IDENTIFYING FUTURE CME PRODUCTIVE ACTIVE REGIONS USING AN AUTOMATICALLY GENERATED CME SOURCE REGION CATALOGUE AND A MACHINE LEARNING MODEL

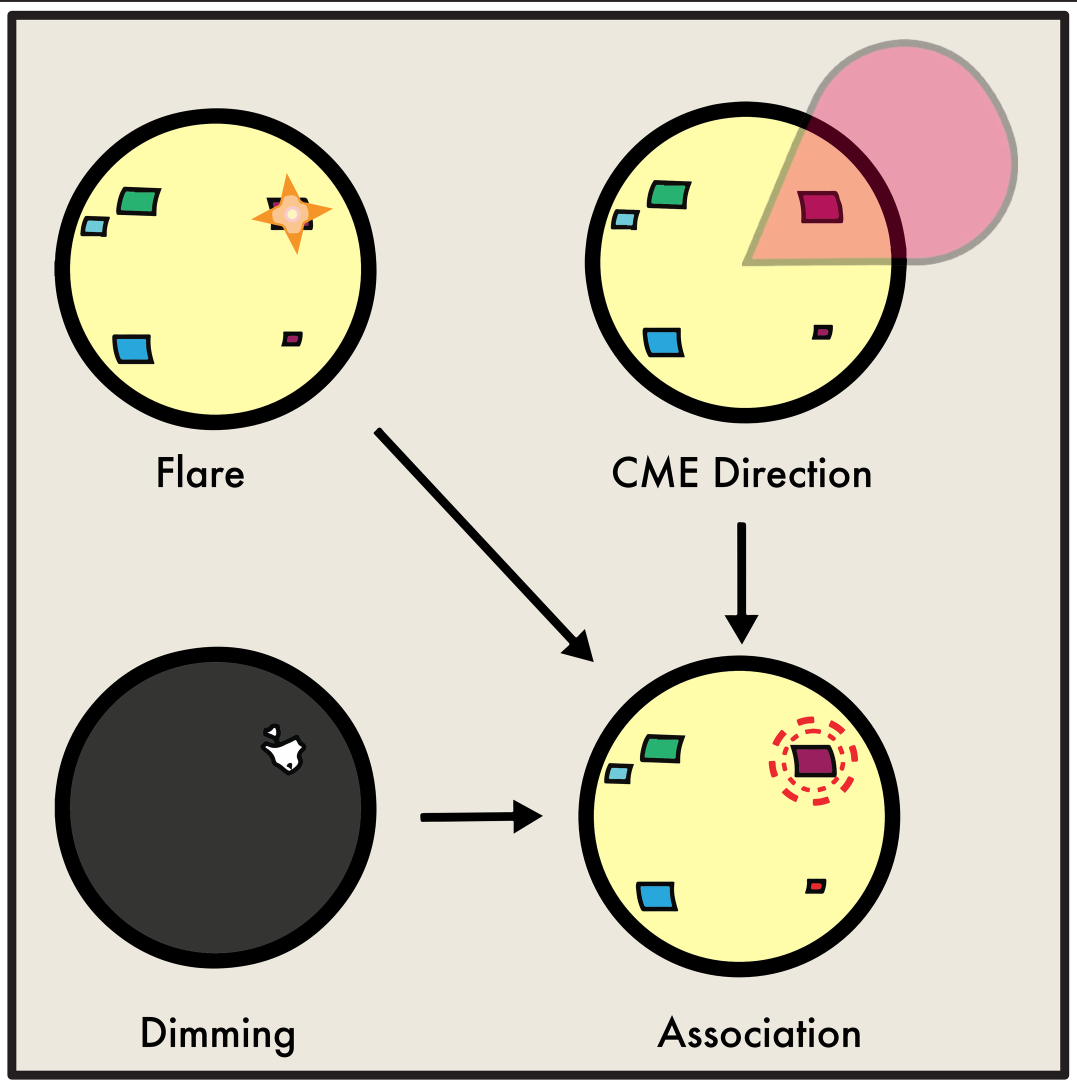
Julio Hernandez Camero[1,2], Lucie M. Green[1,2], Alex Piñel Neparidze[2]

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

1 CATALOGUE

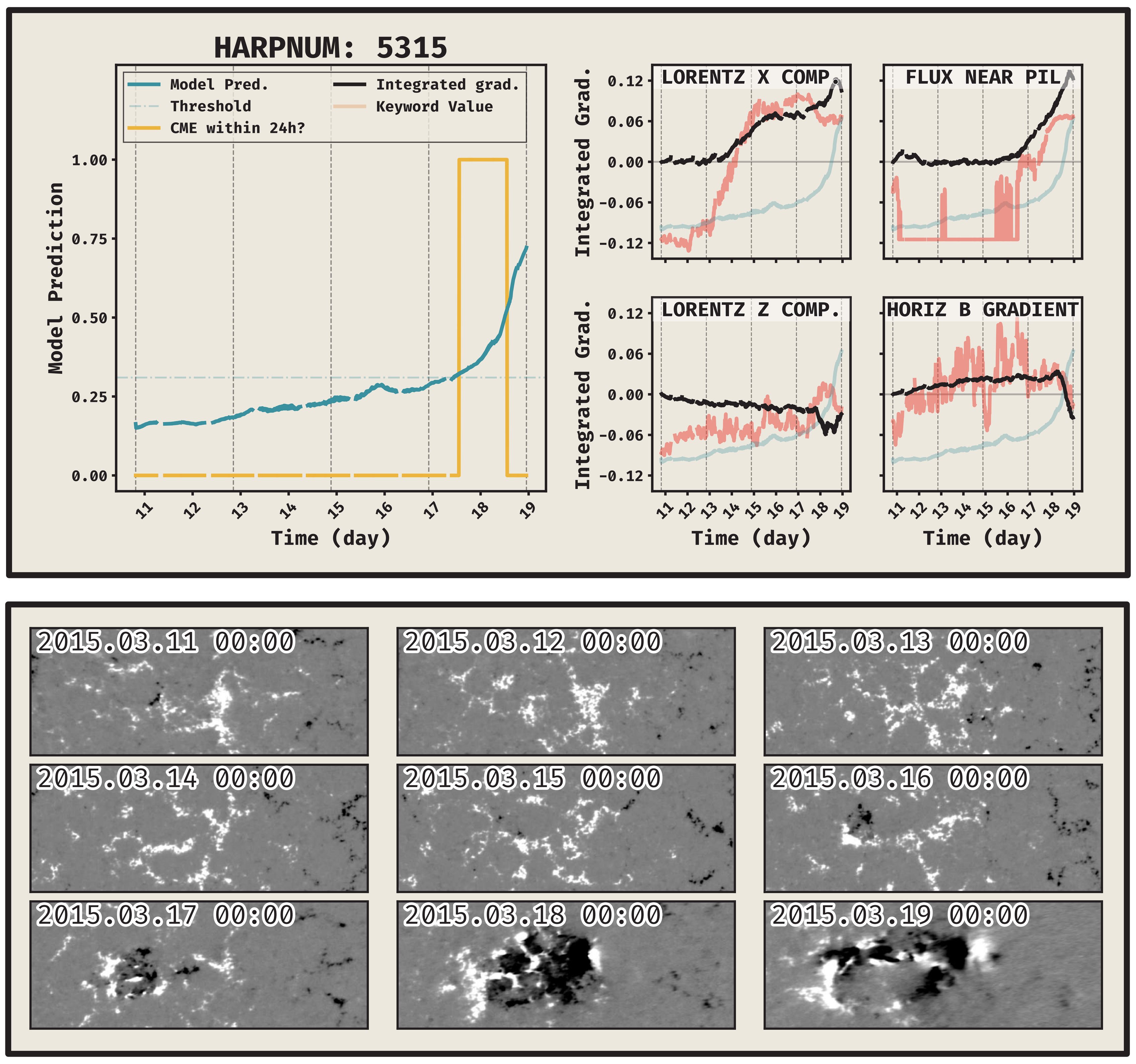
First, we create a catalogue of CME source regions by automatically matching CMEs observed in LASCO to a Spaceweather HMI Active Region Patch (SHARP). We identify potential source regions based on their location on the Sun (see Figure A) and look for post-eruptive signatures such as dimmings and flares. A match is made if any region exhibits a dimming or flare that can be linked to the CME. By running our algorithm on data from 2016 to 2019, we successfully matched 1,132 CMEs, achieving a recovery rate of approximately 57%.

A



A simplified illustration of our automated algorithm for matching a CME to a SHARP region. The match is based on the region's location (top right) and the presence of post-eruptive signatures such as flares (top left) or dimmings (bottom right). A region is matched to a CME if it is in the correct location and exhibits at least one of these signatures.

B FLUX EMERGENCE REGION



Figures B and C: These figures display our model's predictions for CME activity in rolling 24-hour time windows in two distinct active regions over the next 24 hours. In the top left of each figure, the model's overall prediction is shown, while the top right presents the four most significant SHARP keywords contributing to the prediction (arranged in a 2x2 grid). For each keyword, we visualize three components: integrated gradients (black), representing the contribution of the keyword's evolution over the past 24 hours to the prediction; the actual keyword evolution over time (red); and the model's prediction for comparison (blue). The bottom panel shows snapshots of the magnetic field evolution in each region.

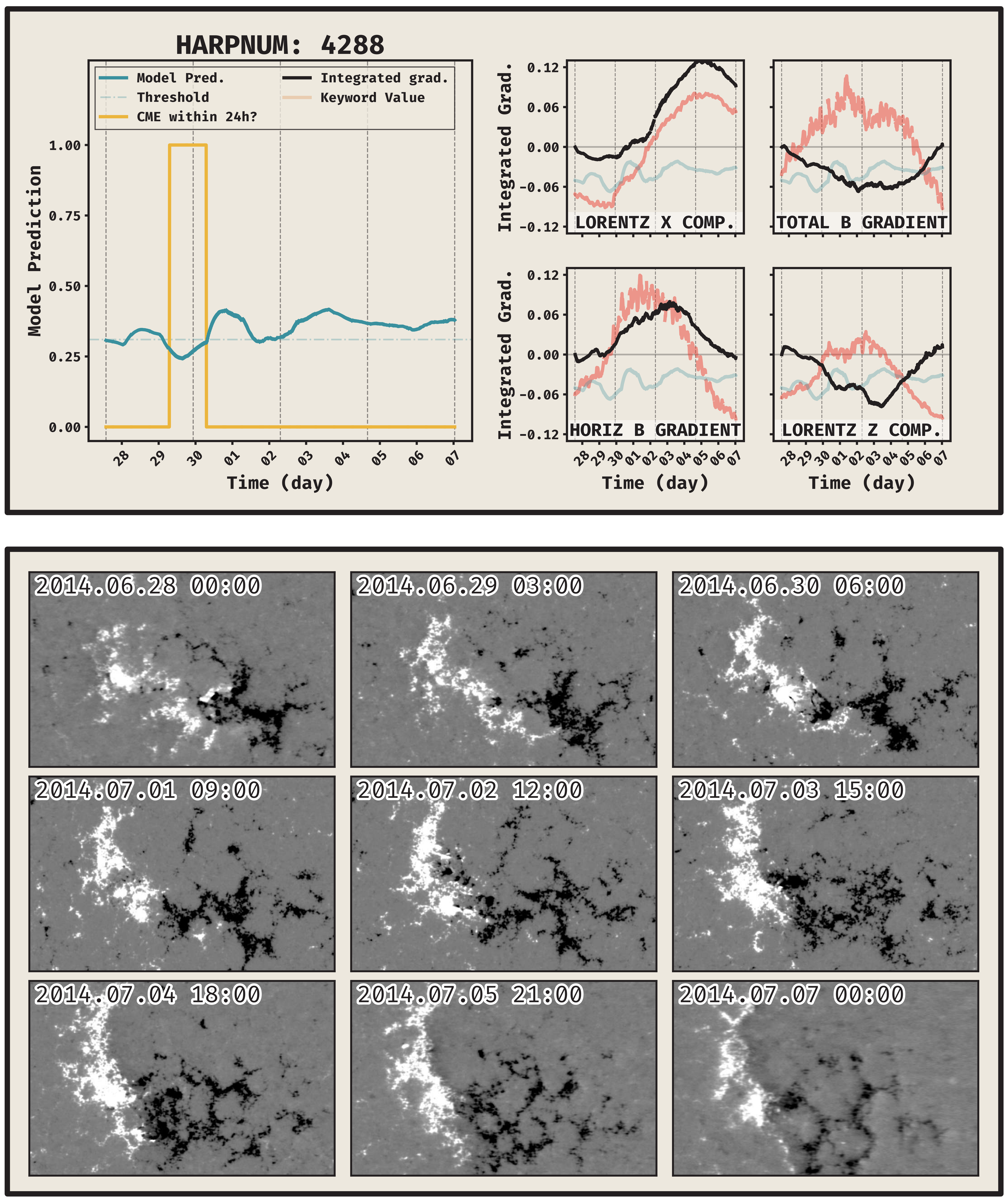
2 MODEL

We then build a machine learning model that aims to forecast CMEs 24 hours in advance, built on a transformer architecture. The input consists of the evolution of 40 features, including SHARP keywords that describe the magnetic field of a SHARP region, along with custom history parameters representing the region's past activity. The output is a binary prediction indicating whether a CME will occur in the next 24 hours. While the model struggles with precise event timing, it excels at distinguishing between CME-quiet and CME-productive regions, except in cases of rapid flux emergence (see Figure B).

KEY TAKEAWAYS

- We have developed an automated algorithm to match CMEs with their source active regions, making both the source code and the resulting catalogue publicly available.
- We trained a transformer-based machine learning model to forecast CMEs 24 hours in advance based on sequences of SHARP keywords.
- We rank regions by forecast accuracy and analyse correlations between keywords and model output. Using integrated gradients, we identify which keyword's evolution most strongly influence the predictions for each region
- The model performs better in regions exhibiting significant flux emergence prior to the eruption compared to those driven by flux cancellation.
- This could be due to the 24-hour window given to the model, which may be more responsive to rapid changes in flux emergence but less sensitive to the slower processes associated with flux cancellation.

C FLUX CANCELLATION REGION



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

MULLARD SPACE SCIENCE LABORATORY