Using Convolutional Neural Networks for Image-Based Forecasting of Coronal Mass Ejections in Active Regions

Julio Hernandez Camero[1,2], Lucie M. Green[1,2], Thomas D. Kitching[1,2]

[1]Mullard Space Science Laboratory, UK, [2]University College London, UK

1 FORECASTING CMEs

Forecasting Coronal Mass Ejections (CMEs) is relevant both from a safety point of view, allowing us to prepare for possible disruptions to our systems, but also from a scientific perspective. Any method that accurately predicts these events has the potential to inform our knowledge about the physical processes behind them.

Thus, efforts to predict CMEs are required that focus on the nature of these events. So far, most works have focused on flare prediction, with CMEs either not being considered or considered as a secondary event. We consider this to be due to the difficulty in associating CMEs to their source regions, and thus a lack of data for producing any forecasting models. Meanwhile, flare catalogues were developed long ago and contain numerous events. Whatever the reason, while flares and CMEs are related, they do not occur on a one-to-one basis. Therefore, it is not the same to ask when a flare will produce a CME as it is asking when an active region in the Sun will produce a CME.

With our work we want to focus on the forecast of CMEs directly.

2 CME SOURCE REGION CATALOGUE

While there are some existing catalogues of varying size relating CMEs to active regions, these are crafted manually. We believe that manual attributions raise two concerns. First, they're time consuming. Second, associations are opaque and users of the catalogue can not scrutinize them and evaluate their trustworthiness on any criteria other than trusting the person who made the association.

To tackle this issue, we propose an automated algorithm that matches CMEs seen in coronagraph data to SDO/HMI SHARP regions using post-eruptive signatures. In particular, we make use of flares [1] and dimmings [2], both known to be associated with CMEs [3,4]. Associations are given a verification level representing our confidence in them. By providing an open association procedure and clearly defined verification levels, associations can easily be scrutinized by users of the catalogue, building trust based on the method rather than on trust in the author.

Overall, we found 1094 CME-CME source region associations between 2010 and 2018. About 20% of these CMEs do not have an associated flare.

3 CNN MODEL

As a first use of our catalogue for the forecasting of CMEs, we implement a binary classification model.

Our input data are cutouts of the three magnetic field components of the SHARP regions from the SDOML dataset [5]. We use different ResNet architectures to obtain the probability of a region producing a CME in the next 48 hours based on the input images and the number of CMEs in the region prior to the input.

Our model obtains a True Skill Score of 0.42, a better performance than random predictions.

These results hint that forecasting CMEs is possible with ML techniques, though further work is in progress to understand what inputs, architectures and metrics can help improve predictions and interpretation of the models.

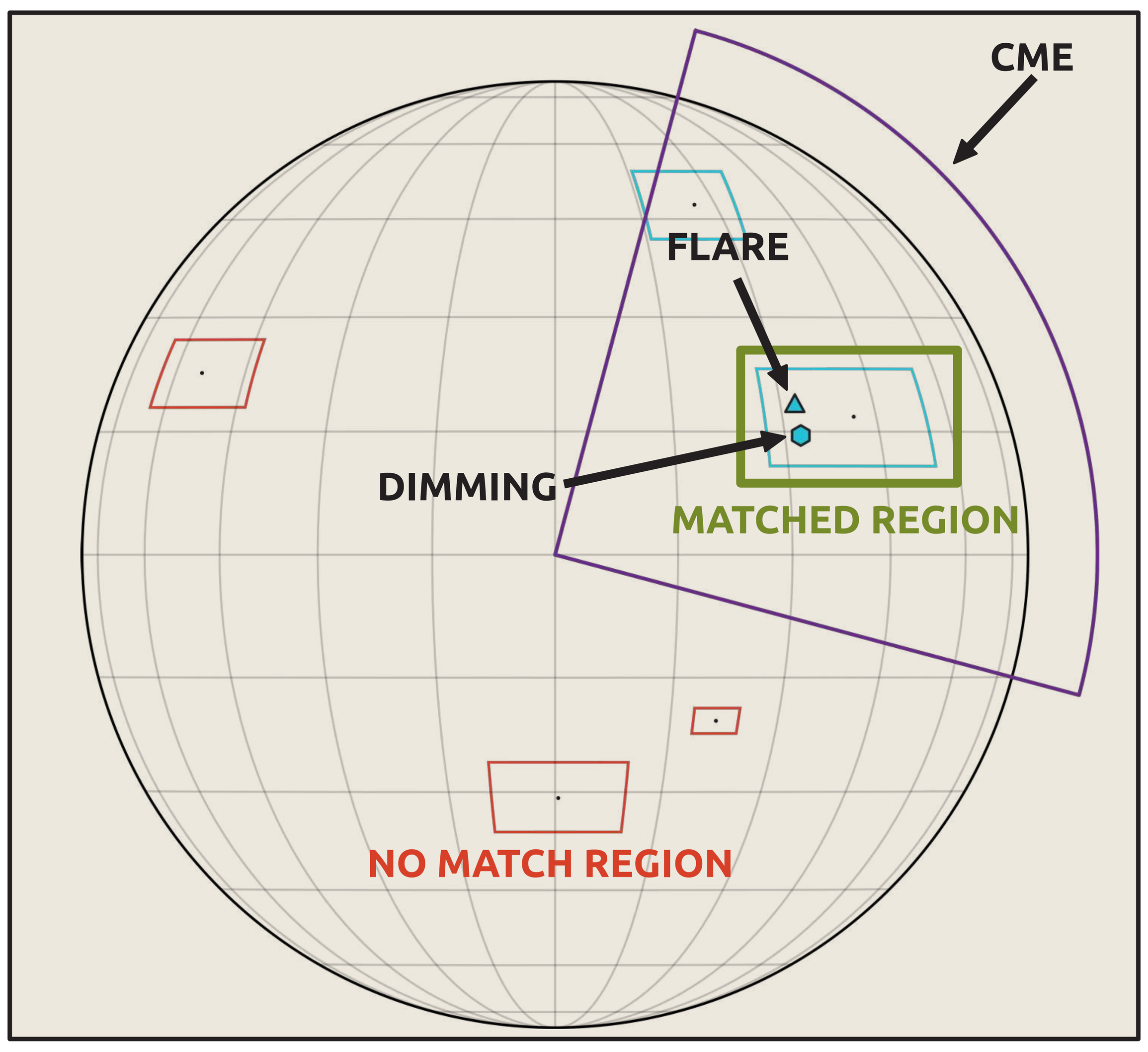

Figure 2: Example CME match to a region with a dimming and a flare. Blue regions are spatially consistent with the CME. Red are not.

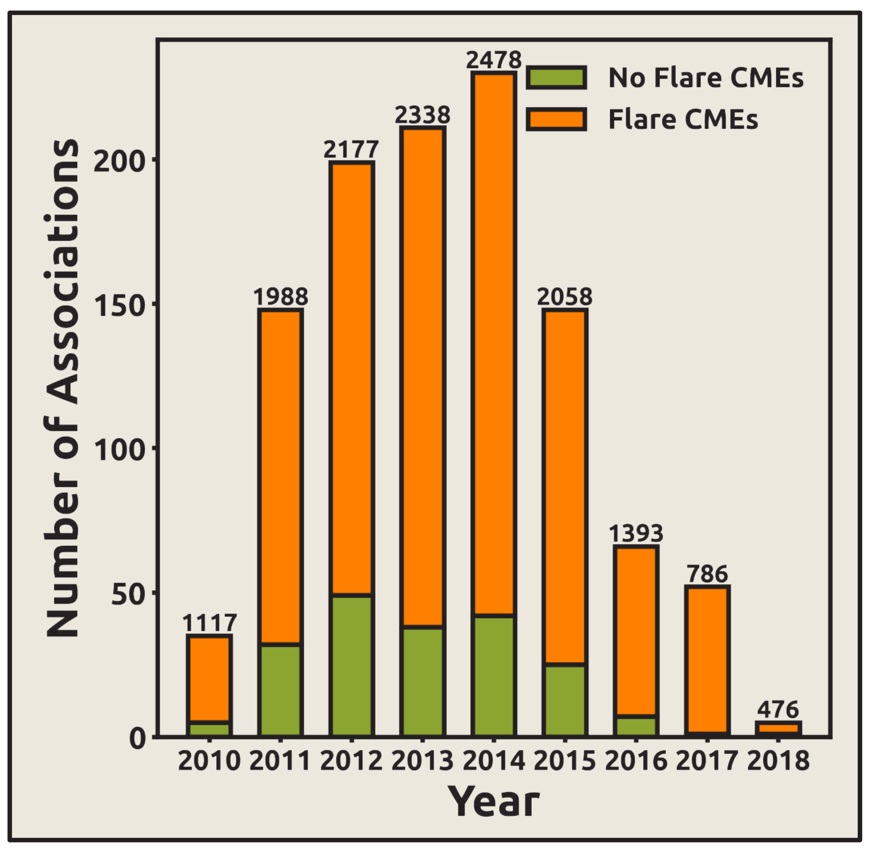

Figure 1: Summary of the number of associations per year, with the total number of CMEs in the LASCO CME catalogue per year over the bars.

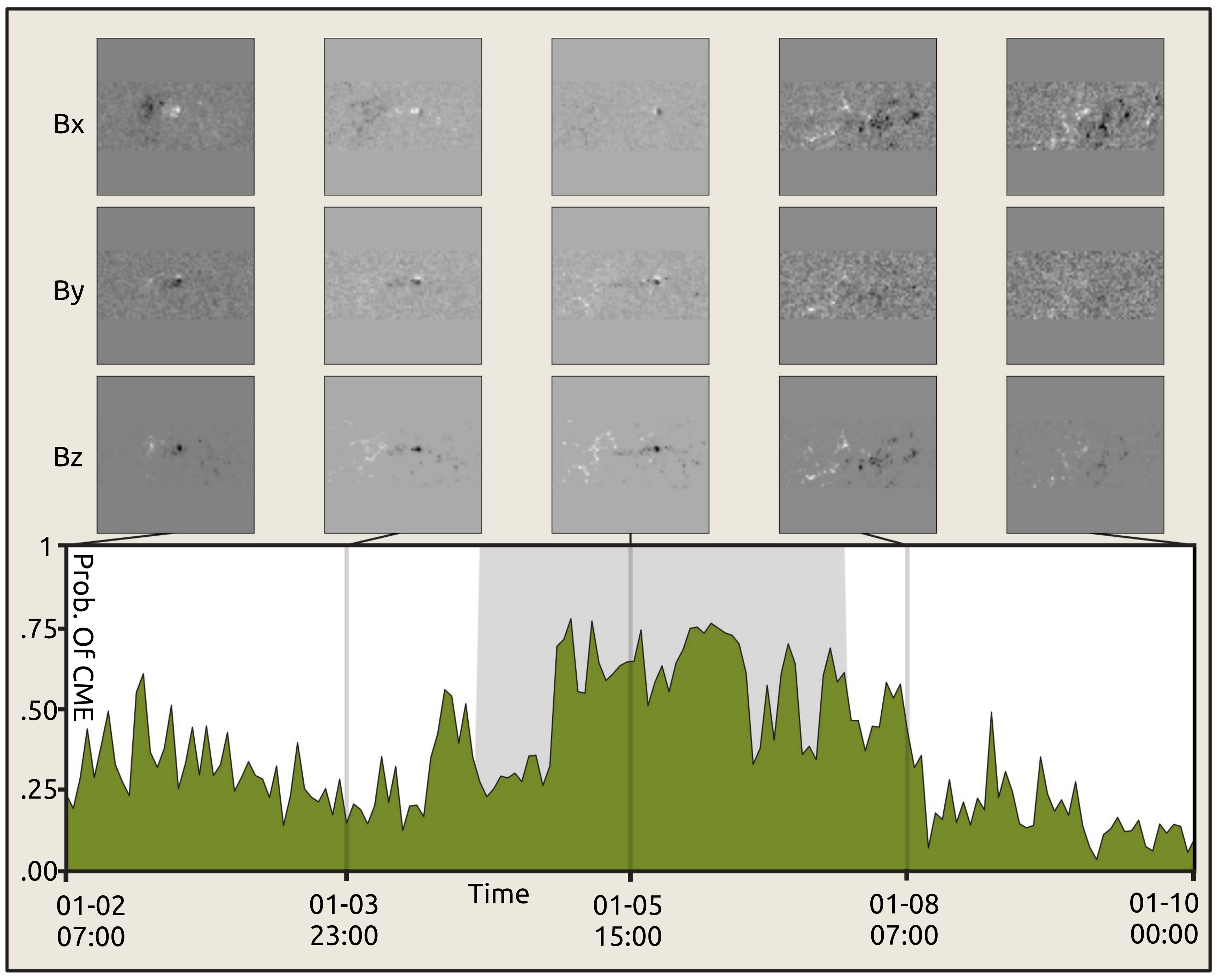

Figure 3: Model predictions for region 3560 with some of the images used to make the predictions. The green shaded area are the model predicted probabilities.

While the model obtains a TSS of 0.42, it is hard to say whether the model really understands CME eruptions when looking at its actual predictions. Using a single metric to describe a model may not be enough.

4 CONCLUSION

We propose an automated algorithm for the association of CMEs with their source regions using post-eruptive signatures. This allows for easier scrutiny of the associations by users of the catalogue.

Using this catalogue, we train a binary classification ML model. The model predicts the probability that a SHARPs region will produce a CME in the next 48 hours, obtaining relatively good results according to widely used metrics. We raise the question whether using a single metric is enough to describe a model's performance.

Next steps will focus on the inclusion of coronal images into models and the use of recurrent neural networks to test whether they improve the performance. We also intend to explore other ways of describing the spatial distribution of features in images without the need to use the images themselves, through topological data analysis and spatial statistics methods [6].

Scan for more information and to contact me!
Or contact me directly at julio.camero.21@ucl.ac.uk

[1] Angryk et al. Sci Data. 2020 Dec;7(1):227.
[2] Kraaikamp E, Verbeeck C. J Space Weather Space Clim. 2015;5:A18.
[3] Low BC. 1996 Jan 1;95:148.
[4] Bewsher D, Harrison RA, Brown DS. A&A. 2008 Feb;478(3):897–906.
[5] Galvez R et al. ApJS. 2019 May 1;242(1):7.
[6] Sun H. Space Weather. 2021 Dec;19(12):e2021SW002837.