature of Toriumi et al. (2012; see their Figure 8). This is based on an implicit assumption that the emerging flux tube would have either recoverable components of both $V_{\text{c}}$ and $V_{\text{h}}$ at all angles or that the disruption would cause a sufficient transient change to the underlying distribution as to be measurable in the appropriate LOS component. A key question therefore is whether $V_{\text{c}}$ is itself recoverable. We believe the answer to be yes, due to the number of ARs with longitudes $< [20^\circ]$ we have been able to recover a signal with both methods used here (e.g., AR 11114 (Figure 7(b))). This is an improvement over the work of Toriumi et al. (2012, 2014), who were not able to find signatures within $[20^\circ]$, though this may be due to their chosen sample. This is an important step toward being able to predict AR emergence with a single method across the whole disk, and not just away from the disk center. Furthermore, we have been able to find a clear signature of emergence at an extreme angle (AR 11142 at 715), further widening the applicable window using these techniques.

In addition, we must also be aware that we may not only be recovering the preemergence signals of the flux that becomes the AR. This is something that is not usually considered by other authors investigating this subject who only consider the emergence of a single flux rope into the quiet Sun, [13:] as opposed to a flux rope emerging into another emerging flux rope or into a preexisting active region. Fortunately, by comparing image sequences of the magnetograms to their Doppler profiles we are able to distinguish which are the relevant signals for us to consider here. In future work, a larger sample of ARs may allow us to automatically distinguish which signals correspond to that of the large-scale preemergence that we have focused in this work.

5. Conclusions

We have analyzed 16 ARs to determine the time of preemergence signatures using Doppler velocity data from the SDO/HMI instrument. The studied ARs cover a large range of longitudes $-71.5^\circ \leq \theta \leq 32.3^\circ$. We initially used a threshold based method to find the time of magnetic emergence, $T_{\text{DEI}}$, in the [200, 300] G window (where possible). To analyze the LOS velocities, we used two different methods, namely, the thresholding method used by Toriumi et al. (2014) and the inflection point method developed here. In addition, this work has generalized the method of Toriumi et al. (2012) to other cadences and Doppler velocity windows. A comparison of the thresholding method applied to 12 minute cadence data to the results of Toriumi et al. (2014) found comparable results with 37.5% of ARs displaying evidence of HDFs, with a mean $\Delta T_{\text{f}}$ of 58 minutes. For the inflection point method, we found at least tentative signals for 62.5% of ARs, with a mean $\Delta T_{\text{f}}$ of 100.8 minutes. This is higher than the mean presented by Toriumi et al. (2014). When comparing the same events we find that our method finds times earlier than the threshold for the Doppler profiles. However, future statistical work is necessary to further constrain the cases in which by-eye analysis is currently required.

We acknowledge support from the Science and Technology Facilities Council (STFC) for the support received through grant Nos. ST/P000304/1 & ST/T00021X/1. C.J.N. acknowledges support from the European Space Agency (ESA) as an ESA Research Fellow. SDO/HMI data are provided courtesy of NASA/SDO and the HMI science team.

ORCID iDs

T. Rees-Crockford @ https://orcid.org/0000-0003-4243-1776
C. J. Nelson @ https://orcid.org/0000-0003-1480-8356
M. Mathioudakis @ https://orcid.org/0000-0002-7725-6296

References

Acheson, D. J. 1979, SoPh, 62, 23
Archontis, V., Moreno-Insertis, F., Galsgaard, K., Hood, A., & O'Shea, E. 2004, A&A, 426, 1047
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
BIPM 2019, Le Système international d’unités/The International System of Units (“The SI Brochure”) (9th ed.; Paris: Bureau international des poids et mesures), http://www.bipm.org/en/si/si_brochure/
Birch, A. C., Braun, D. C., Leka, K. D., Barnes, G., & Javornik, B. 2013, ApJ, 762, 131
Bogart, R. S. 2007, AN, 328, 352
Braun, D. C. 1995, in ASP Conf. Ser. 76, GONG 1994. Helio- and Astro-Seismology from the Earth and Space, ed. R. K. Ulrich, E. J. Rhodes, Jr., & W. Dappen (San Francisco, CA: ASP), 250
Cheung, M. C. M., Rempel, M., Title, A. M., & Schüssler, M. 2010, ApJ, 720, 233
Cheung, M. C. M., Schüssler, M., & Moreno-Insertis, F. 2007, A&A, 467, 703
Grigoriev, V. M., Ermakova, L. V., & Khilystova, A. I. 2007, AstL, 33, 766
Hale, G. E., Ellerman, F., Nicholson, S. B., & Joy, A. H. 1919, ApJ, 49, 153
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Ilonidis, S., Zhao, J., & Kosovichev, A. 2011, Sci, 333, 993
Kawaguchi, I., & Kitai, R. 1976, SoPh, 46, 125
Otsuji, K., Kitai, R., Ichimoto, K., & Shibata, K. 2011, PASJ, 63, 1047
Parker, E. N. 1955, ApJ, 121, 491
Prescott Adams, R., & MacKay, D. J. C. 2007, arXiv:0710.3742
Scherrer, P. H., Schou, J., Bush, R. L., et al. 2012, SoPh, 275, 207
Schuck, P. W., Antioco, S. K., Leka, K. D., & Barnes, G. 2016, ApJ, 823, 101
Schunker, H., Braun, D. C., Birch, A. C., Burston, R. B., & Gizon, L. 2016, A&A, 595, A107
Syntelis, P., Archontis, V., & Hood, A. 2019, ApJ, 874, 15
The SunPy Community, Barnes, W. T., Bobra, M. G., et al. 2020, ApJ, 890, 68
Toriumi, S., Hayashi, K., & Yokoyama, T. 2012, ApJ, 751, 154
Toriumi, S., Hayashi, K., & Yokoyama, T. 2014, ApJ, 794, 19
Toriumi, S., & Yokoyama, T. 2010, ApJ, 714, 505
Toriumi, S., & Yokoyama, T. 2013, A&A, 553, A55
Zharkov, S., & Thompson, M. J. 2008, SoPh, 251, 369