SEARCH: SEgmentation of polAR Coronal Holes

Ajay Kumar Tiwari[1], Benoit Tremblay[2, 3, 4], Andong Hu[1], Linnea M. Wolniewicz[3], Michael Kirk[5, 6], Silvina Guidoni[7], Brent Smith[8], Matthew Penn[9], and Tanmoy Samanta[10]

[1]Centrum Wiskunde & Informatica, Amsterdam, NL
[2]Laboratory for Atmospheric & Space Physics, Boulder, CO, USA
[3]CU Boulder, Boulder, CO, USA
[4]National Solar Observatory, Boulder, CO, USA
[5]ASTRA, Louisville, CO, USA
[6]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[7]American University, Washington, DC, USA
[8]Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA
[9]NVIDIA, Santa Clara, CA, USA
[10]George Mason University, Fairfax, VA, USA

Abstract

The identification of solar coronal holes (CHs) observed in Extreme Ultraviolet (EUV) intensity images of the Sun is key in improving our understanding of their association with solar magnetic fields and heliophysics. In particular, CHs at the poles of the Sun are a notorious source of fast solar wind and thus warrant further study, most notably in the context of space-weather forecasting. This has consequently led to the development of various segmentation methods for their identification, including supervised machine learning. We introduce the SEARCH project to combine EUV data from the three vantage points offered by the SoHO, STEREO-A and STEREO-B satellites during the 2010-2014 epoch to produce synchronic maps and apply unsupervised learning methods including clustering and convolutional neural networks for the segmentation of CHs. SEARCH segmentation maps provide a venue to explore the relationship between CH pole areas, geomagnetic activity, and the magnetic activity cycle (dynamo process) of the Sun and Sun-like stars.

1 Introduction

The polar regions of the Sun remain elusive to the current constraints of solar imagery. No spacecraft has yet left the ecliptic plane by more than a few degrees from a heliographic view. This lack of direct measurement of solar poles has necessitated indirect methods for understanding the polar environment. Magnetic fields at the poles are notably used to forecast upcoming solar magnetic activity cycles (Schatten et al., 1978; Schatten, 2005; Svalgaard et al., 2005; Pesnell and Schatten, 2018). Additionally, the influence of polar CHs extends well beyond the immediate solar atmosphere. For instance, high-speed solar wind streams have long been known to originate from CHs (Nolte et al., 1976; Harvey and Sheeley, 1978; Wang and Sheeley, 1990).

Identification of CHs and Active Regions (ARs), the most important visible features on the solar disk, is crucial for understanding and making predictions for space weather. CHs are thought to be dominated by the open magnetic field lines that thread through the solar atmosphere plasma and extend into the heliosphere. CHs are defined by their low emissivity in extreme ultraviolet (EUV) and soft X-ray wavelengths and their low plasma density (Bohlin, 1977), which result in dark regions at those wavelengths. ARs are typically associated with closed, confined magnetic field lines and

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

have been studied extensively due to their role in the eruptive processes such as flares and Coronal Mass Ejections (CMEs). The correlation between the solar magnetic cycle and the apparent size and shape of polar CHs is a matter of ongoing investigation (*e.g.*, Waldmeier, 1981; Bravo and Stewart, 1994; Harvey and Recely, 2002; Kirk et al., 2009). These studies have utilized digitally-assembled global maps of the Sun, such as synoptic maps (*e.g.*, Hess Webber et al., 2014; Lowder et al., 2014; Karna et al., 2014).

Although intensity-based thresholding methods of identifying CHs are the most common (*e.g.*, the CHIMERA catalogue: Garton et al., 2018), other venues have been explored. Verbeeck et al. (2014) used a probability distribution function of EUV image pixel intensities and a clustering technique to segment CHs with the Spatial Possibilistic Clustering Algorithm (SPoCA). Recently, supervised machine learning (ML) was used to identify CH boundaries in full-disk images (Illarionov and Tlatov, 2018) and synoptic maps (Illarionov et al., 2020) of the Sun at 193 Å such as the one in Figure 1a. Illarionov and Tlatov (2018) trained a convolutional neural network (CNN) on segmentation maps that were generated by the Kislovodsk Mountain Astronomical Station using a semi-automatic procedure. The drawback of supervised ML is its reliance on catalogues of consistently identified CHs for training. At best, supervised ML will emulate the database it was trained on. With no existing ground truth for CH boundaries, there are considerable challenges in determining the accuracy of models and catalogues.

In this paper, we address the identification of CHs in an unsupervised learning setting, *i.e.* without the use of models or databases and the biases they may have. We introduce the open-source Segmentation of polAR Coronal Holes (SEARCH[1]) project which will include (a) a data reduction pipeline to fetch and preprocess EUV images from the SoHO and STEREO satellites and combine them into synchronic maps, and (b) unsupervised learning algorithms ranging from clustering to segmentation to identify polar CHs in synchronic maps.

2 Synchronic Maps of the Sun

Synchronic maps take advantage of the multiple viewing angles of complementary spacecrafts to fully map the Sun in the EUV wavelengths at a given time. Synchronic maps are advantageous over synoptic maps because they do not rely on the solar rotation to assemble a full-Sun image and account for the dynamic evolution of CHs over that 27-day period. The multi-viewing angles capture entirely the boundaries of polar CHs, provided the timely alignment of the satellites' orbits (*e.g.*, Figure 1b).

The advantage of using synchronic maps in the detection of the coronal holes is shown in Caplan et al. (2016). The aforementioned study provides single-wavelength synchronic maps between June 2010 to August 2014 located at the Predictive Science Inc. (PSI) database[2]. PSI synchronic maps combine data at 195 Å captured by the Exteme UltraViolet Imagers (EUVIs: Wuelser et al., 2004) on board the Solar TErrestrial RElations Observatory Ahead (STEREO-A) and Behind (STEREO-B) and data at 193 Å captured by the Atmospheric Imaging Assembly (AIA: Lemen et al., 2012, see Figure 1a) on board the Solar Dynamics Observatory (SDO).

The definition of CH boundaries is ambivalent because they appear differently in different wavelength. As part of the SEARCH collaboration, work is underway to utilize the available multi-wavelength, multi-instrument legacy EUV data to create a database of synchronic maps and explore CH boundaries in a holistic way. To this end, we combine data from the Extreme ultraviolet Imaging Telescope (EIT: Delaboudinière et al., 1995) on board the Solar and Heliospheric Observatory (SoHO) and the STEREO-A and -B EUVIs. For all three instruments, we process images in three wavelength-filter channels: 171 Å, 195 Å and 304 Å. Preprocessing steps of the data include wavelet enhancement of the images to reduce noise and increase their contrast (*e.g.*, see Stenborg et al., 2008) and the homogenization of data from multiple sources through the histogram-matching algorithm described in Hamada et al. (2020). The aforementioned steps will result in over 2000 three-channel full-Sun synchronic maps during 2010-2014 that will be used for the automated detection of CHs with SEARCH. However, for the purposes of this paper we test the validity of the unsupervised learning methods included in SEARCH on PSI single-wavelength synchronic maps.

[1] SEARCH Github page: `https://github.com/heliohackweek/polar_coronal_holes`. [2] Predictive Science Inc. synchronic maps: `http://www.predsci.com/chd/chd_maps.html`.

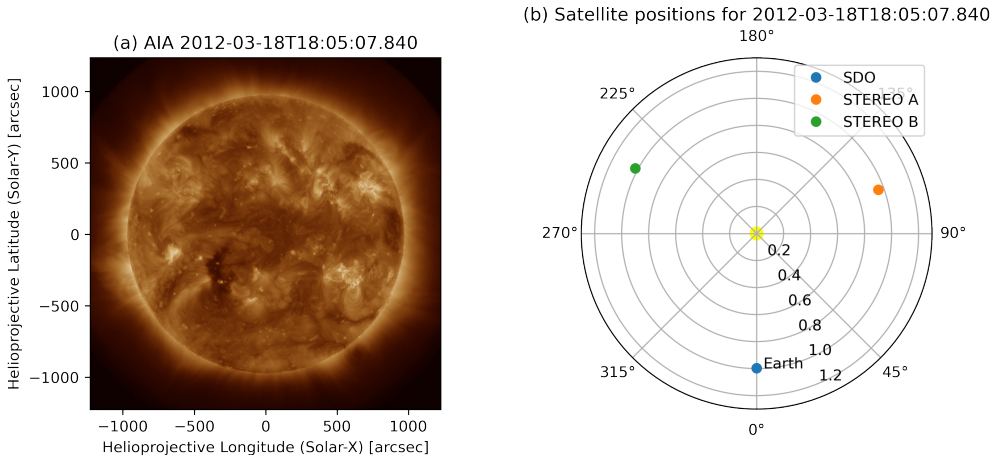

Figure 1: (a) Full-disk image of the Sun captured by SDO/AIA at 193 Å on 2012-03-18 at 18:05(UT). (b) Positions of the SDO and STEREO satellites at the time of observation in (a).

3 Unsupervised Learning for Segmentation

Unsupervised learning is performed on synchronic maps in one of two techniques: (1) Clustering with the K-means algorithm and (2) a recently developed W-net convolutional neural network (CNN). The K-means algorithm iteratively assigns pixel values in single-channel or three-channel EUV images to a given number of classes by minimizing distances between cluster points and maximizing the distance between clusters. The number of classes is optimized using the elbow method (Joshi and Nalwade, 2013). Each class is then compared and associated with the different structures observed in EUV images of the Sun such as ARs (appearing as bright structures) and CHs (appearing as dark structures). This approach is comparable to the SPoCA-suite, but for synchronic maps.

CNNs excel at identifying features at different spatial scales and segmenting images. For example, U-nets are fully CNNs that inherit their name from their U-shaped architecture, consisting of a contracting branch for context and an expanding branch for localization. U-nets were previously shown to be effective at identifying CH boundaries in a supervised learning framework (Illarionov and Tlatov, 2018; Illarionov et al., 2020). The recently developed W-net, an architecture that merges two U-nets to form an autoencoder, segments images in a fully unsupervised framework (Rajendrakumar Gare et al., 2020). More specifically, the first U-net is an encoder to produce a segmentation map from an input image, whereas the second U-net is a decoder to reconstruct the input image from the segmentation map. A fully-connected conditional random field (CRF) mask is then applied to smooth the segmentation maps. We have adapted the W-net architecture for single-channel and three-channel synchronic maps, but no further optimization has been performed at the moment.

4 Preliminary Results

First order comparisons are limited to existing catalogues and methods as there is no objective way to define CH boundaries. The unsupervised learning methods included in SEARCH are implemented and tested using single-wavelength synchronic maps from the PSI database for CH evolution (see Section 2). The 6004 PSI synchronic maps can feature missing data due to the alignment of the satellites and the combination process. The database also includes publicly-available CH segmentation maps that were identified using EZSEG, a region-growing-style algorithm described in Caplan et al. (2016). The intersection over union metric (IoU) and the structural similarity index (SSIM) were used to measure the accuracy and the similarity of the K-means and W-net predictions with respect to the existing PSI CH detections, as described below. In this context, the IoU and SSIM metrics must be interpreted as a measure of consistency with existing methods, not as a measure of performance.

The K-means technique was trained using a set of 15 images to identify 6 clusters, *i.e.* the optimal number computed by the Elbow method, and then tested on approximately 6000 maps. Any missing data were removed from the K-means input EUV images as well as from PSI CH masks when performing comparisons. Among identified clusters are features that resemble closely CHs (*e.g.*,

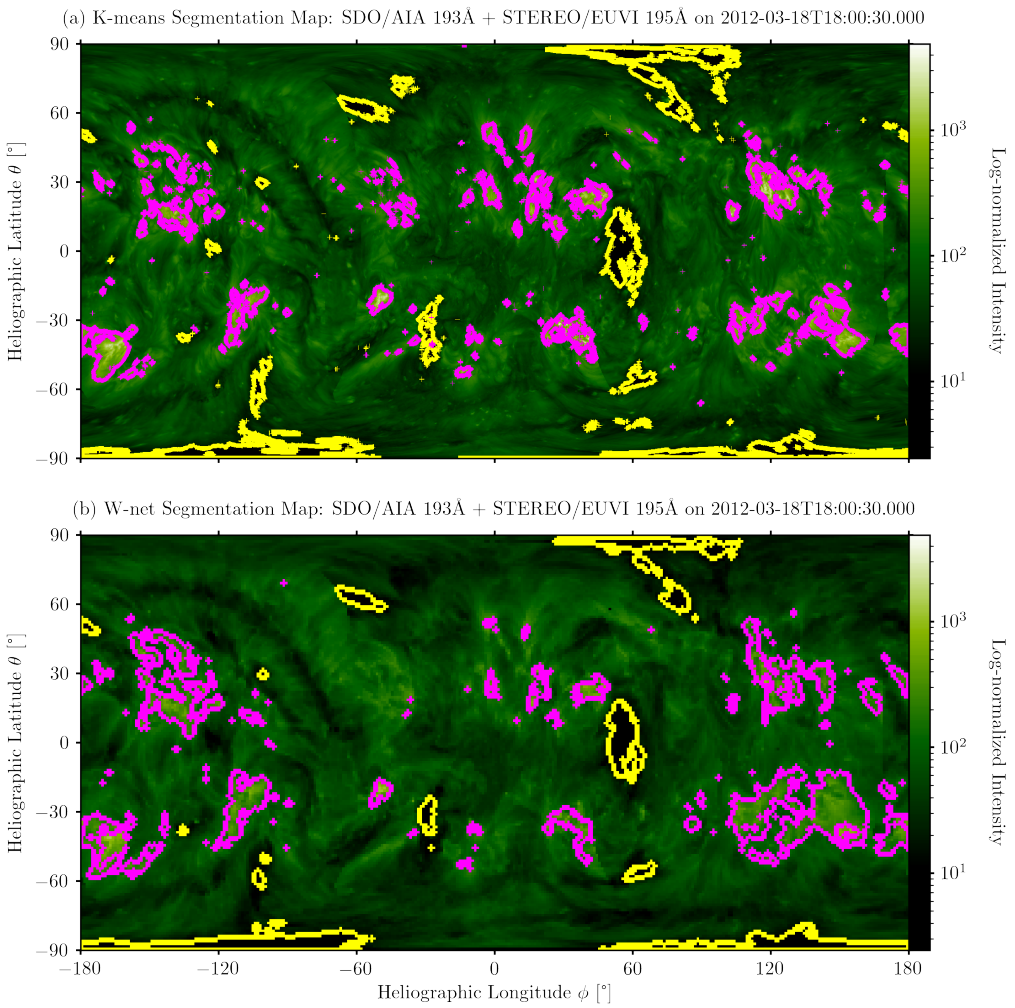

Figure 2: Segmentation maps generated by (a) the K-means method using 6 clusters and (b) the W-net from a synchronic map at 195 Å (displayed as colored background) found in the PSI database. Clusters that best identify CHs (resp. ARs) are segmented in yellow (resp. in magenta).

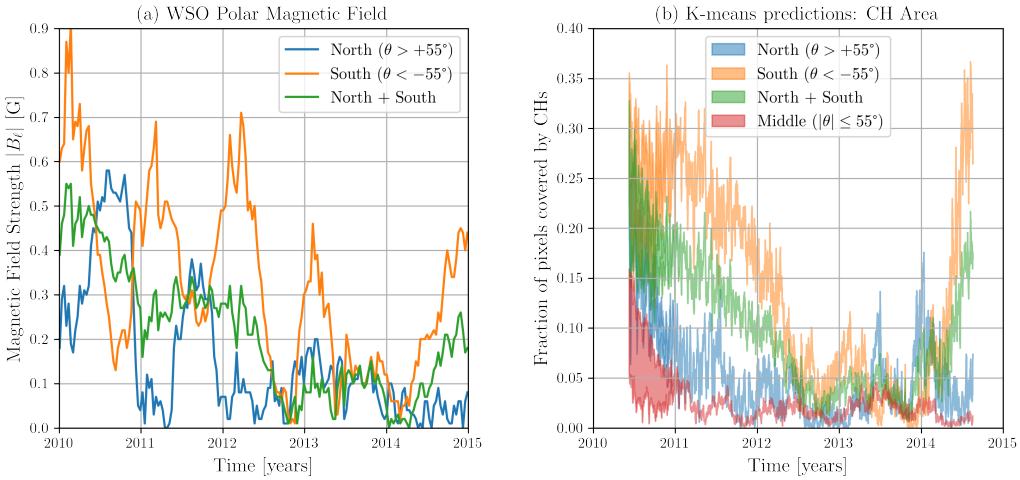

Figure 3: (a) Wilcox Solar Observatory polar magnetic field observations from 2010 to 2015. Poles are defined where the latitude $|\theta| > 55°$. (b) Evolution of polar CH areas inferred using K-means.

see yellow contours in Figure 2a) and, as a bonus, ARs (magenta contours). Mean values of SSIM and IoU when K-means is tested against the entire PSI database are 0.96 and 0.60 respectively, suggesting a good agreement. Metrics change slightly when considering regions with missing data in the comparison because the PSI database CHs were identified using a region-growth method.

Over 1500 PSI synchronic maps were used to train the W-net. Input EUV images and reference CH masks were downsampled by a factor of 10 in each spatial dimension for memory considerations. Pixels where data is missing were thresholded to the minimum intensity value. Preliminary tests performed on 4000 low-resolution synchronic maps with little to no missing data show promising results as the W-net identified clusters for both CHs and ARs (yellow and magenta contours in Figure 2b) which resemble closely that of the K-means method (Figure 2a). Mean values of 0.95 and 0.53 are achieved for the SSIM and IoU. However, W-net predictions from EUV images with a non-negligible fraction of missing data (*e.g.*, the PSI maps from the year 2010) result in departure from the expected solution (not shown). This can be explained by the downsampling and upsampling operations performed by the W-net. We expect that optimization of the data preparation and network architecture will further improve the W-net's ability at segmenting CHs and ARs.

4.1 Science Application: CH Area and Polar Magnetic Fields

The physics of CHs is still poorly constrained and even more so at the Sun's poles. The magnetic fields within a polar CH and the dynamics of the solar surface that drives the formation of this region are largely unknown. We investigated the correlation between polar magnetic field strengths and CH area between 2010 and 2015. The Wilcox Solar Observatory (WSO) has been performing observations[3] of the Sun's polar magnetic fields since 1976. Figure 3a shows measurements of the line-of-sight magnetic field strength that were collected for latitudes $|\theta| > 55^\circ$ and then corrected for projection effects using a 20 nHz low-pass filter. We estimated the area of (polar) CHs during the same time window by computing the fraction of pixels that were labeled as CHs through unsupervised learning. Figure 3b shows the possible range of CH areas estimated through K-means as a function of time. For each band of latitudes, the lower curve assumes that there are no CHs where data is unavailable (*i.e.*, lower limit) and the upper curves assumes that all unavailable data belongs to CHs (*i.e.*, upper limit). As the polar magnetic field strength in Figure 3a decreases with time and reaches a minimum some time between 2013 and 2014 (before undergoing a change in polarity), so does the area of polar CHs (Figure 3b). This is especially apparent for the South pole of the Sun (orange curve). Similar variations were found for polar CH areas estimated using the W-net (see Appendix A). However, it is clear that the area of polar CHs do not track the measured magnetic field variations at time scales less than a year nor are size and field strength linearly correlated. For example, the size of the southern polar hole was measured to be approximately of equal size in mid-2010 and mid-2014, yet the southern field strength in mid-2014 was nearly half of that measured in mid-2010. This means CH size measurement are a good proxy for the long-term changes in coronal structure, but more work is needed to causally link the two phenomena.

5 Conclusion and Future Work

We presented two unsupervised machine learning methods, *i.e.* K-means and W-net CNN, that segment synchronic maps to determine solar CH boundaries. These methods are in good agreement with other publicly-available CH segmentation maps. In addition to CHs, both methods identified features consistent with ARs. Promising preliminary results warrant further exploration of single- and multi-wavelength synchronic maps, most notably for the W-net, which was not optimized. Future venues to explore with the SEARCH project include (1) the relationship between polar CH area and geomagnetic activity using data from the Advanced Composition Explorer (ACE), (2) the correlation between CHs (open magnetic flux sources) and space weather using legacy EIT synoptic maps which cover a full 22-year solar magnetic cycle, and (3) ARs identification and segmentation. Previous studies performed on synoptic maps have suggested a relationship between solar dynamo and the size of polar CHs (*e.g.*, Kirk et al., 2009; Waldmeier, 1981; Bravo and Stewart, 1994; Harvey and Recely, 2002). Improvements to our understanding of the solar dynamo will also extend to Sun-like stars. Finally, this study highlights the need for a multi-stereoscopic view of the Sun[4] to further our understanding of the Sun and the space weather (Morton et al., 2016).

[3] Wilcox Solar Observatory polar field observations: `http://wso.stanford.edu/Polar.html` [4] SULIS: `https://sulis.space`

6 Broader Impact

Space weather refers to conditions on the Sun and the extended solar atmosphere (heliosphere) that interacts with the immediate surrounding of the Earth's atmosphere. The conditions of this atmosphere can influence the performance and reliability of space-based activities including and not limited to satellite navigation, communication and space travel, in a similar way to meteorological weather and how it affects terrestrial activities. The modern society has become ever so reliant on space based satellites, and this dependence is continuing to grow with the advent of newer and faster modes of communications such as 5G. Such technological dependence is increasingly vulnerable to disturbances from outside of the Earth system. Space weather and the importance of understanding space weather drivers has been established over the past decades. Various government agencies have started paying attention to the space weather, the House of Congress recently passed the Promoting Research and Observations of Space Weather to Improve the Forecasting of Tomorrow Act, 2020 (PROSWIFT Act 2020) emphasising and acknowledging the importance of space weather and space weather drivers in the future. Understanding the nature of polar coronal holes and the associated large scale open-magnetic fluxes are one of the most important aspects of understanding space weather.

Acknowledgments and Disclosure of Funding

This project was first conducted as part of the NASA HelioHackweek[5] 2020. The authors would like to thank their fellow SEARCH team members Emily Mason, Jake Strang, Jamie Staeben, and Luisa Capannolo for participating in this project and contributing to the discussions. The authors also thank the organizers of this event at NASA and at the NASA Center for Climate Simulation (NCCS). We would also like to acknowledge the team at NVIDIA who provided us with the necessary resources during the course of this event. A.K.T would like to thank Shivangee Rathi, for valuable discussion and for Python troubleshooting. Finally, B.T. would also like to thank Prof. Benjamin Brown for his help with the SEARCH project.

A Appendix: W-net Predictions of Polar CH Areas as a Function of Time

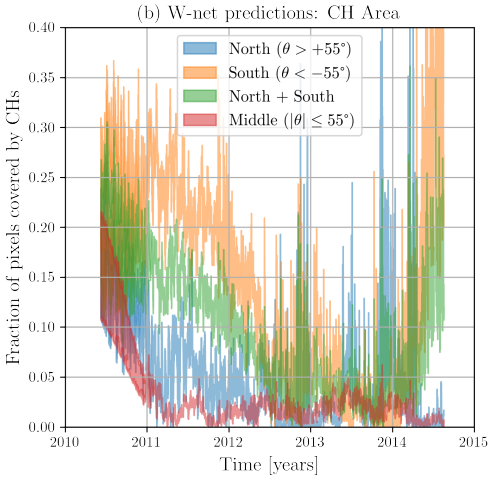

Figure 4: Evolution of polar CH areas inferred by the W-net from synchronic maps that were downsampled by a factor of 10 in each dimension. Poles are defined where the latitude $|\theta| > 55°$.

[5] NASA HelioHackweek 2020 event page: `https://heliohackweek.github.io/`.

References

Bohlin, J. D., 1977. An observational definition of coronal holes. In J. B. Zirker, ed., Coronal Holes and High Speed Wind Streams, 27–69.

Bravo, S., and G. Stewart, 1994. Evolution of Polar Coronal Holes and Sunspots during SOLAR-CYCLE-21 and SOLAR-CYCLE-22. *Solar Phys.*, **154**(2), 377–384. 10.1007/BF00681106.

Caplan, R. M., C. Downs, and J. A. Linker, 2016. Synchronic Coronal Hole Mapping Using Multi-instrument EUV Images: Data Preparation and Detection Method. *Astrophys. J.*, **823**(1), 53. 10.3847/0004-637X/823/1/53, 1510.04718.

Delaboudinière, J. P., G. E. Artzner, J. Brunaud, A. H. Gabriel, J. F. Hochedez, et al., 1995. EIT: Extreme-Ultraviolet Imaging Telescope for the SOHO Mission. *Solar Phys.*, **162**(1-2), 291–312. 10.1007/BF00733432.

Garton, T. M., P. T. Gallagher, and S. A. Murray, 2018. Automated coronal hole identification via multi-thermal intensity segmentation. Journal of Space Weather and Space Climate, **8**, A02. 10.1051/swsc/2017039, 1711.11476.

Hamada, A., T. Asikainen, and K. Mursula, 2020. New Homogeneous Dataset of Solar EUV Synoptic Maps from SOHO/EIT and SDO/AIA. *Solar Phys.*, **295**(1), 2. 10.1007/s11207-019-1563-y, 1910.13720.

Harvey, J. W., and N. R. Sheeley, 1978. Coronal holes, solar wind streams, and geomagnetic activity during the new sunspot cycle. *Solar Phys.*, **59**(1), 159–173. 10.1007/BF00154939.

Harvey, K. L., and F. Recely, 2002. Polar Coronal Holes During Cycles 22 and 23. *Solar Phys.*, **211**(1), 31–52. 10.1023/A:1022469023581.

Hess Webber, S. A., N. Karna, W. D. Pesnell, and M. S. Kirk, 2014. Areas of Polar Coronal Holes from 1996 Through 2010. *Solar Phys.*, **289**(11), 4047–4067. 10.1007/s11207-014-0564-0.

Illarionov, E. A., A. Kosovichev, and A. Tlatov, 2020. Machine-learning approach to identification of coronal holes in solar disk images and synoptic maps. arXiv e-prints, arXiv:2006.08529. 2006.08529.

Illarionov, E. A., and A. G. Tlatov, 2018. Segmentation of coronal holes in solar disc images with a convolutional neural network. *Mon. Not. Roy. Astron. Soc.*, **481**(4), 5014–5021. 10.1093/mnras/sty2628, 1809.05748.

Joshi, K. D., and P. S. Nalwade, 2013. Modified K-Means for Better Initial Cluster Centres.

Karna, N., S. A. Hess Webber, and W. D. Pesnell, 2014. Using Polar Coronal Hole Area Measurements to Determine the Solar Polar Magnetic Field Reversal in Solar Cycle 24. *Solar Phys.*, **289**(9), 3381–3390. 10.1007/s11207-014-0541-7.

Kirk, M. S., W. D. Pesnell, C. A. Young, and S. A. Hess Webber, 2009. Automated detection of EUV Polar Coronal Holes during Solar Cycle 23. *Solar Phys.*, **257**(1), 99–112. 10.1007/s11207-009-9369-y, 0901.1158.

Lemen, J. R., A. M. Title, D. J. Akin, P. F. Boerner, C. Chou, et al., 2012. The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Phys.*, **275**(1-2), 17–40. 10.1007/s11207-011-9776-8.

Lowder, C., J. Qiu, R. Leamon, and Y. Liu, 2014. Measurements of EUV Coronal Holes and Open Magnetic Flux. *Astrophys. J.*, **783**(2), 142. 10.1088/0004-637X/783/2/142, 1502.06038.

Morton, R. J., E. Scullion, D. S. Bloomfield, J. A. McLaughlin, S. Regnier, S. W. McIntosh, S. Tomczyk, and P. Young, 2016. Exploring Coronal Dynamics: A Next Generation Solar Physics Mission white paper. arXiv e-prints, arXiv:1611.06149. 1611.06149.

Nolte, J. T., A. S. Krieger, A. F. Timothy, R. E. Gold, E. C. Roelof, G. Vaiana, A. J. Lazarus, J. D. Sullivan, and P. S. McIntosh, 1976. Coronal holes as sources of solar wind. *Solar Phys.*, **46**(2), 303–322. 10.1007/BF00149859.

Pesnell, W. D., and K. H. Schatten, 2018. An Early Prediction of the Amplitude of Solar Cycle 25. *Solar Phys.*, **293**(7), 112. 10.1007/s11207-018-1330-5.

Rajendrakumar Gare, G., J. Li, R. Joshi, M. Prashant Vaze, R. Magar, M. Yousefpour, R. L. Rodriguez, and J. M. Galeotti, 2020. W-Net: Dense Semantic Segmentation of Subcutaneous Tissue in Ultrasound Images by Expanding U-Net to Incorporate Ultrasound RF Waveform Data. arXiv e-prints, arXiv:2008.12413. 2008.12413.

Schatten, K., 2005. Fair space weather for solar cycle 24. *Geophys. Res. Lett.*, **32**(21), L21106. 10.1029/2005GL024363.

Schatten, K. H., P. H. Scherrer, L. Svalgaard, and J. M. Wilcox, 1978. Using Dynamo Theory to predict the sunspot number during Solar Cycle 21. *Geophys. Res. Lett.*, **5**(5), 411–414. 10.1029/GL005i005p00411.

Stenborg, G., A. Vourlidas, and R. A. Howard, 2008. A Fresh View of the Extreme-Ultraviolet Corona from the Application of a New Image-Processing Technique. *Astrophys. J.*, **674**(2), 1201–1206. 10.1086/525556.

Svalgaard, L., E. W. Cliver, and Y. Kamide, 2005. Sunspot cycle 24: Smallest cycle in 100 years? *Geophys. Res. Lett.*, **32**(1), L01104. 10.1029/2004GL021664.

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

Waldmeier, M., 1981. Cyclic Variations of the Polar Coronal Hole. *Solar Phys.*, **70**(2), 251–258. 10.1007/BF00151332.

Wang, Y. M., and J. Sheeley, N. R., 1990. Magnetic Flux Transport and the Sunspot-Cycle Evolution of Coronal Holes and Their Wind Streams. *Astrophys. J.*, **365**, 372. 10.1086/169492.

Wuelser, J.-P., J. R. Lemen, T. D. Tarbell, C. J. Wolfson, J. C. Cannon, et al., 2004. EUVI: the STEREO-SECCHI extreme ultraviolet imager. In S. Fineschi and M. A. Gummin, eds., Telescopes and Instrumentation for Solar Astrophysics, vol. 5171 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 111–122. 10.1117/12.506877.