Multiline flare prediction based on the IRIS spectral lines Mg II h&k, Si IV, and C II with machine learning

Jonas Zbinden,

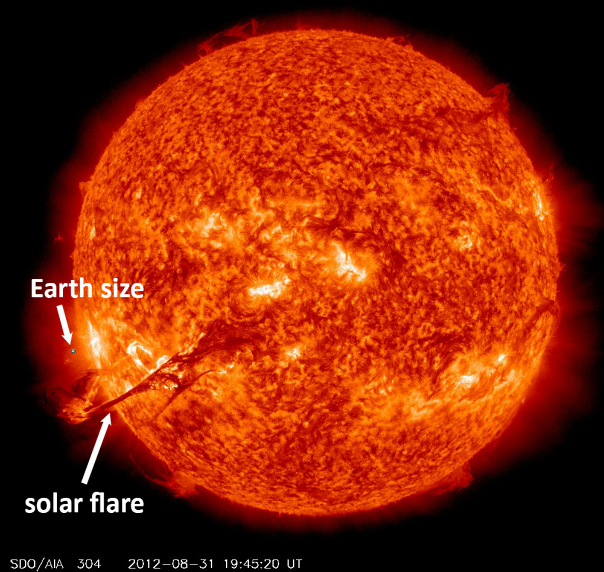
Astronomical Institute

University of Bern

SCOSTEP - 2023

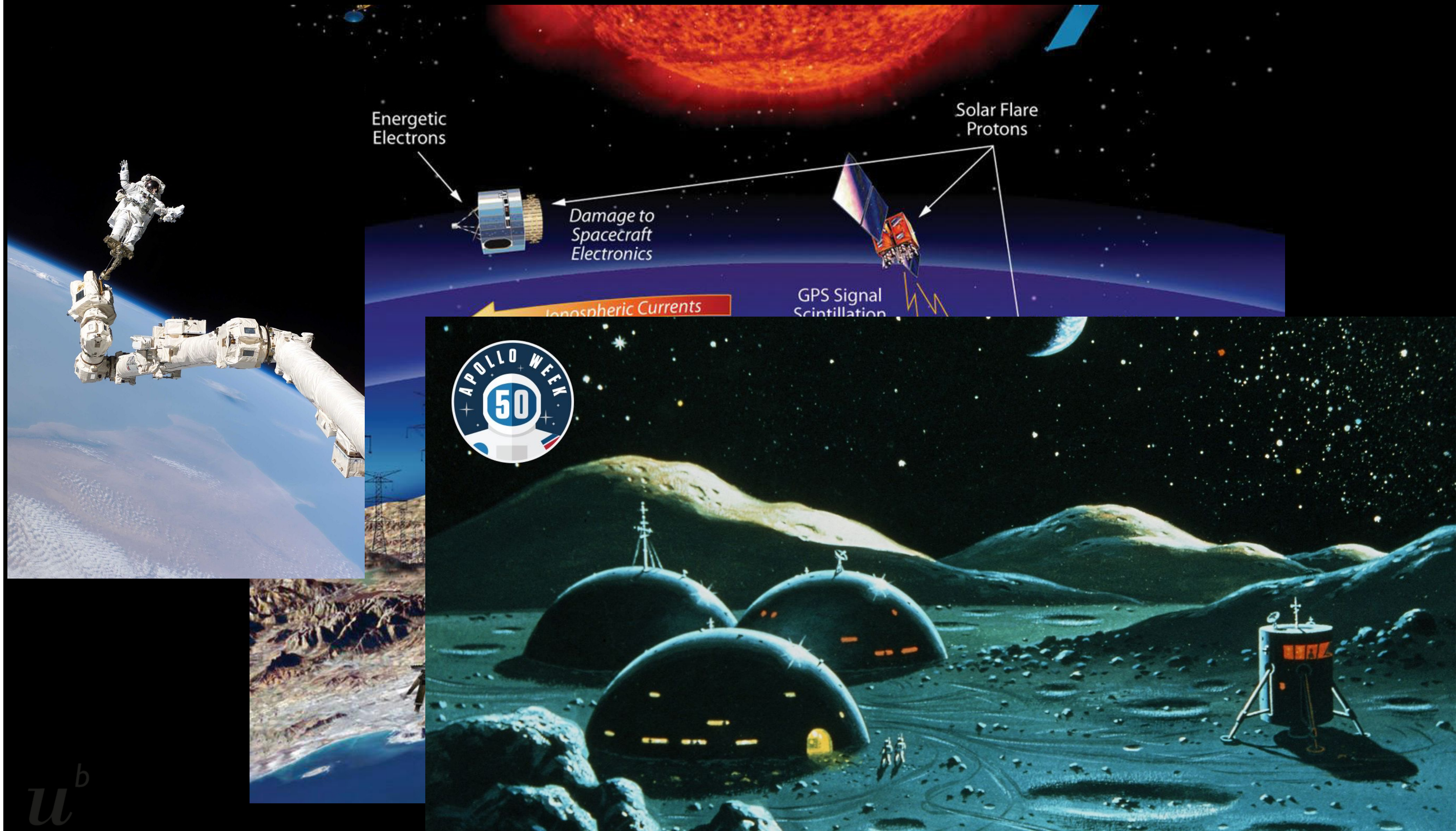
16 May 2023

Image Credit: Lucia Kleint



UNIVERSITY
OF BERN

Solar flare prediction – Why should we care?



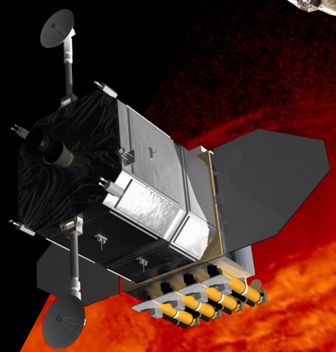
http://sml.unige.ch/
jonas.zbinden@unibe.ch

UNIVERSITY
OF BERN

Probing the solar atmosphere in height

IRIS

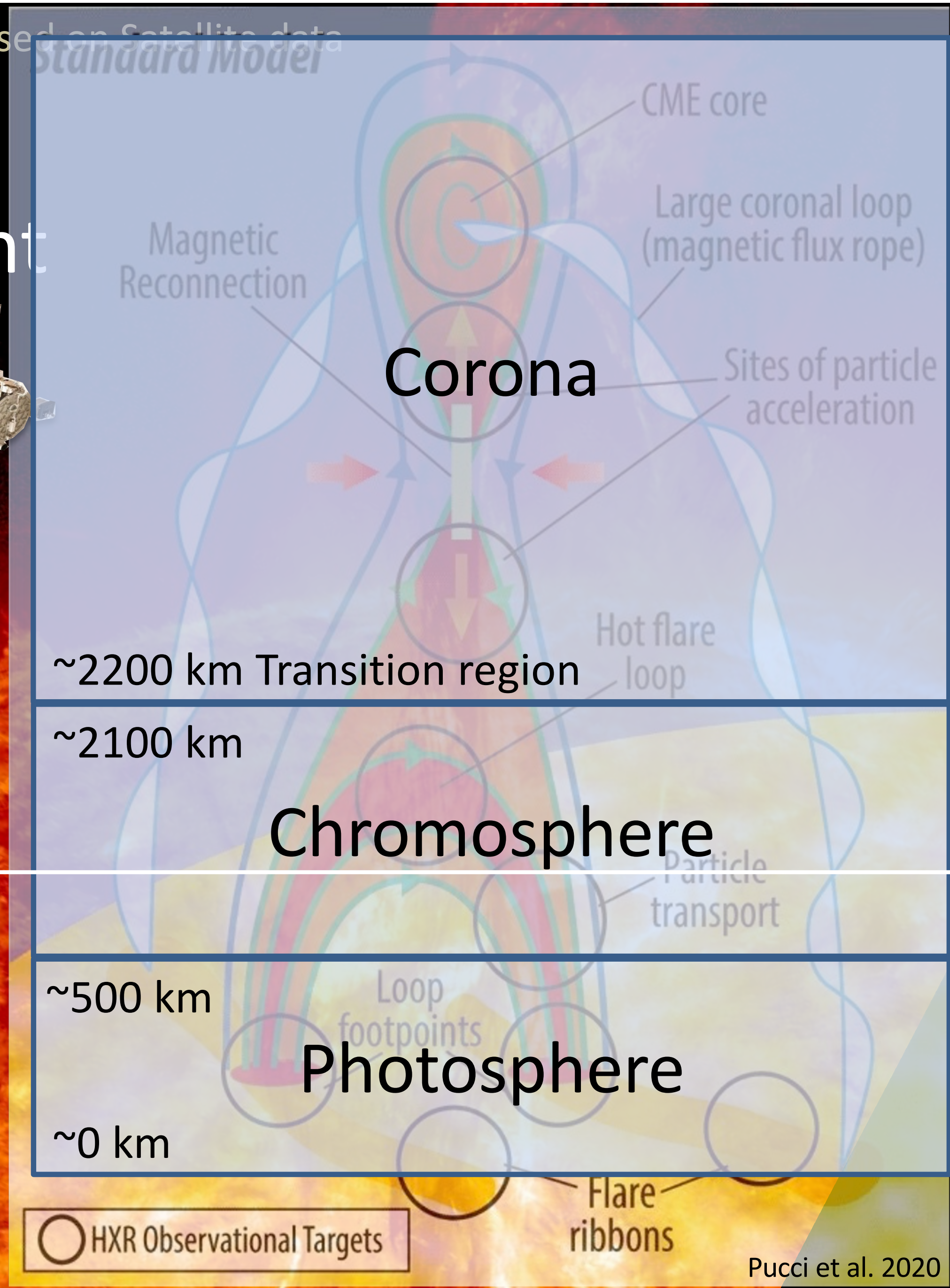
SDO



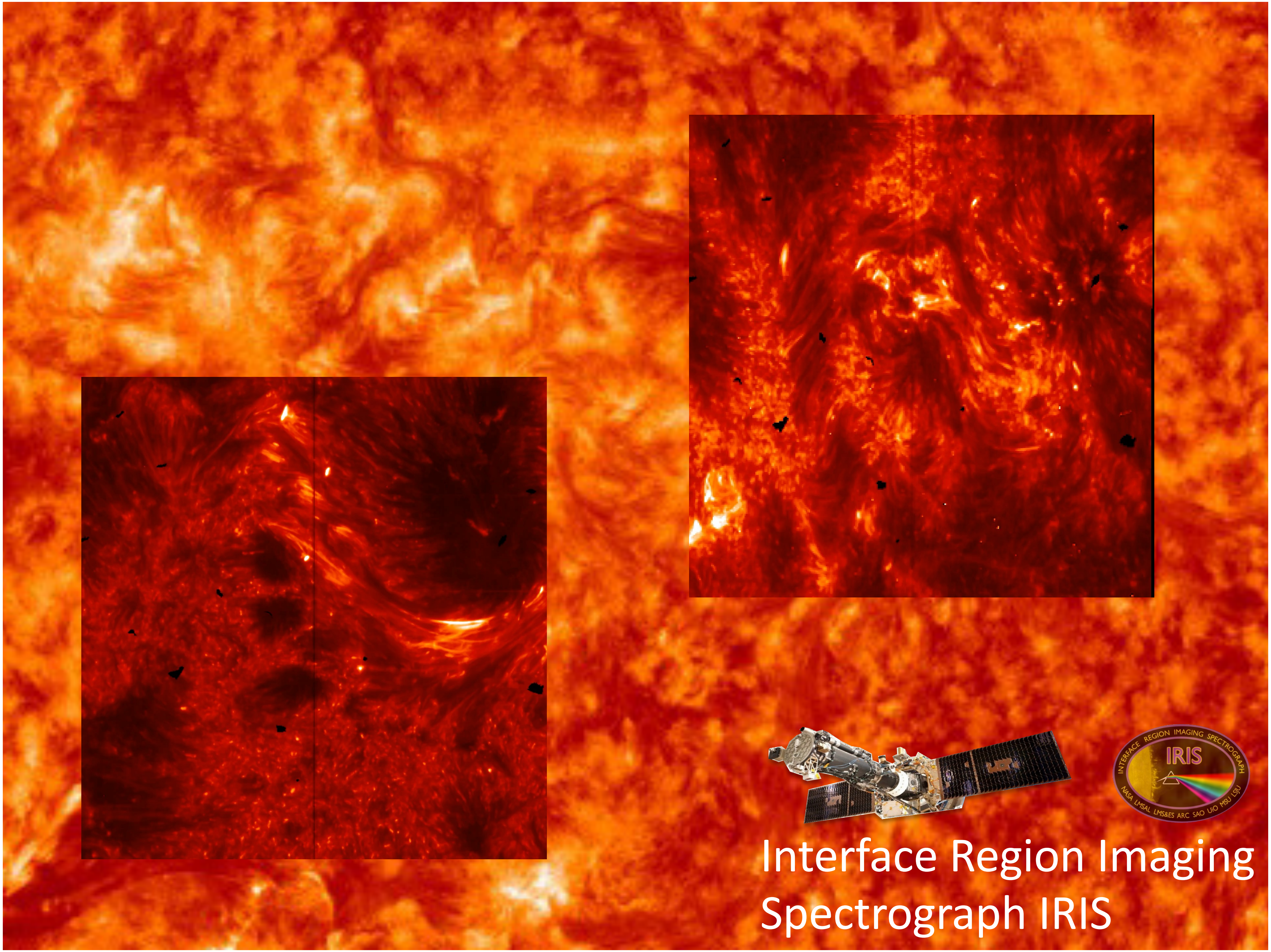
Can only be observed from space

Can be observed from ground

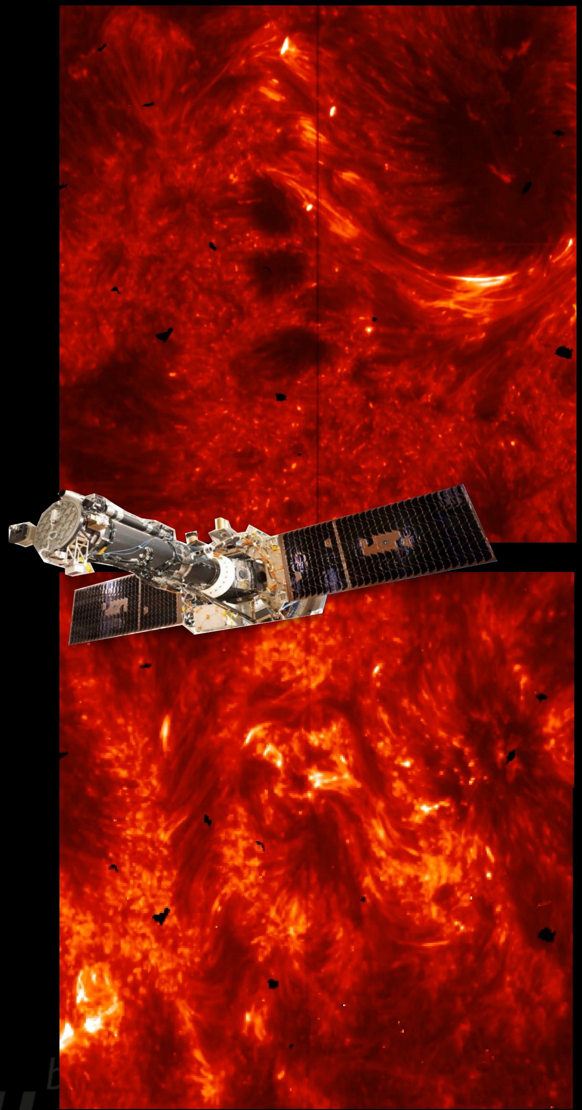
DKIST

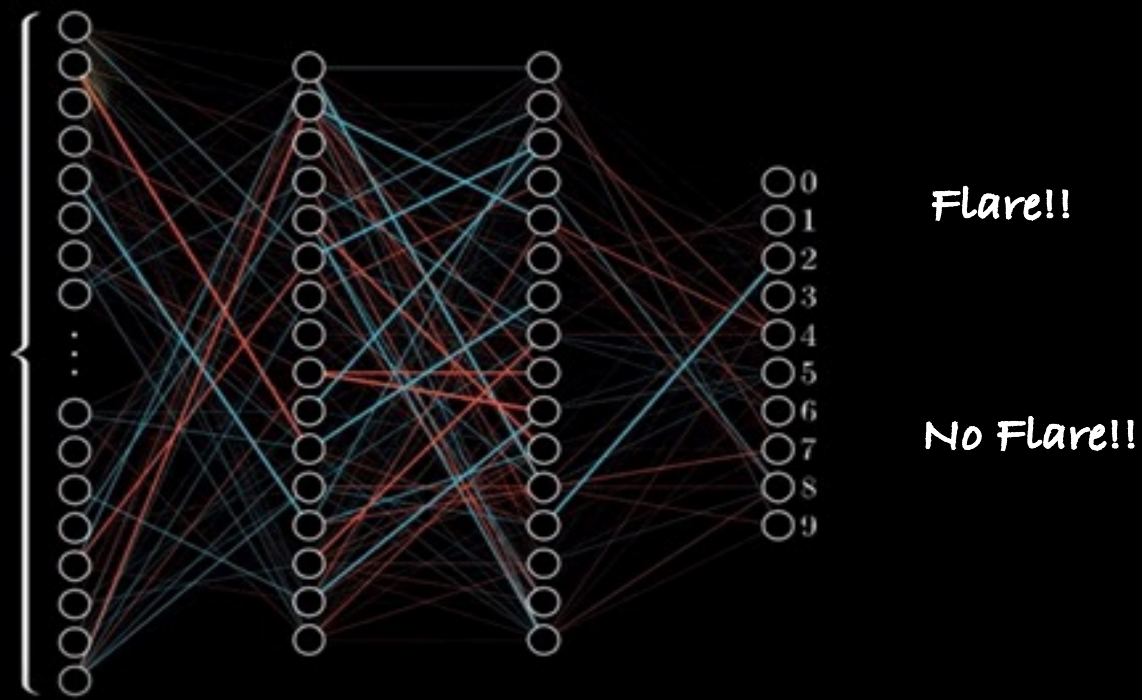


http://sml.unige.ch/
jonas.zbinden@unibe.ch


Interface Region Imaging
Spectrograph IRIS

How well can we predict flares based on IRIS spectral lines?





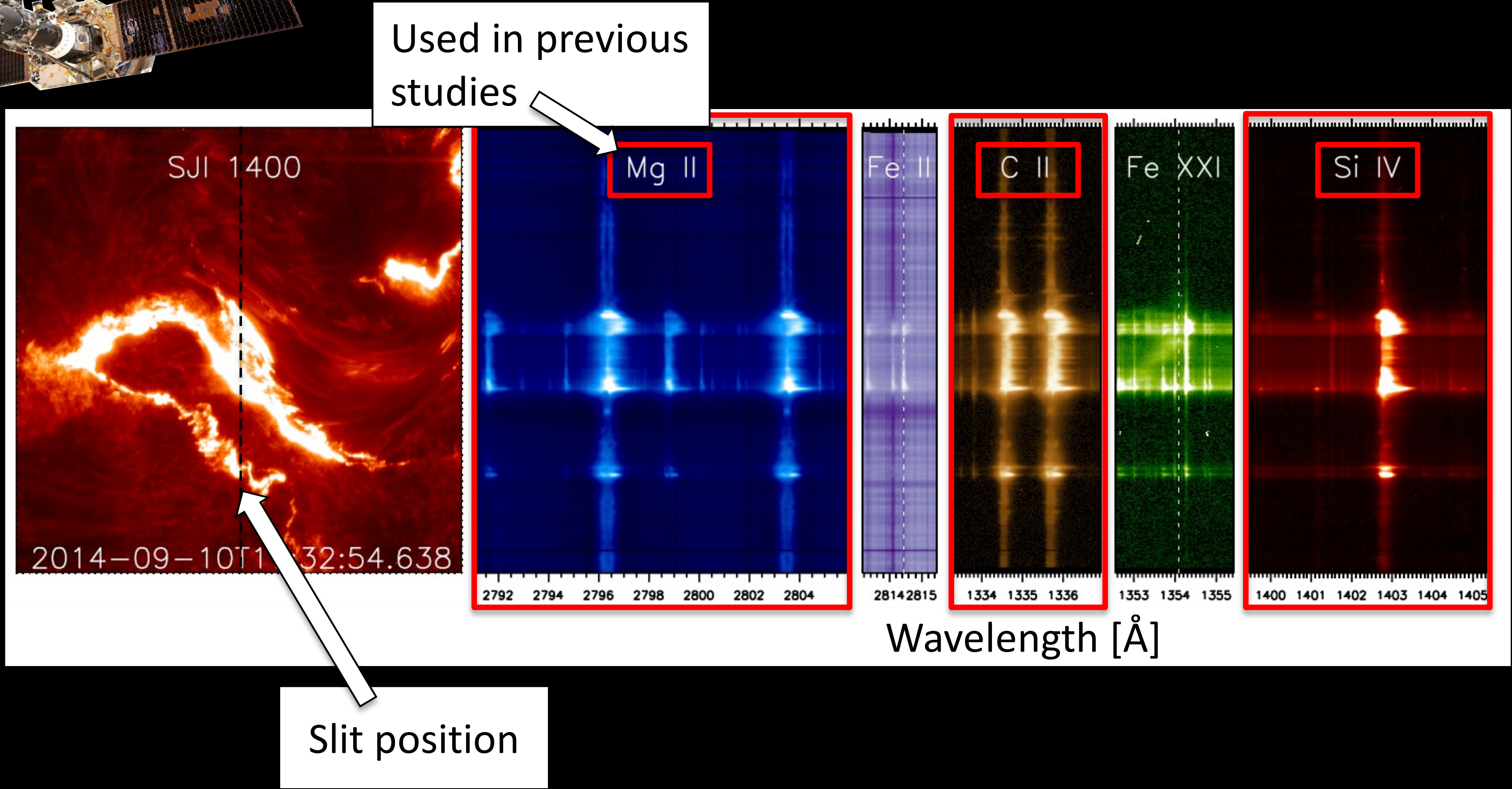
Proof of concept: Panos et al. 2020 showed prediction of flares possible on short timescales with Mg II k.

Explainable AI: Panos et al. 2023 show which part of spectra are most important for flare prediction.

UNIVERSITY OF BERN

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Spectra collected with IRIS:



http://sml.unige.ch/
jonas.zbinden@unibe.ch

Dataset

- Observations from previous studies for proof of concept

PF: 19 obs, 32 flares, ~25 minutes - 1 hour before flare
AR: 18 obs, ~40+ hours

- Full set of observations:

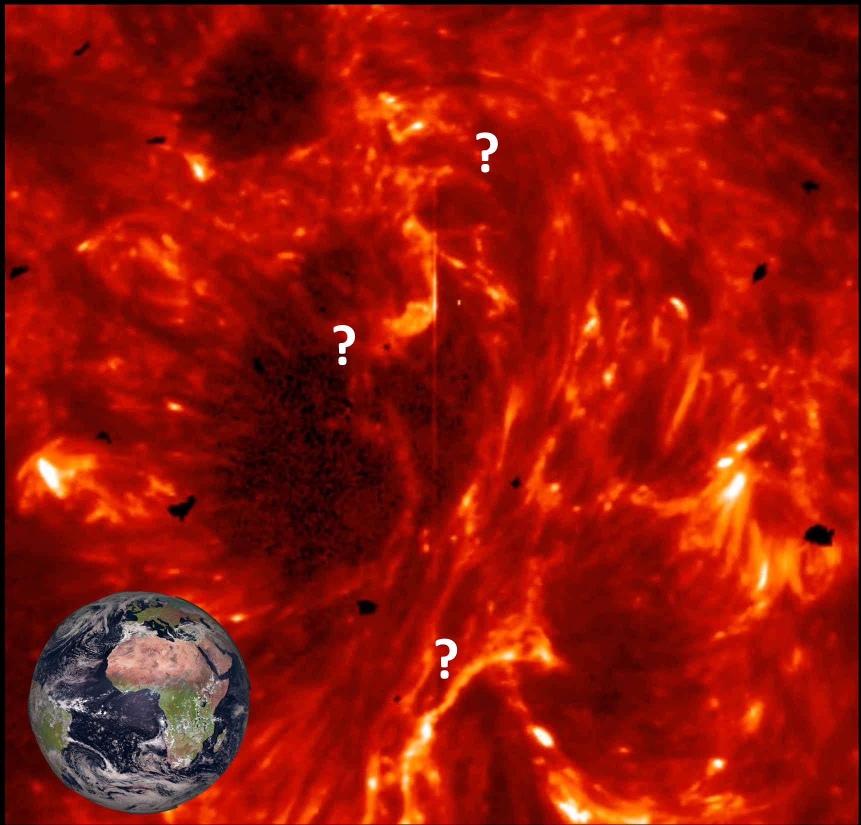
PF: 50 obs, 73 flares, ~25 minutes - 1 hour before flare
AR: 30 obs, 50+ hours
5 to 10 Million spectra in each line

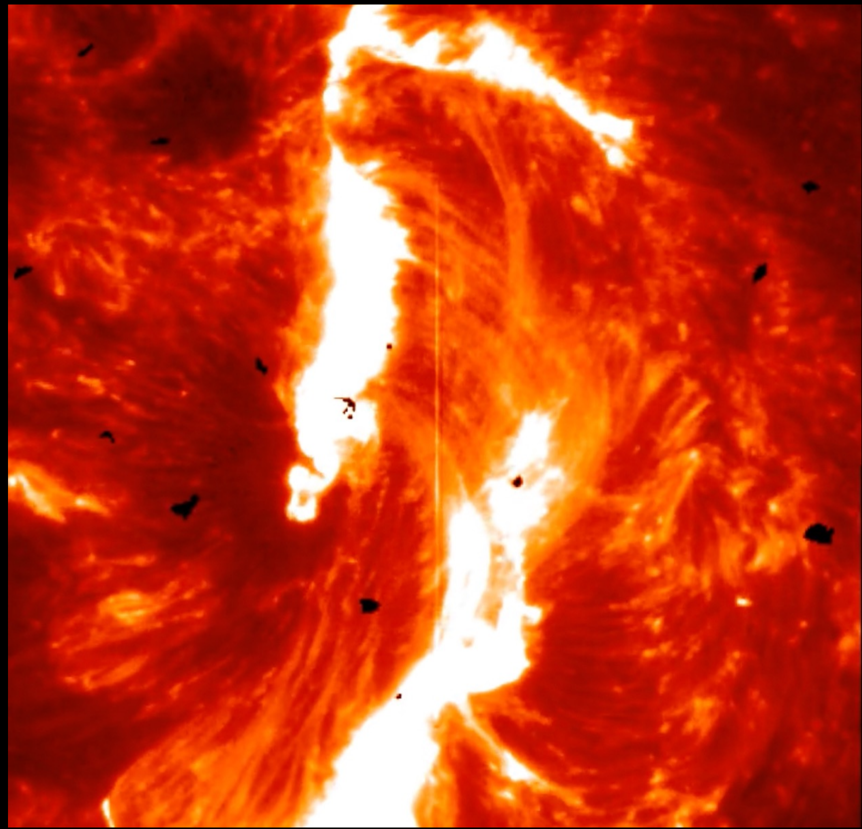
- **Dataset split into Groups 1 and 2**

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Masking

Where will the flare happen?

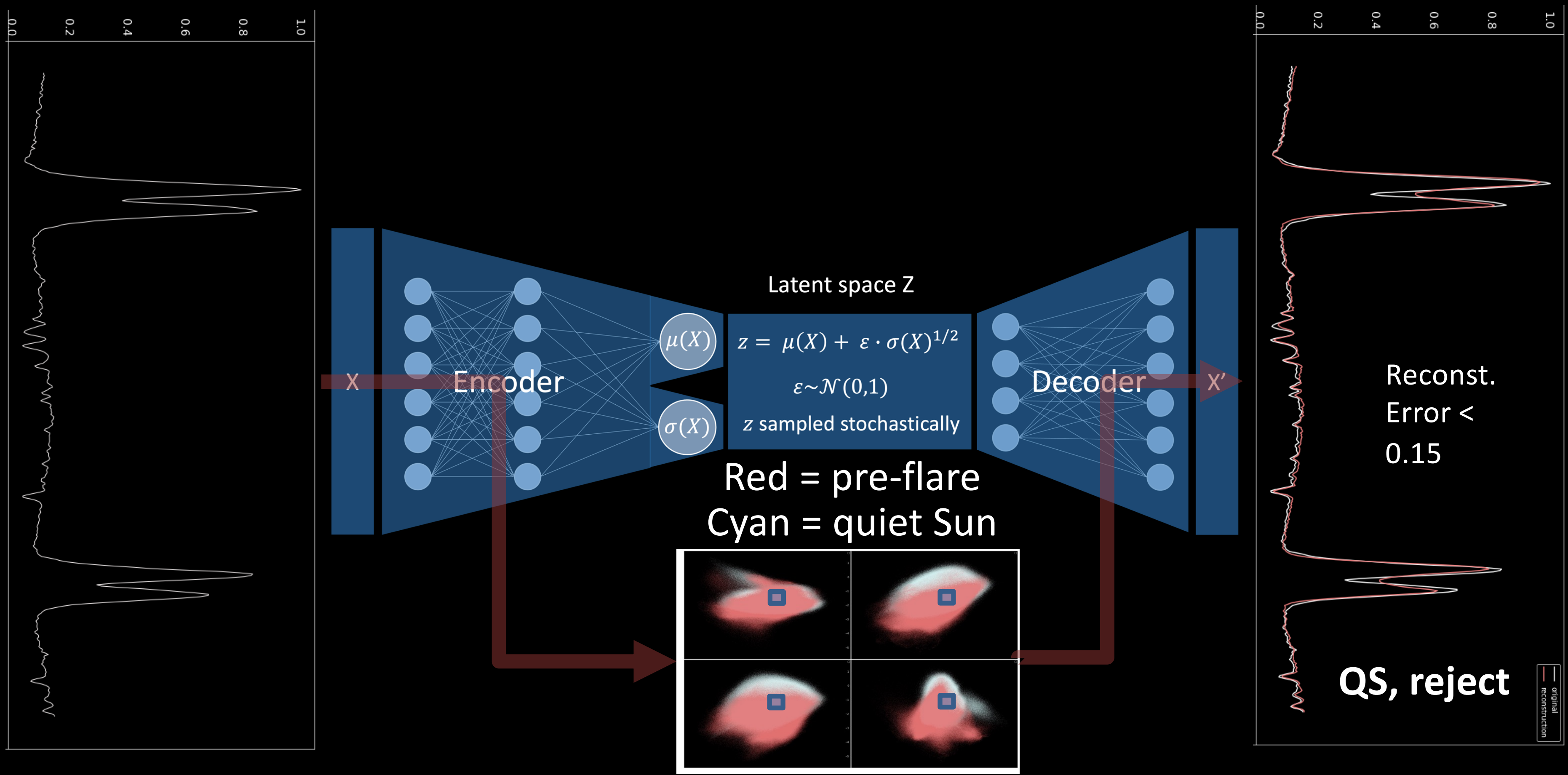




UNIVERSITY OF BERN

http://sml.unige.ch/
jonas.zbinden@unibe.ch

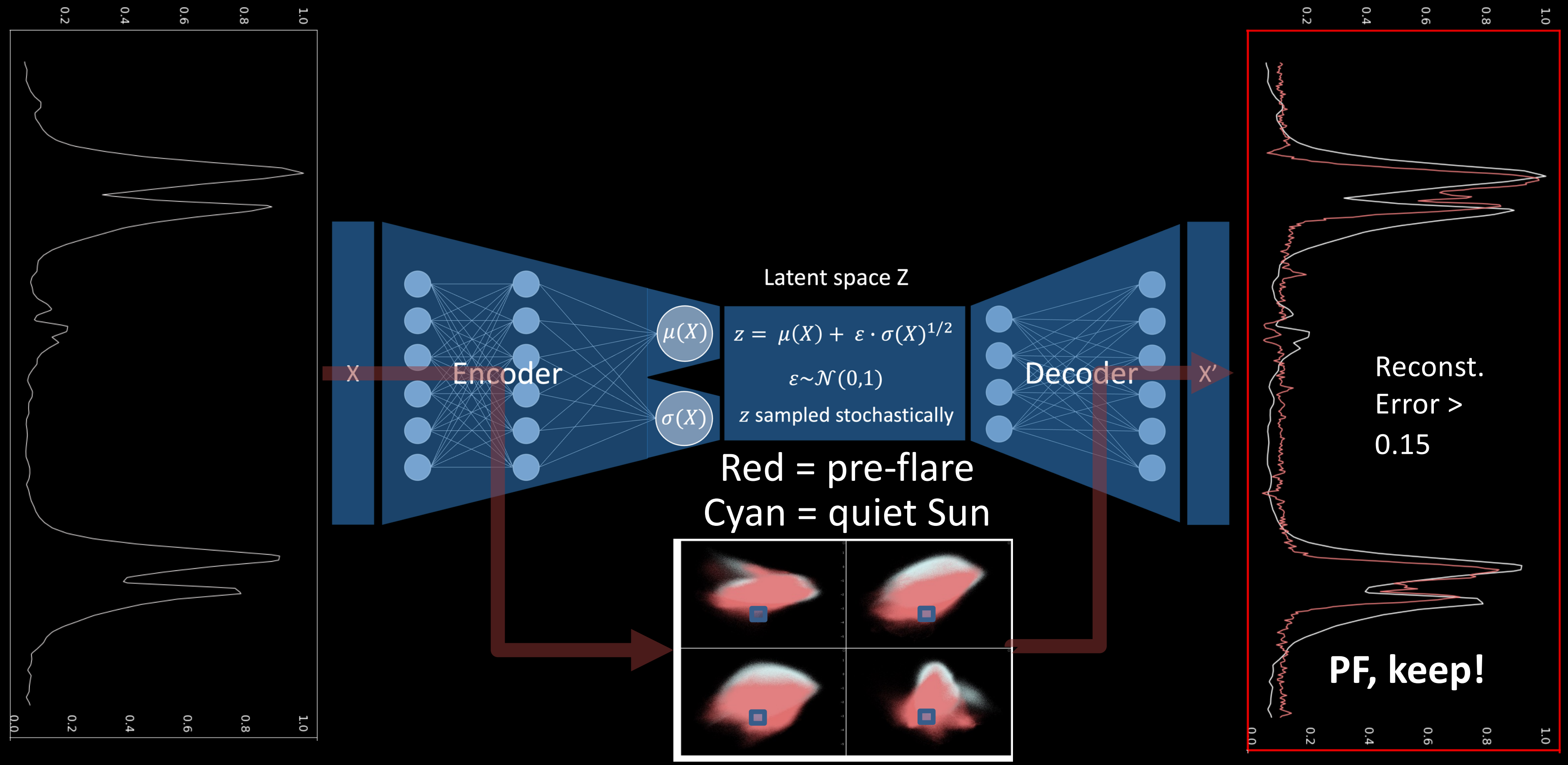
Variational Autoencoder
X = Spectrum, z = vector in 4-dimensional latent space



UNIVERSITY
OF BERN

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Variational Autoencoder
X = Spectrum, z = vector in 4-dimensional latent space

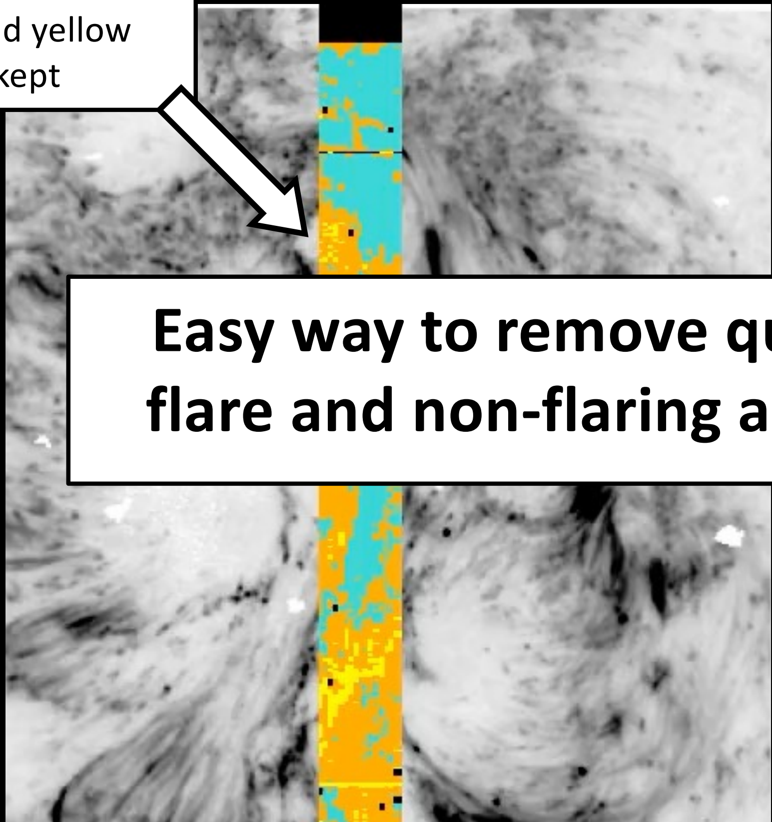


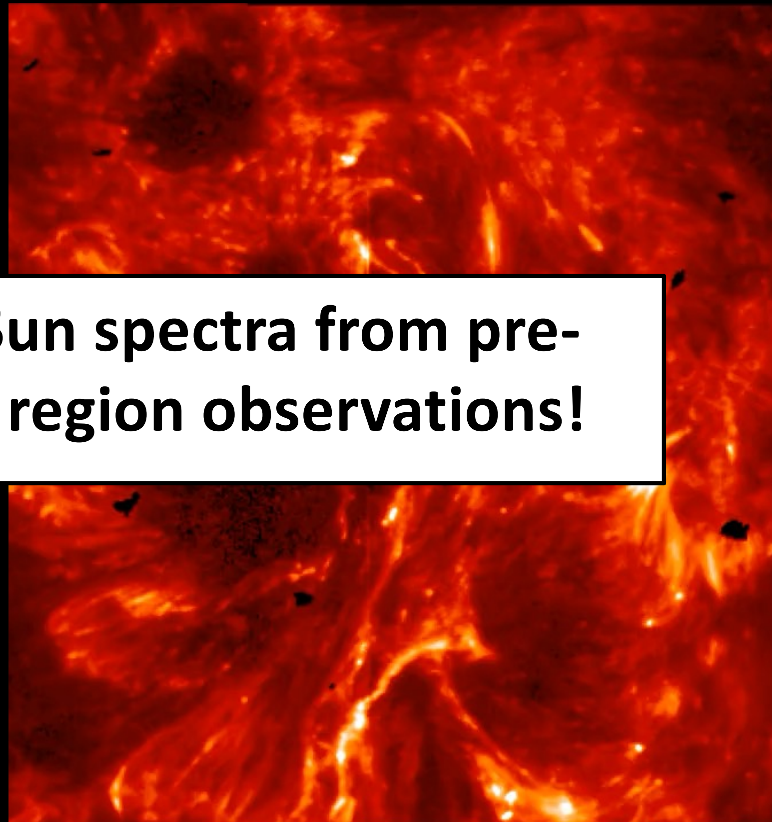
UNIVERSITY OF BERN
http://sml.unige.ch/
jonas.zbinden@unibe.ch

Masking

Filtering data with Variational Autoencoder (VAE) trained on quiet sun spectra

Orange and yellow areas are kept





Easy way to remove quiet Sun spectra from pre-flare and non-flaring active region observations!

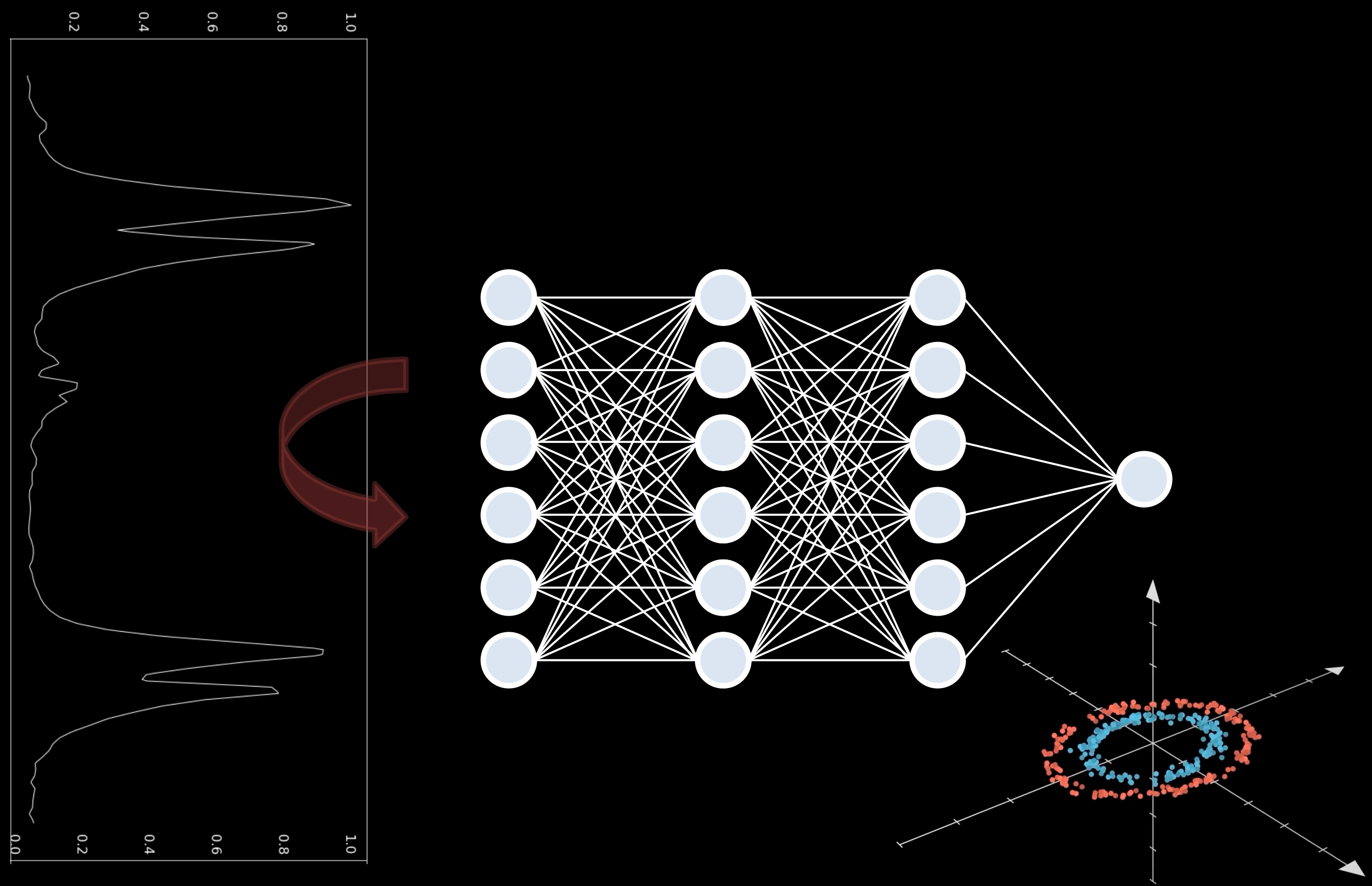
Blue: Quiet Sun
Orange: less quiet sun
Yellow: Definitely not quiet sun

UNIVERSITY OF BERN

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Jonas Zbinden

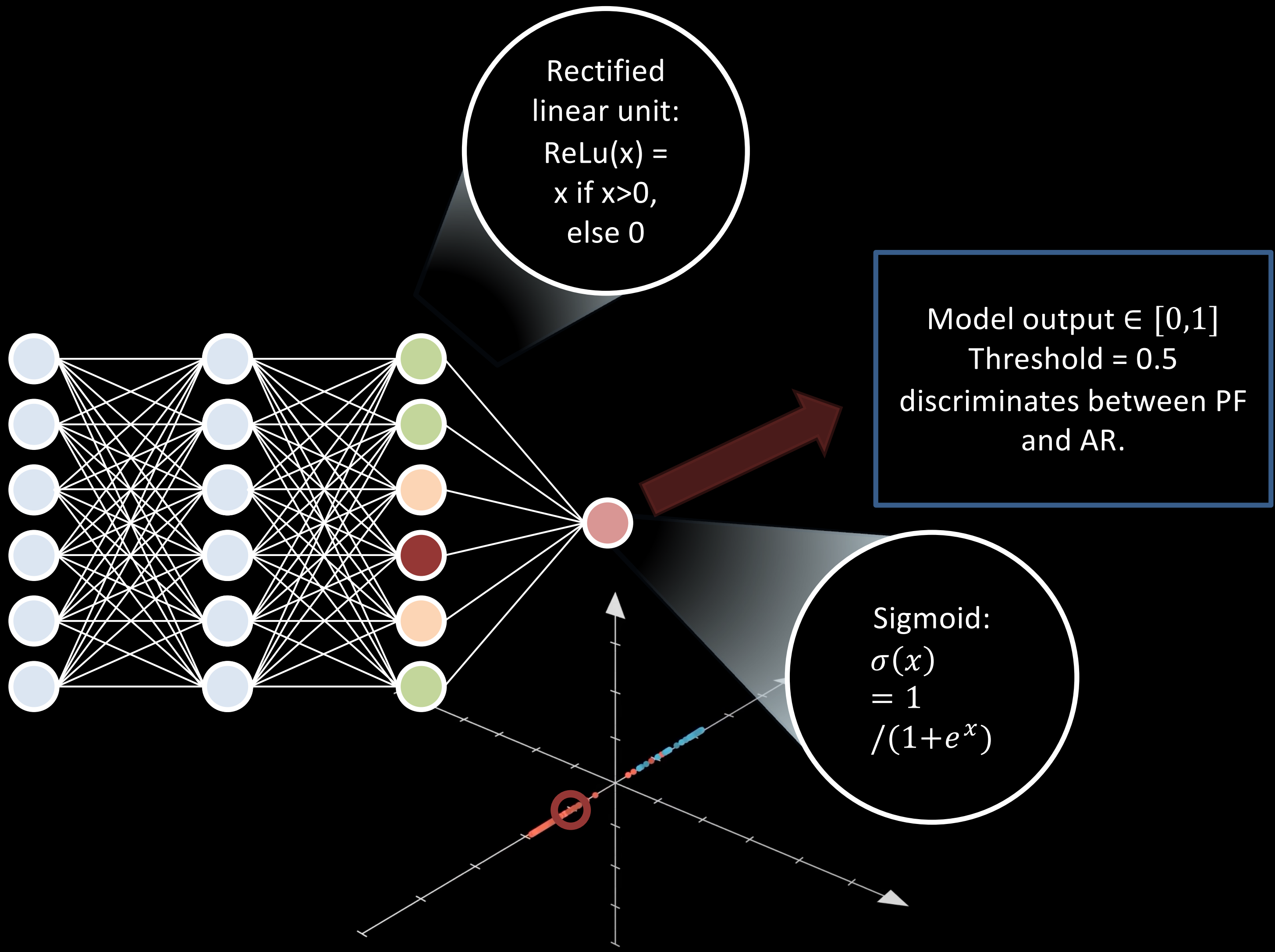
Neural Network (deep, fully connected)



Sam Sartor: youtube.com/watch?v=CfAL_cL3SGQ

UNIVERSITY OF BERN

http://sml.unige.ch/
jonas.zbinden@unibe.ch



UNIVERSITY OF BERN
http://sml.unige.ch/
jonas.zbinden@unibe.ch

Prediction – Testing models

Accuracy: Problems with datasets with class imbalance

$$\text{ACC} = \frac{TP+TN}{P+N} = \in [0,1]$$

Class imbalance invariant score True Skill Statistics TSS:

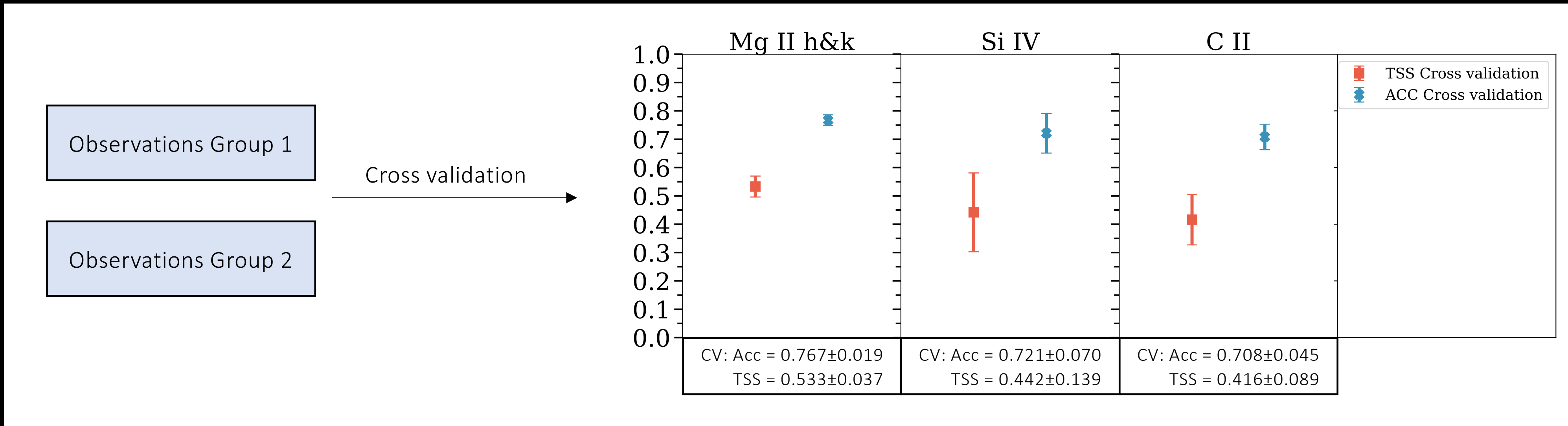
$$\text{TSS} = \frac{TP}{P} + \frac{TN}{N} - 1 = \in [-1,1]$$

Score between -1 and 1

- 1: perfect predictions
- 0: random guessing
- -1: model is confused (opposite outputs to labels)

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Results Single spectral lines experiment:



- Best score Panos et al. 2020: **TSS ~ 0.6** on Observation group 1

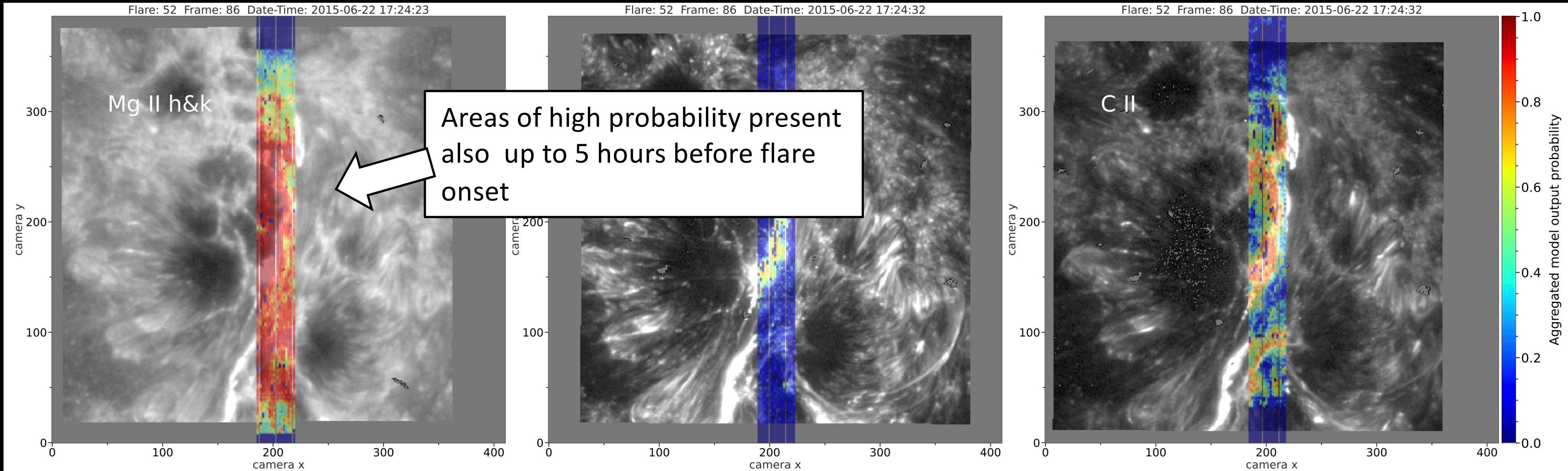
  ⟶ We reached **TSS = 0.712** on Observation group 1

  ⟶ **Mg II h&k** line highest score in all experiments

UNIVERSITY OF BERN

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Results Single spectral lines experiment:



UNIVERSITY OF BERN

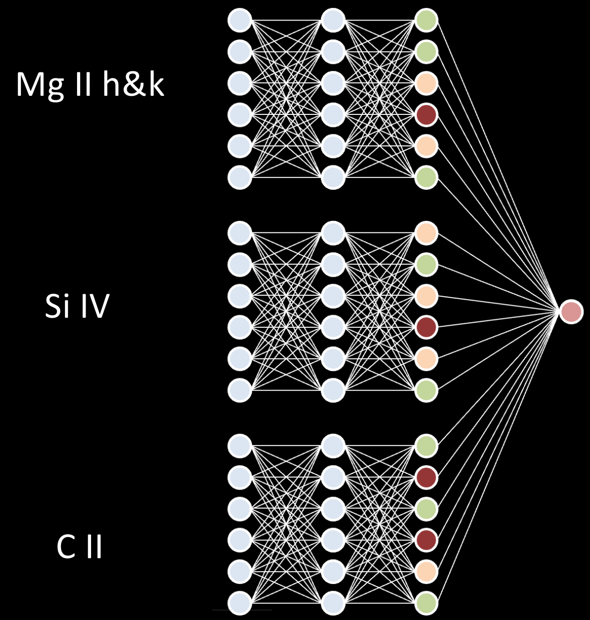
http://sml.unige.ch/
jonas.zbinden@unibe.ch

Combined spectral lines experiment

Dataset of ONLY shared pixels

Experiments: **Full set of observations**
- Mg II h&k on this data set
- Combination of spectral lines

Conceptual model architecture



- Best score **TSS = 0. 534** (Mg II h&k TSS=0.463 shared pixels)
- Shared pixels introduce selection bias
- Information gain is **minimal** (Best score Mg II h&k TSS=0.533)

UNIVERSITY OF BERN
http://sml.unige.ch/
jonas.zbinden@unibe.ch

Summary of Results

- **Best** model based on **Mg II h&k (TSS = 0.533) for 1 hour before flare onset**
  - Investigating the spectra indicative of an occurring flare can be associated with **heating in the mid- to upper chromosphere**
- **C II** and **Si IV** have some predictive information but less than **Mg II h&k (models generalize badly)**
- Each observation has **unique properties** that can affect the training and testing of the models
- **Combining** spectral lines marginally improves scores
- VAE (or other sophisticated **masking methods**) can lift some of the mixing of PF, AR and QS spectra

UNIVERSITY OF BERN

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Conclusions

- Extension of previous studies to an operational setup
- First time tested the potential of the spectral lines **Si IV** and **C II** for flare prediction and **combining** all three spectral lines.
- **No correlations** found between model outputs and observational properties! (Intensity, GOES X-ray flux, flare magnitude,...)
  - ➢ The models learn information **exclusively** on the **shape** of the **spectra**
  - ➢ The model outputs are **scale invariant**

Do the probability outputs change over time (increasing closer to flare onset)?

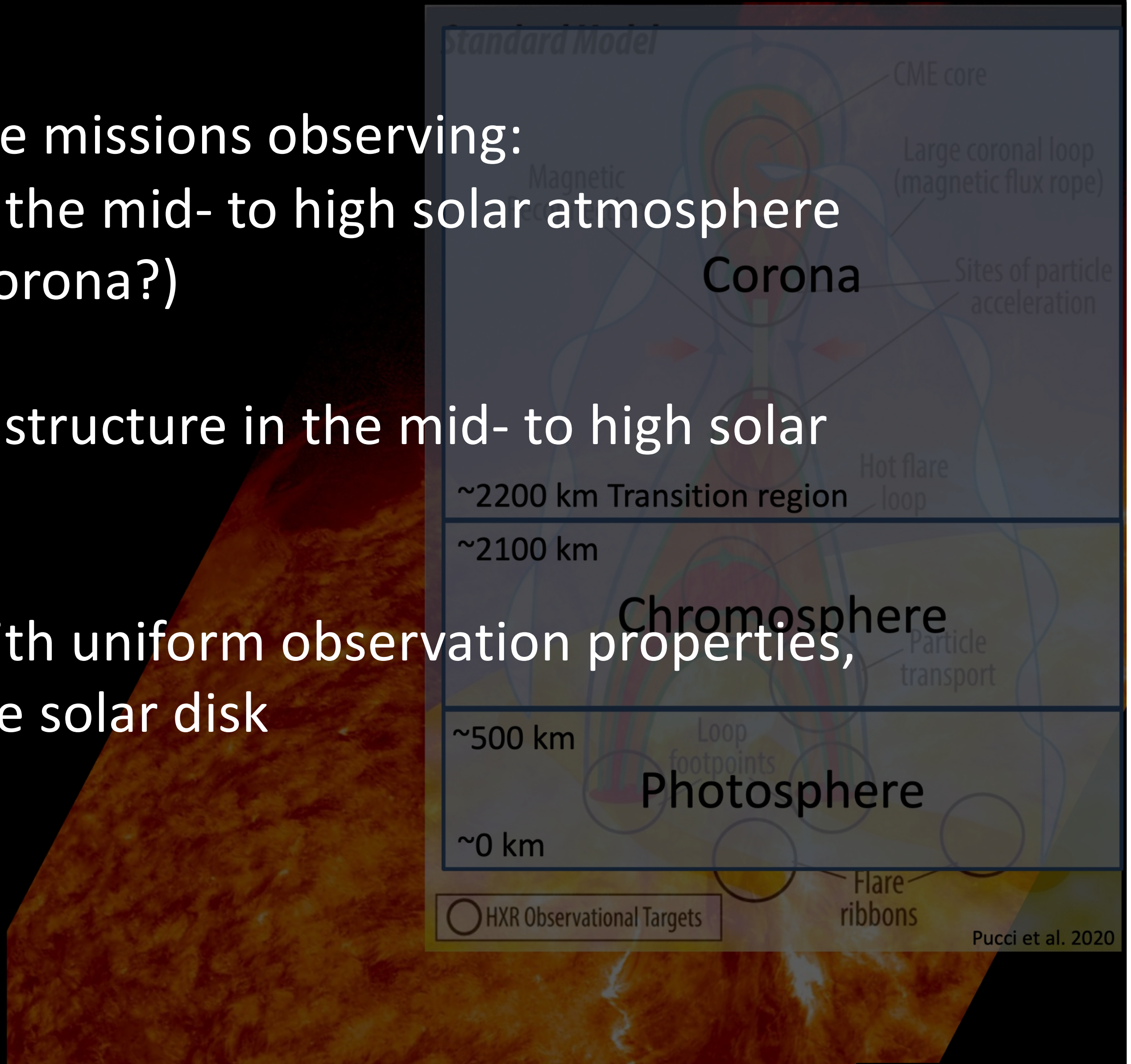
What is the minimum resolution and field of view to capture the differences between the pre-flaring and non-flaring active regions?

http://sml.unige.ch/
jonas.zbinden@unibe.ch

Conclusions for future space missions

We need future space missions observing:

- spectra forming in the mid- to high solar atmosphere (Chromosphere, Corona?)
- the magnetic field structure in the mid- to high solar atmosphere
- long time series with uniform observation properties, ideally of the entire solar disk



Pucci et al. 2020

http://sml.unige.ch/
jonas.zbinden@unibe.ch