Interpreting Transformer-Based CME Forecasting and the Role of Flare Associations

Julio Hernandez Camero, Lucie M. Green

Department of Climate and Space Physics
University College London

7th April 2025

Introduction

Coronal Mass Ejections (CMEs)

- CMEs are one of the main drivers of space weather.
- Forecasting which active regions in the Sun produce them and when can help:
  - Better prepare our infrastructure for potential impacts.
  - Potentially better understand the physical mechanisms behind these eruptions.

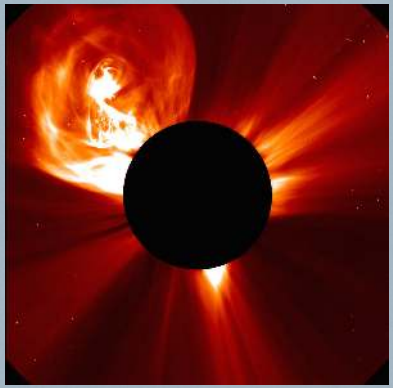


Image credit: ESA/NASA SOHO/LASCO

Introduction

Coronal Mass Ejections (CMEs)

- CMEs are one of the main drivers of space weather.
- Forecasting which active regions in the Sun produce them and when can help:
  - Better prepare our infrastructure for potential impacts.
  - Potentially better understand the physical mechanisms behind these eruptions.

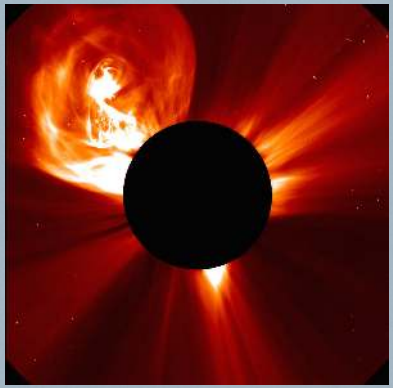


Image credit: ESA/NASA SOHO/LASCO

Introduction

Coronal Mass Ejections (CMEs)

- CMEs are one of the main drivers of space weather.
- Forecasting which active regions in the Sun produce them and when can help:
  - Better prepare our infrastructure for potential impacts.
  - Potentially better understand the physical mechanisms behind these eruptions.

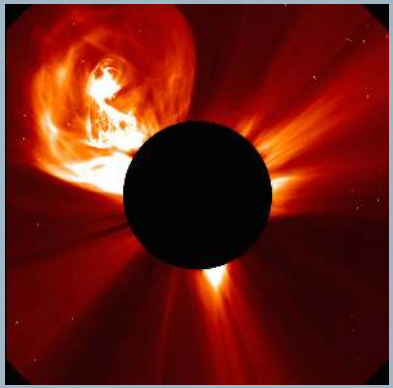


Image credit: ESA/NASA SOHO/LASCO

Introduction

Coronal Mass Ejections (CMEs)

- CMEs are one of the main drivers of space weather.
- Forecasting which active regions in the Sun produce them and when can help:
  - Better prepare our infrastructure for potential impacts.
  - Potentially better understand the physical mechanisms behind these eruptions.

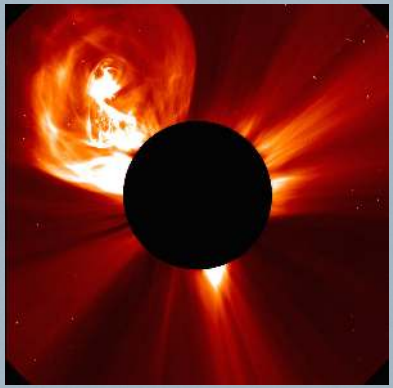
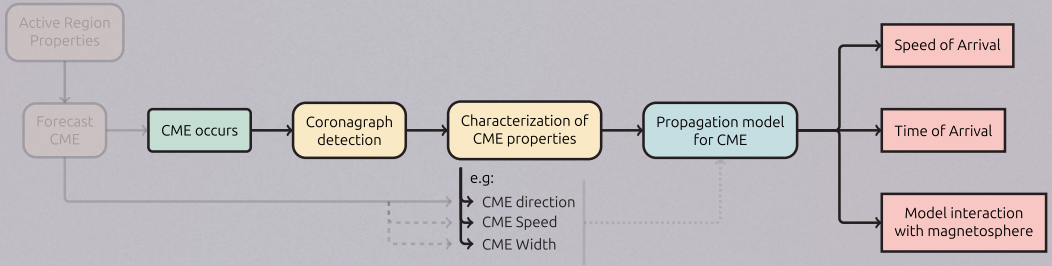


Image credit: ESA/NASA SOHO/LASCO

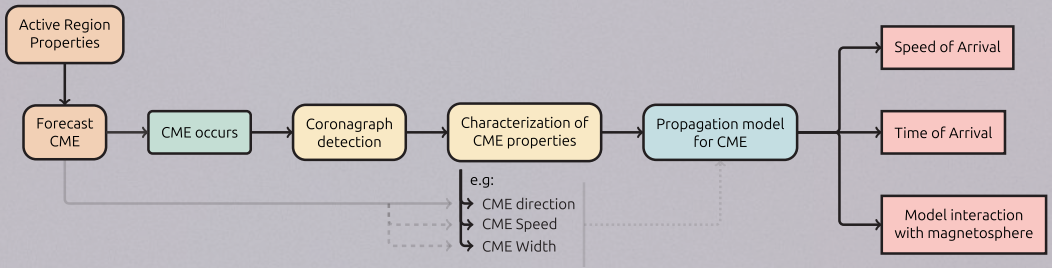
Introduction

Simple CME impact pipeline



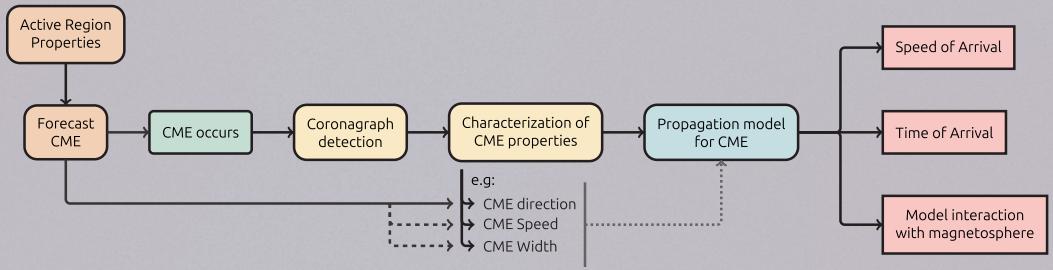
Introduction

Simple CME impact pipeline



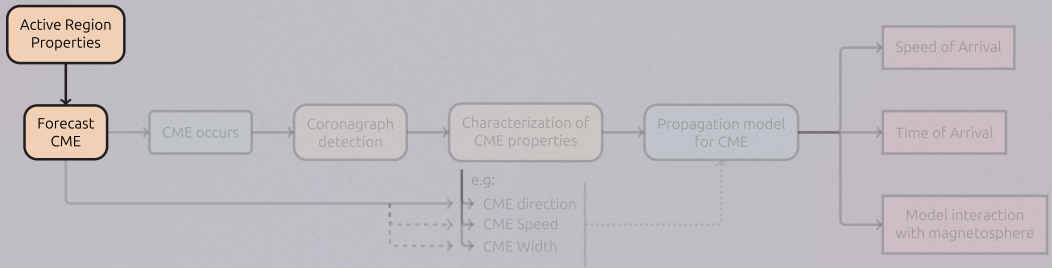
Introduction

Simple CME impact pipeline



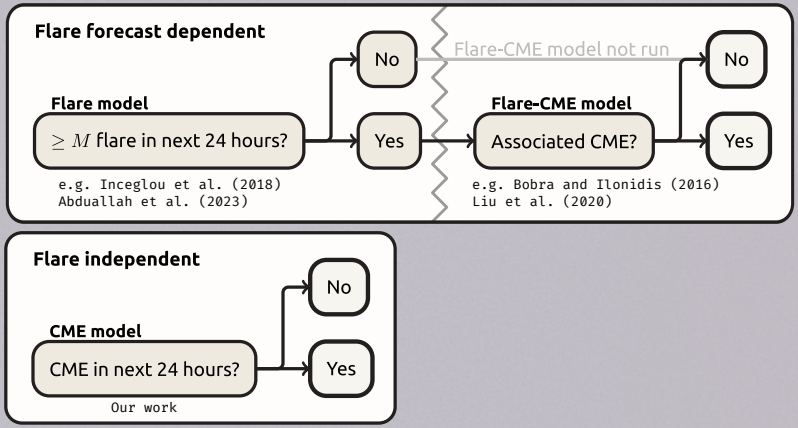
Introduction

Simple CME impact pipeline



Introduction

Options for forecasting CMEs



Model setup

SHARP keywords (Bobra et al. (2018), Angryk et al. (2020))

- We produce forecasts for Space weather HMI Active Region Patches (SHARPs) between 2010 and 2018.
- Track magnetic field concentrations and *tend* to coincide with NOAA active regions (not always).
- Parameters describing the magnetic field are calculated every 12 minutes. These are our inputs.
- CMEs matched to regions in our previous work (find in **julhcam.com**).

Model setup

SHARP keywords (Bobra et al. (2018), Angryk et al. (2020))

- We produce forecasts for Space weather HMI Active Region Patches (SHARPs) between 2010 and 2018.
- Track magnetic field concentrations and *tend* to coincide with NOAA active regions (not always).
- Parameters describing the magnetic field are calculated every 12 minutes. These are our inputs.

- CMEs matched to regions in our previous work (find in **julhcam.com**).

Model setup

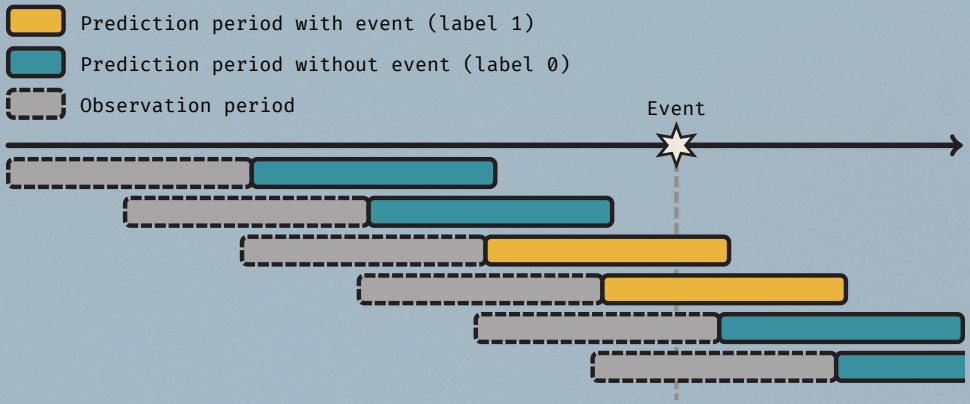
SHARP keywords (Bobra et al. (2018), Angryk et al. (2020))

- We produce forecasts for Space weather HMI Active Region Patches (SHARPs) between 2010 and 2018.
- Track magnetic field concentrations and *tend* to coincide with NOAA active regions (not always).
- Parameters describing the magnetic field are calculated every 12 minutes. These are our inputs.

- CMEs matched to regions in our previous work (find in **julhcam.com**).

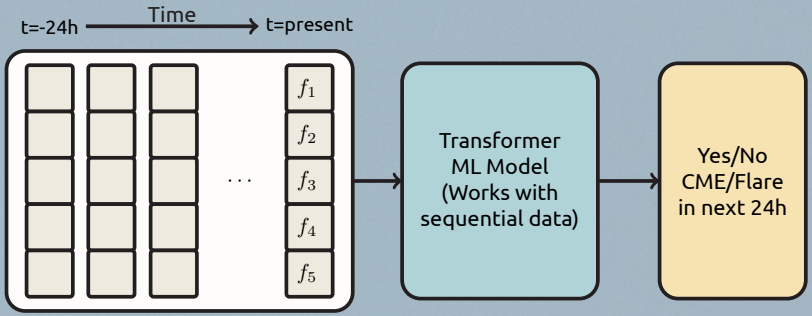
Model setup

Forecasting setup



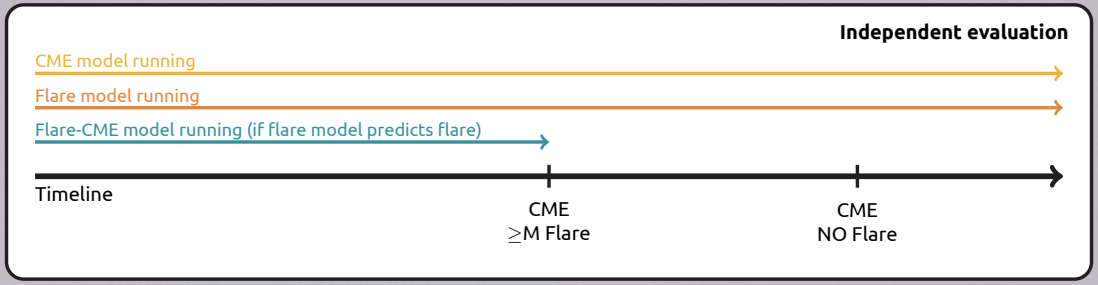
Model setup

Forecasting setup



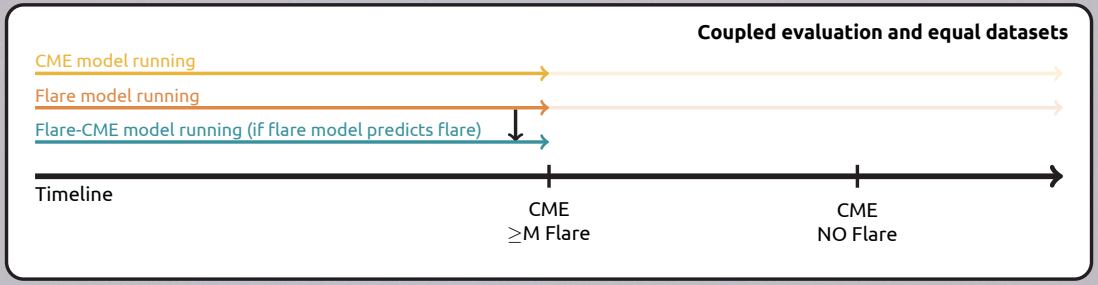
Model performance

Two kinds of evaluations



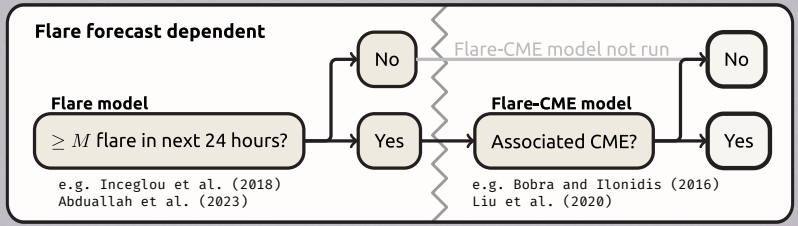
Model performance

Two kinds of evaluations



Model performance

Options for forecasting CMEs



Model performance

Overall performances

Table: Individual model performance

Forecast Type	TSS
Flare	0.82
CME	0.52
Flare-CME	0.13

Table: Combined model performance. Evaluated on equal datasets.

Forecast Type	TSS
CME	0.00
Flare + Flare-CME	0.13

Model performance

Some takeaways and hypotheses

- Models that forecast CMEs struggle more: perhaps not as many signatures in photospheric magnetic field?
- TSS inflates the performance of the models without knowledge of event distribution in dataset (look out for submitted paper).
- When evaluated in 24h prior to $\geqslant$M flare, flare-CME model does better than CME.
  - Intuitively makes sense, but both are looking for CMEs
  - Hint that models are looking for different signatures
  - **Next: Can we use explainability methods to see where those two models differ?**

Model performance

Some takeaways and hypotheses

- Models that forecast CMEs struggle more: perhaps not as many signatures in photospheric magnetic field?
- TSS inflates the performance of the models without knowledge of event distribution in dataset (look out for submitted paper).
- When evaluated in 24h prior to $\geqslant$M flare, flare-CME model does better than CME.
  - Intuitively makes sense, but both are looking for CMEs
  - Hint that models are looking for different signatures
  - **Next: Can we use explainability methods to see where those two models differ?**

Model performance

Some takeaways and hypotheses

- Models that forecast CMEs struggle more: perhaps not as many signatures in photospheric magnetic field?
- TSS inflates the performance of the models without knowledge of event distribution in dataset (look out for submitted paper).
- When evaluated in 24h prior to $\geqslant$M flare, flare-CME model does better than CME.
  - Intuitively makes sense, but both are looking for CMEs
  - Hint that models are looking for different signatures
  - **Next: Can we use explainability methods to see where those two models differ?**

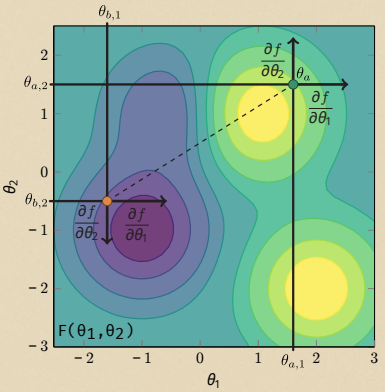
Work in progress: Explainability

Integrated Gradients (IGs) (Sundararajan et al. (2017))

- Attribution method: How much did each of the inputs contribute to the prediction?

$$\mathrm{IG}_i(\theta_{a,i}) = (\theta_{a,i} - \theta_{b,i}) \int_0^1 \frac{\partial F\big(\theta_a + \alpha\,(\theta_b - \theta_a)\big)}{\partial \theta_i}\, d\alpha$$

- Has nice properties like:
  - Completeness
  - Independent of model implementation
  - Sensitivity



Toy model in 2D $F(\theta_1, \theta_2)$

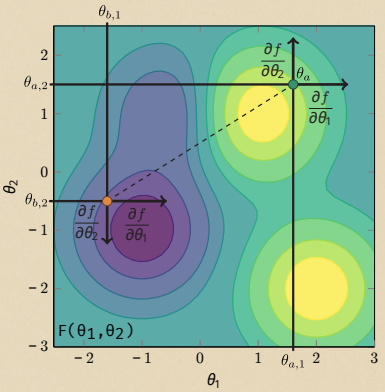
Work in progress: Explainability

Integrated Gradients (IGs) (Sundararajan et al. (2017))

- Attribution method: How much did each of the inputs contribute to the prediction?

$$\mathrm{IG}_i(\theta_{a,i}) = (\theta_{a,i} - \theta_{b,i}) \int_0^1 \frac{\partial F\big(\theta_a + \alpha\,(\theta_b - \theta_a)\big)}{\partial \theta_i}\, d\alpha$$

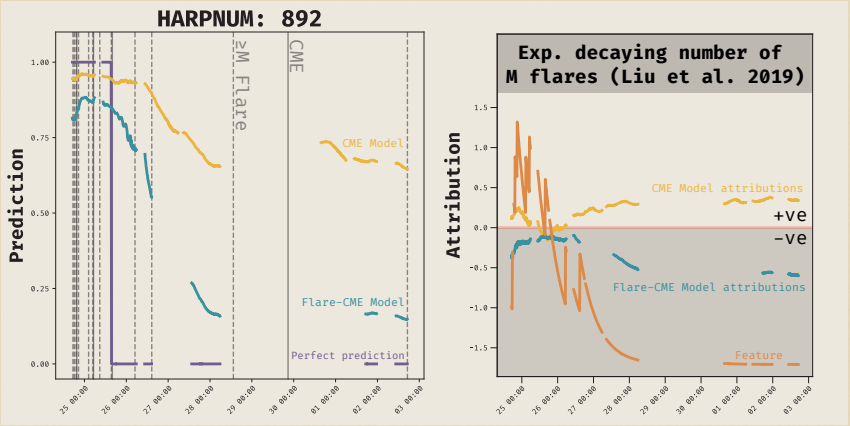
- Has nice properties like:
  - Completeness
  - Independent of model implementation
  - Sensitivity



Toy model in 2D $F(\theta_1, \theta_2)$

Work in progress: Explainability

IGs in practice



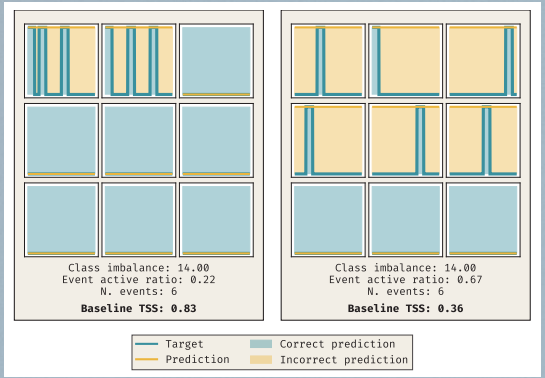
Work in progress: Explainability

Current challenges

- How to choose a meaningful baseline?
- Is a line integral valid from a physical perspective?
- How to test any explanation against what the model really does?

Backup slides

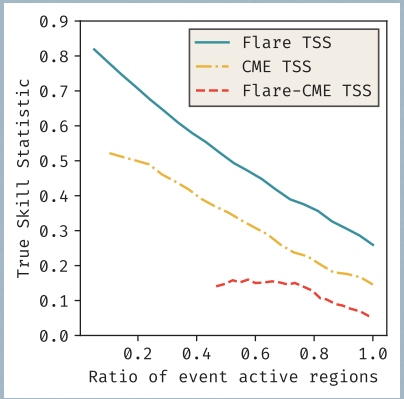
Event active regions



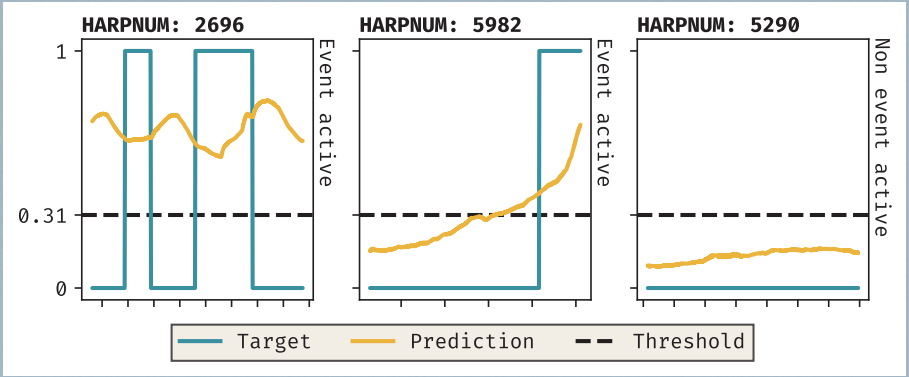
Dataset	Ratio
Flare	0.05
CME	0.10
Flare-CME	0.47

Backup slides

Active regions TSS

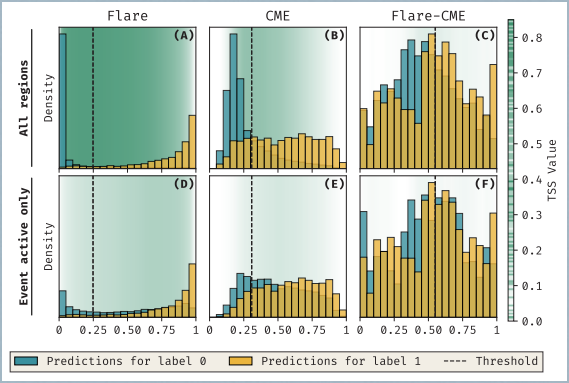


CME predictions examples



Backup slides

Distribution of predictions



Backup slides

Dataset splits