Machine Learning Tools for Predicting Solar Energetic Particle Hazards: Machine Learning Part
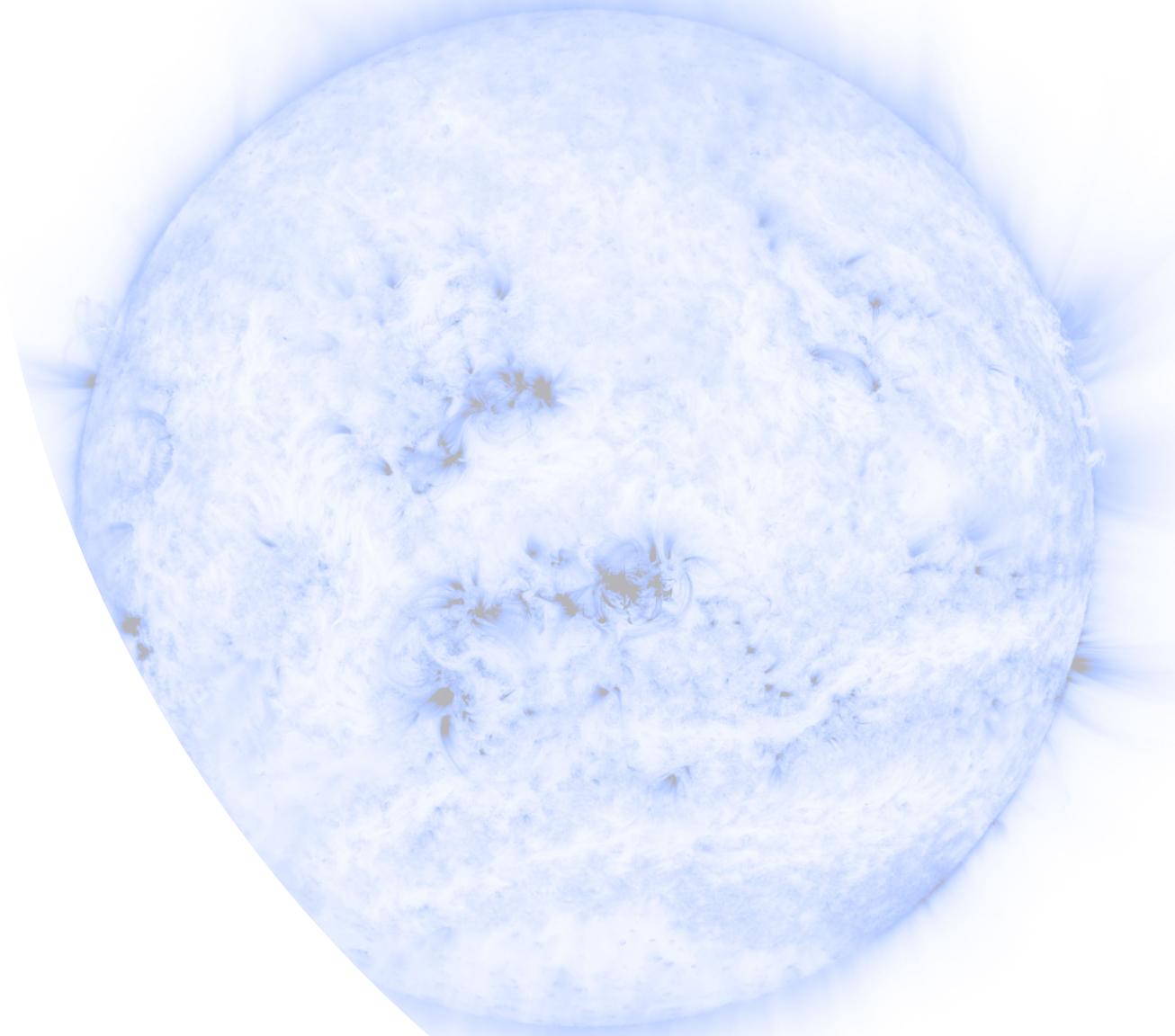
Viacheslav Sadykov
and ESI Team

NASA Ames Research Center

Bay Area Environmental Research Institute

ESI project continuation review   November 20, 2020

Outline

- Introduction: prediction of Solar Proton Events (SPEs)

- Progress Report: development of "all-clear" forecasts of SPEs

- Conclusions

Solar Energetic Particles (SEPs) and Solar Proton Events (SPEs)

➢ Solar Energetic Particle (SEP) events can be defined as significant enhancements of the particle flux coming from the Sun with respect to the stable background

➢ Solar Proton Events (SPEs) represent a major subclass of SEPs

➢ The terms "SEP event" and "SPE" are equivalent for this presentation and represent enhancements of energetic proton fluxes as measured by near-Earth satellites (GOES)
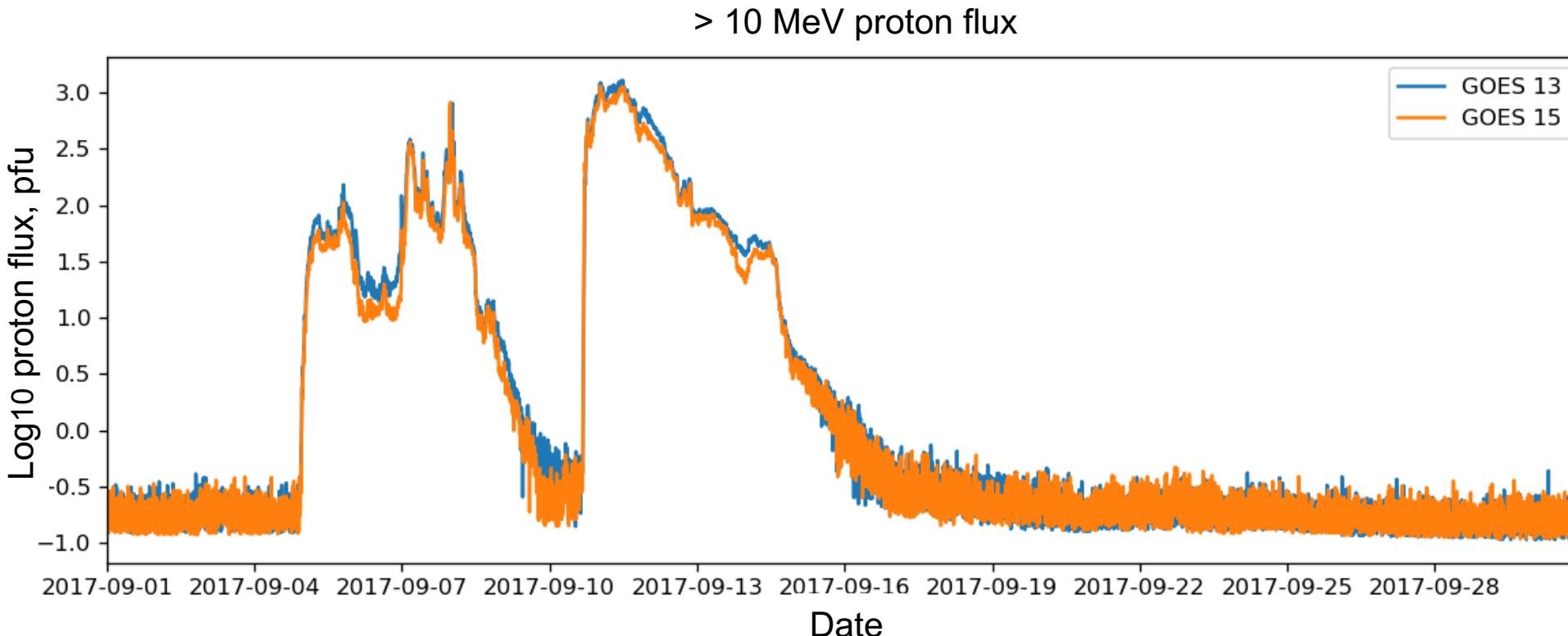
> 10 MeV proton flux

An example of > 10 MeV proton flux measurements by the GOES-13 and GOES-15 satellites

Why is predicting solar proton events challenging?

➢ Severe class-imbalance ratio. The ratio of SPE-active to SPE-quiet days is:

    ➢ **1/34** for > 10 MeV > 10 pfu events

    ➢ **1/155** for > 100 MeV > 1 pfu events

➢ SPE onset may occur significantly later than the initiating flare.

➢ The locations of SPE initiations on the Sun are not known precisely. Some events are initiated on the far side of the solar disk.

Statistics of SPE days (June 2010 - December 2019)

■ SEP-quiet days

■ Days with > 10 MeV > 10 pfu flux

■ Days with > 100 MeV > 1 pfu flux
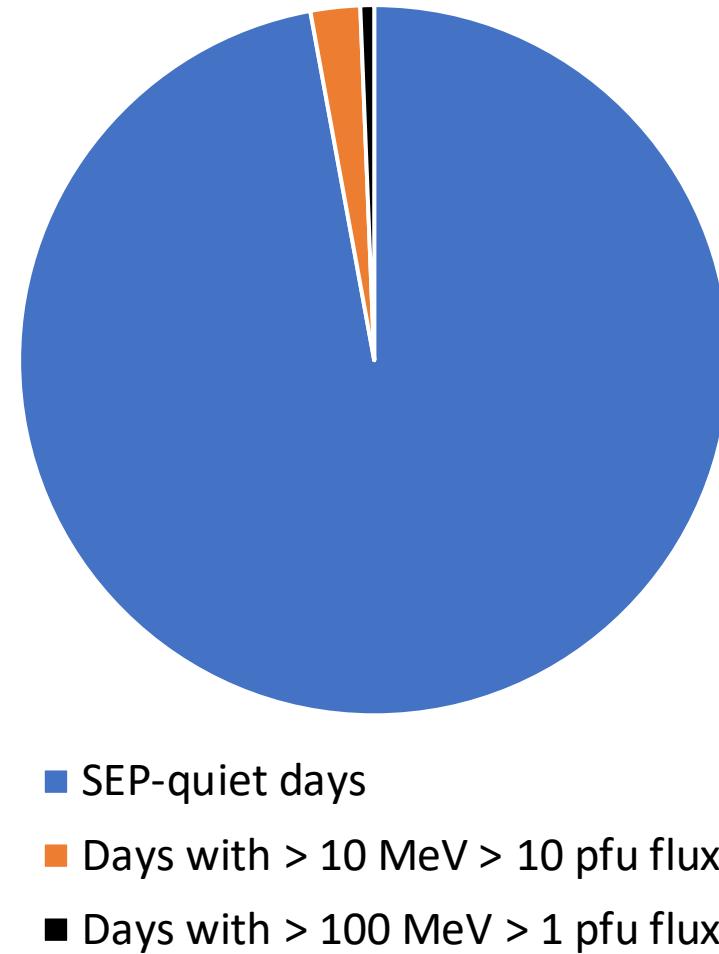
Year 1 Milestones of the Proposed Research

Two milestones were planned to be accomplished for Year 1:

➢ Complete the Task 1 and create a fully functional online-accessible database of SPE-related data, metadata, and data products. The database will be available online for broader research community from the NJIT web servers and NASA Helioportal.

➢ Perform initial development of the parameter-dependent "all-clear" SPE forecasts for different timescales, particle flux and energy thresholds, and subject to availability of different data sources. The priority will be given to development of daily forecasts of SPEs, and comparison with the operational forecasts from SWPC NOAA.

This presentation reports the accomplishment of

the 2nd Milestone of Year 1.
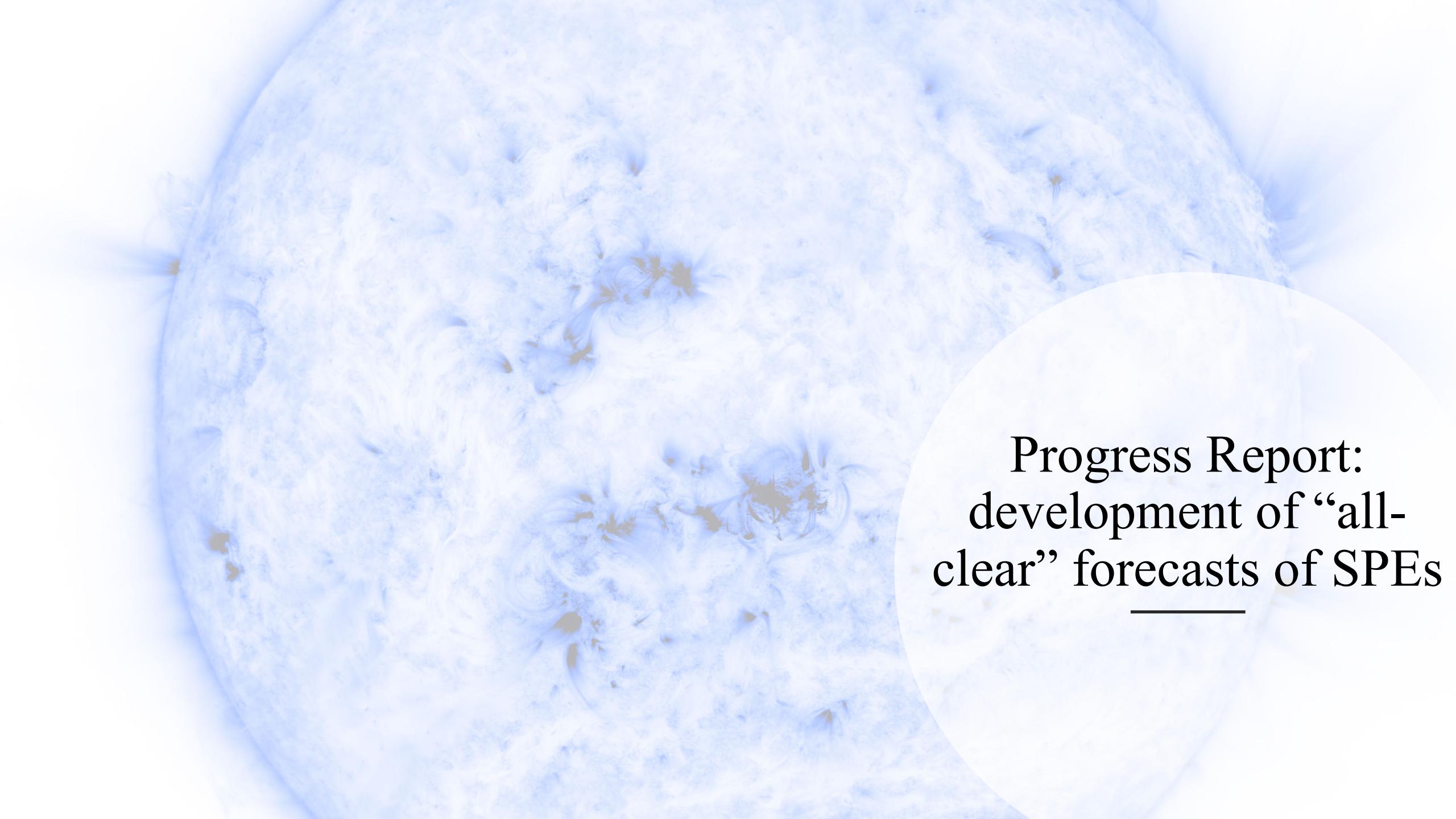
Progress Report: development of "all-clear" forecasts of SPEs

What does it mean to predict an SPE event?

➢ In the framework of this study, to predict an SPE event means, for example:

- To predict at 12 AM UT whether the measured peak flux of > 10 MeV protons will exceed 10 particle flux units during the next day. Timeframe: June 2010 – December 2019.

➢ Defined in this way, the predictions can be compared directly with the SWPC NOAA operational daily forecasts.
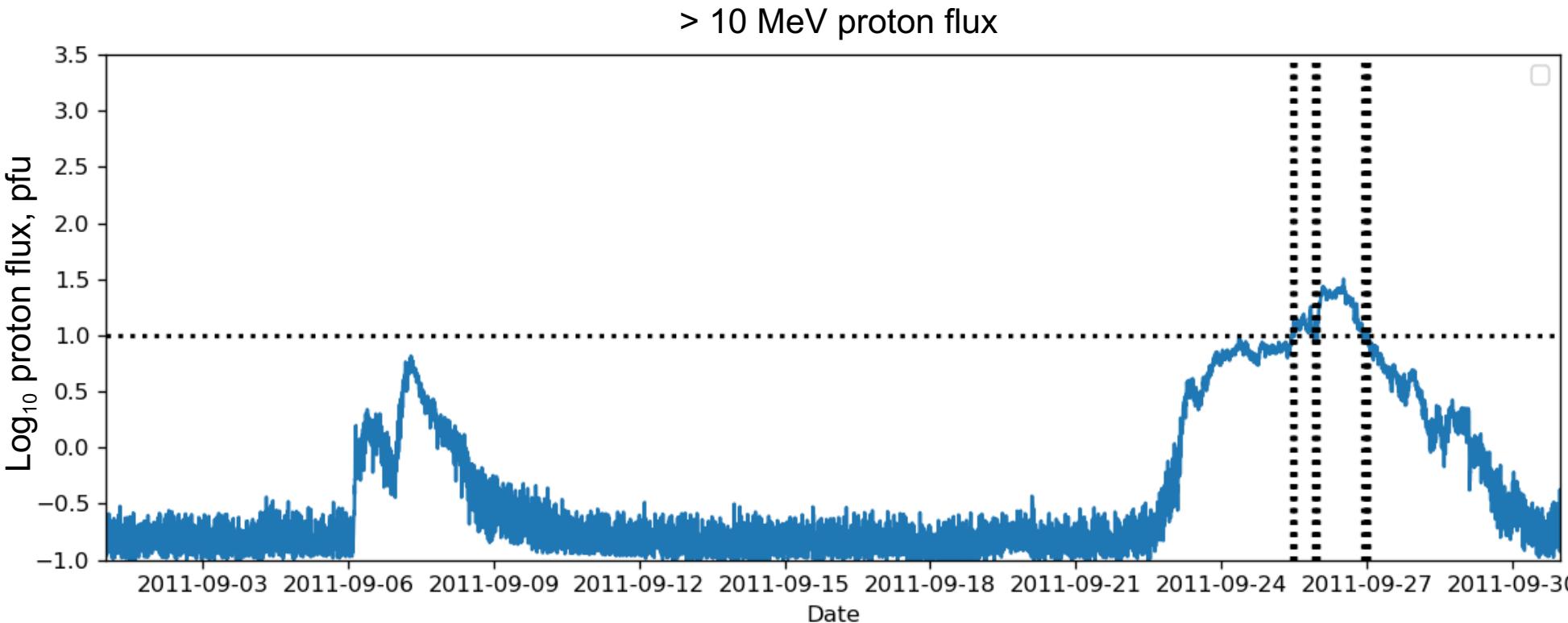
> 10 MeV proton flux

An example of > 10 MeV proton flux measurements by the GOES-15 satellite

Machine learning perspective on the problem

➢ The problem is a classic binary classification problem.

➢ To solve this problem, we employ neural networks and minimize the cross-entropy loss function during training. Given the true label y (1 or 0) and the predicted probability of the event p:

$$Loss = -(y \cdot log(p) + (1 - y) \cdot log(1 - p))$$

➢ The binary classification results may be represented as a confusion table:

Confusion Matrix	Prediction: SPE event	Prediction: no SPE event
Reality: SPE event	True Positives (TP)	False Negatives (FN)
Reality: no SPE event	False Positives (FP)	True Negatives (TN)

➢ The binary outcomes can be combined to form metrics like True Skill Statistics (TSS)

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}$$

About the SWPC NOAA operational SPE forecasts

➢ Issued at 22:00 PM UT for the next day

➢ Mainly statistics-based (utilize lookup tables and event prehistory)

➢ Data utilized for SPE forecasts: integrated SXR flux, AR locations, presence of type-II and type-IV radio bursts

➢ 1% is the smallest probability level issued.

➢ The calculated probabilities can be corrected by forecasters based on their experience.

➢ The daily forecasts for the whole Sun are available online

➢ Major problem: during 2010-2020, 14 out of 101 SPE days happened when a 1% chance of the event was predicted. It is problematic to build all-clear forecasts based on that data.

Working with AR information

➤ The energy released during transient events is (in most cases) initially stored in non-potential configurations of magnetic fields in active regions (ARs).

➤ SHARP features represent the properties of the vector magnetic field extracted for AR patches (Bobra et al. 2014).

➤ We utilize the last reliable daily median values of the SHARP AR parameters and assume the AR to have these parameters while traveling behind the limb.

Sun

Zone of reliable measurements of magnetic fields

Interplanetary magnetic field lines

Earth
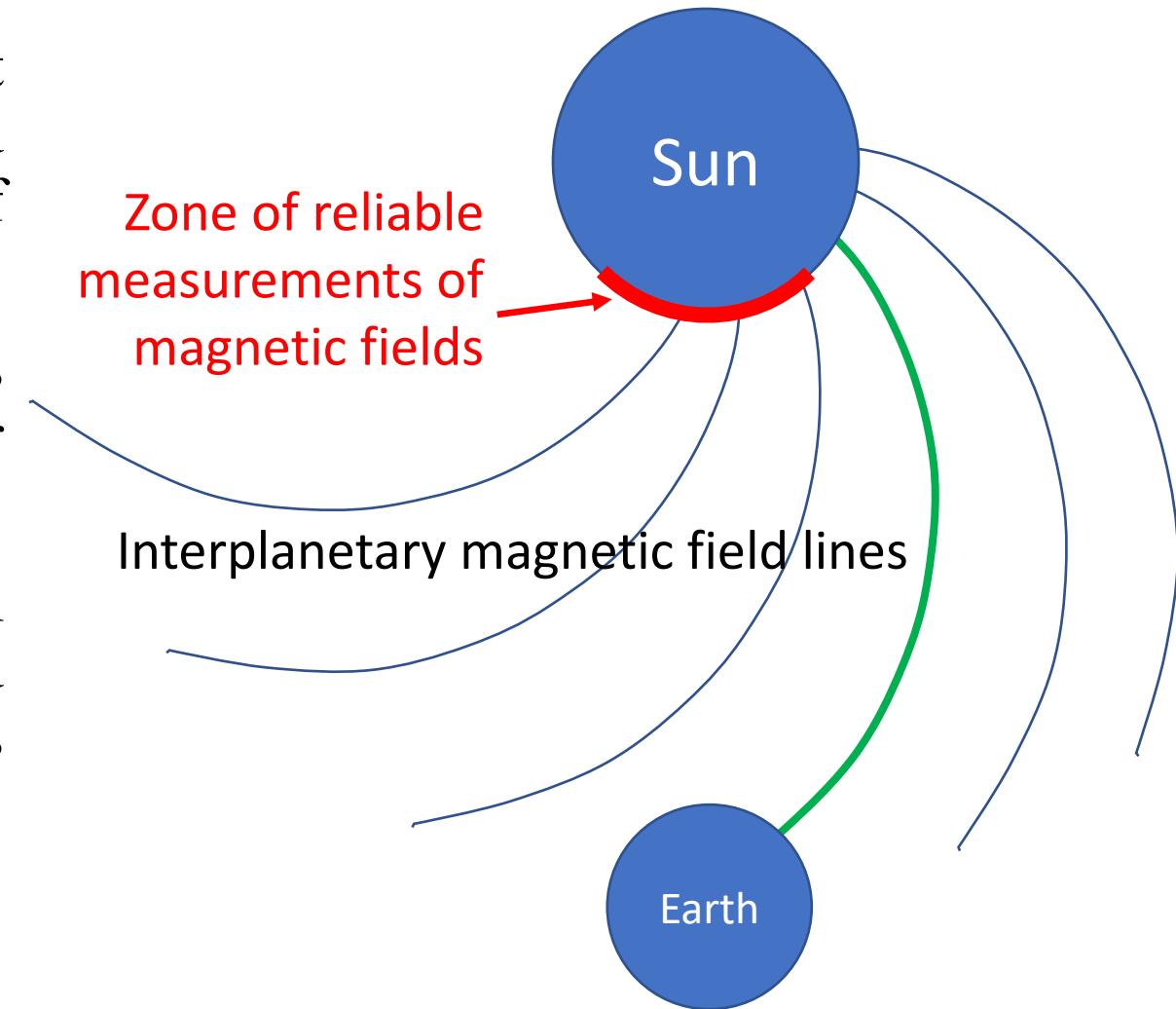
Extracted features

➢ Median values of SHARP properties for 10 ARs with the largest unsigned magnetic fluxes present on the Sun (including ARs behind the limb)

➢ Daily properties of SEP flux (mean, median, min, max, and last values, calculated for >10 MeV flux only)

➢ Daily properties of SXR flux (mean, median, min, max values, for fluxes in both the 0.5 – 4 Å and 1 – 8 Å channels)

➢ Statistics of Radio Bursts (number of type-II and type-IV bursts)

➢ Comparison with: SWPC NOAA daily operational forecasts

# Neural network architecture for whole-Sun SEP prediction

➢ The architecture is implemented with Python PyTorch

➢ AR features are processed in "AR Blocks". The weights are shared between the blocks.

➢ The number of AR Blocks remains the same for each day. The ARs with the highest magnetic fluxes serve as input.

➢ Whole-Sun features do not need to be linked to the ARs.

➢ The presented architecture allows us to address the problem of undefined-origin ARs for some SEP events.

AR Block Structure

AR Radial Magnetic Field Features

AR location (coordinates)

AR Block 1

AR Block 2

...

AR Block N

Whole-Sun features

Event Prehistory

AR block input:
SHARP features of AR, and AR location

AR block output:
Several outputs characterizing AR's "capability" to produce the SEP event

Main network input:
AR block outputs, features from the integrated observations of the Sun, event prehistory

Main network output:
A probability of the event to happen within the certain timeframe (24 hours for the proposed investigation)

Train-test separation and learning strategy

➢ Time periods in the training data set: 2010-2013, 2016, end of 2018-2019 (66 SPE days)

➢ Test data set: 2014-2015, 2017-beginning of 2018 (35 SPE days)

➢ An early stopping criterion is implemented on the test data set to prevent overfitting.

➢ The developed architecture is much more stable with respect to the fully-connected implementation.

➢ The procedure was performed 5 times for each investigated setup.
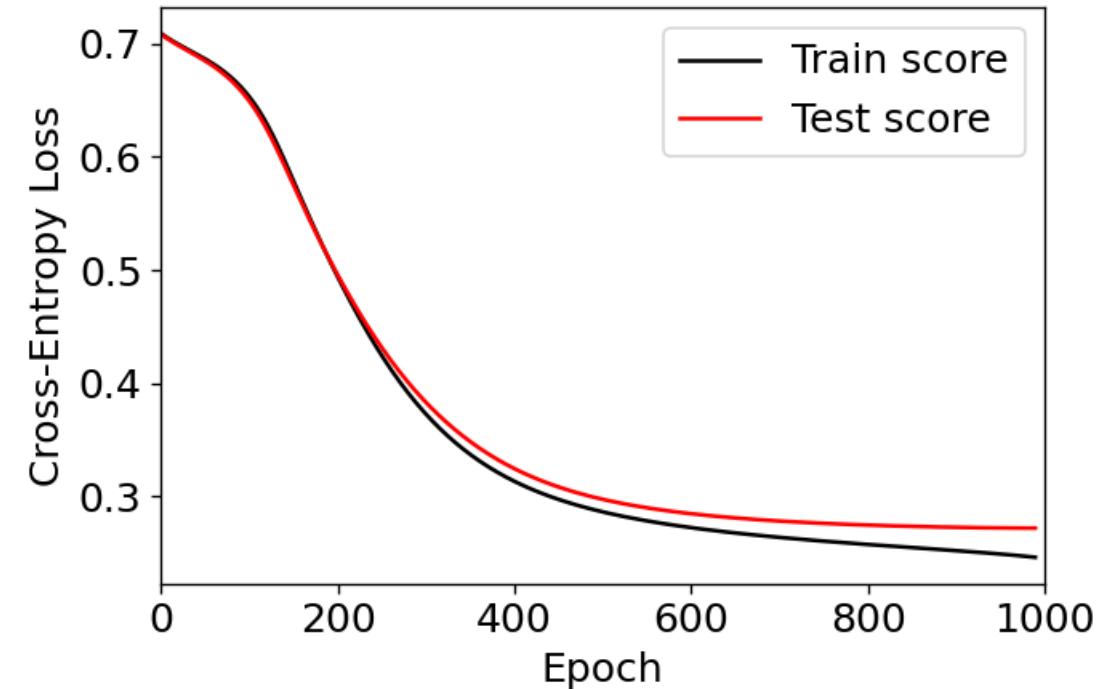
Important note: our goal is not to evaluate our predictor on the unknown data but to investigate how much we can learn from the available data in principle.

An example of cross-entropy loss for train and test data sets during the training progress

Inclusion/exclusion of parameters during the testing phase

➢ We would like to investigate how inclusion/exclusion of various parameters affects the prediction:

  ➢ Instead of adapting the network architecture to variable input, we "erase" the information for excluded descriptors (i.e., set the corresponding input to a constant unchanging value)

➢ Questions to be investigated:

  ➢ Comparison of the neural network prediction with SWPC NOAA forecast

  ➢ Exploration of the prediction solely based on SHARP properties

  ➢ Understanding the role of SHARP and proton flux properties in the prediction

 ➢ Exploration of the Receiver Operating Characteristic (ROC) curves

ROC curve obtained by varying the neural network output decision threshold
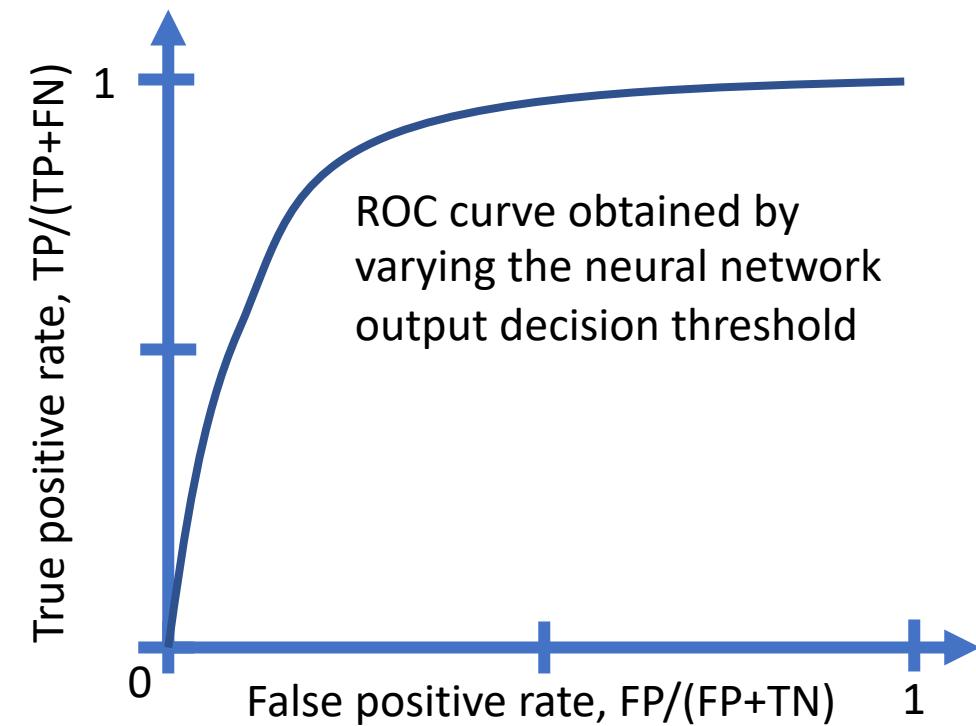
Comparison of ML prediction with SWPC NOAA and persistence forecasts

	Cross-entropy loss	TSS	Best testing example			
			TP	TN	FP	FN
ML prediction	0.271±0.005	0.775±0.004	31	1054	124	4
SWPC NOAA forecast	0.317	0.772	29	1111	67	6
Persistence model	-	0.647	23	1166	12	12

➢ Both the ML prediction and SWPC NOAA forecast are better than the persistence model (in terms of TSS).

➢ The ML prediction has the same TSS score as the SWPC NOAA forecast but has a lower cross-entropy loss.

➢ Let us now look at Receiver Operating Characteristic (ROC) curves for the network

ROC curves for the forecasts

➢ Although the TSS scores of both forecasts were the same, the ROC curves show a difference.

➢ The ML-based forecast clearly outperforms the SWPC NOAA forecast at the higher true positive rates relevant to all-clear forecasts.

➢ 4 SPE events were totally missed by SWPC NOAA forecasts: the corresponding issued probabilities were 1% (lowest-issued probability).
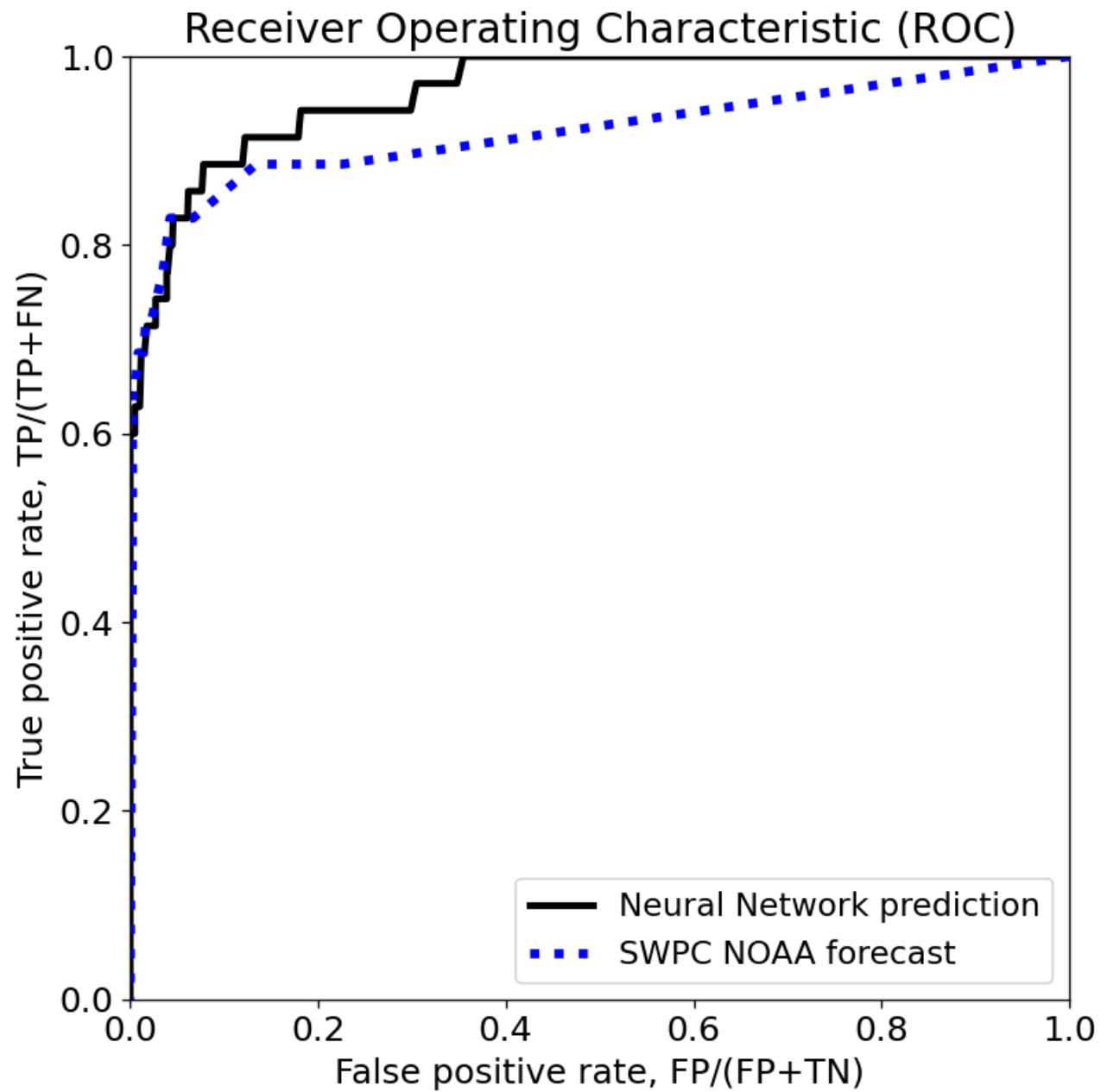
Understanding: ML prediction based on SHARP AR parameters only

	Cross-entropy loss	TSS	Best testing example			
			TP	TN	FP	FN
SHARP characteristics	0.477±0.003	0.574±0.006	32	787	391	3
Coordinates excluded	0.480±0.001	0.553±0.003	32	758	420	3
No behind-the-limb extension	0.589±0.018	0.473±0.013	30	742	436	5

➢ The neural network learns almost nothing if no behind-the-limb extension of active regions is implemented.

➢ The benefits from including AR coordinates are doubtful. Let's look at ROC curves for these forecasts.

ROC curves (AR-based predictions)

➢ Inclusion of AR coordinates does not improve the prediction (although it has a higher TSS score).

➢ AR-based predictions are worse than the SWPC NOAA operational forecasts in the region of low false positive rates.

➢ However, there are certain advantages of AR-based forecasts: it is possible to predict all events and have the false positive rate never equal 1.
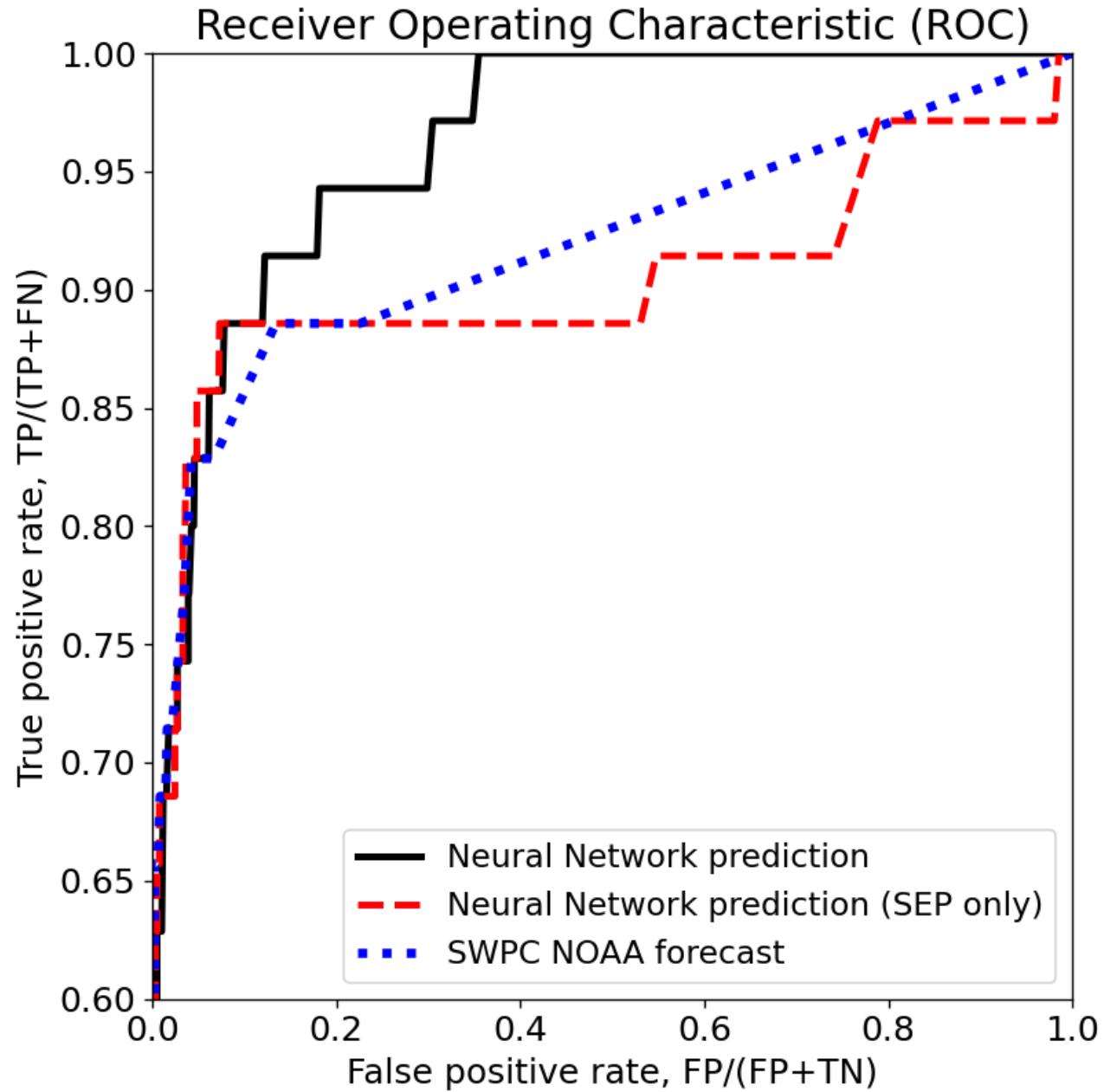

Receiver Operating Characteristic (ROC)

— AR-based prediction
- - - AR-based prediction (no coordinates)
····· SWPC NOAA forecast
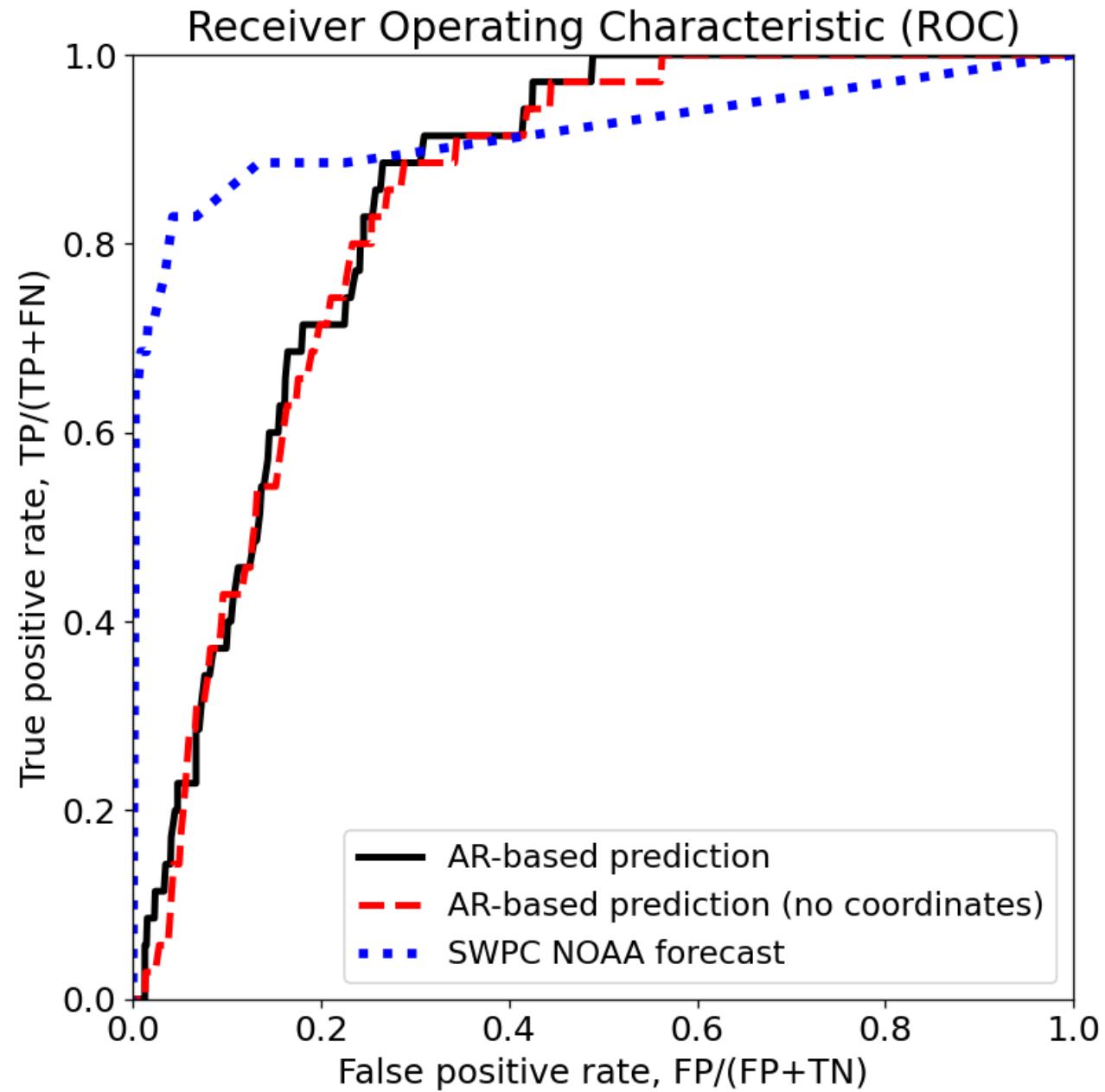
Understanding: inclusion/exclusion of features for ML prediction of SEPs

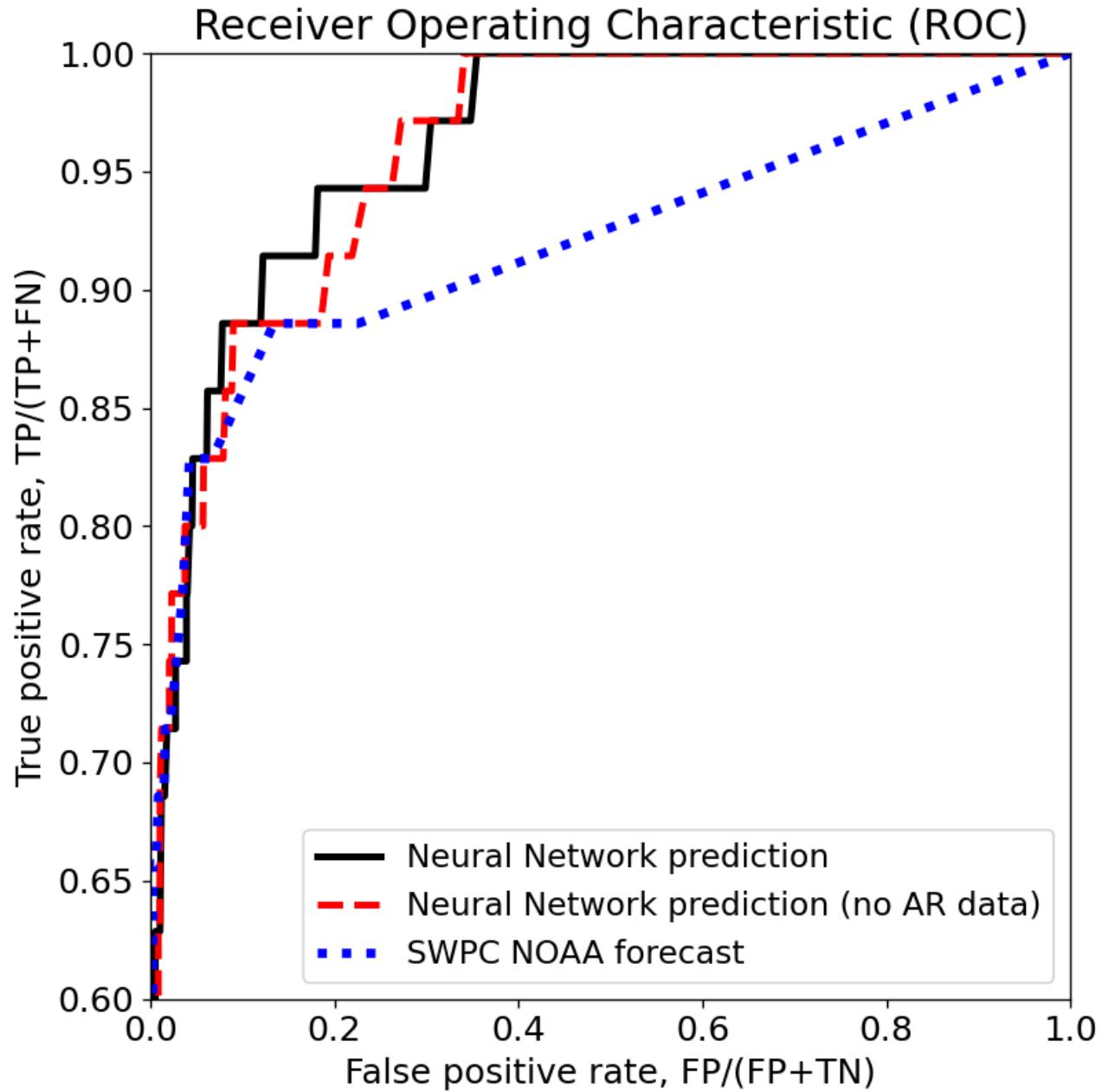|	Cross-entropy loss	TSS	Best testing example			
			TP	TN	FP	FN
All properties included	0.271±0.005	0.775±0.004	31	1054	124	4
AR information excluded	0.265±0.001	0.772±0.001	31	1046	132	4
SEP information excluded	0.497±0.008	0.499±0.007	31	734	444	4
SXR information excluded	0.282±0.015	0.765±0.015	31	1049	129	4
Radio burst information excluded	0.269±0.003	0.777±0.001	31	1054	124	4
SEP descriptors only	0.312±0.003	0.788±0.001	29	1131	47	6

ROC curves (SEP characteristics)

➢ Inclusion of SEP characteristics is the most critical for network performance.

➢ The predictions behave very similarly to SWPC NOAA forecasts if trained on SEP characteristics only.

Receiver Operating Characteristic (ROC)

True positive rate, TP/(TP+FN) vs False positive rate, FP/(FP+TN)

— Neural Network prediction
- - - Neural Network prediction (SEP only)
····· SWPC NOAA forecast

ROC curves (SEP characteristics)

➢ Exclusion of AR characteristics does not significantly affect the predictions

➢ There are two possible explanations:

  ➢ All the necessary information is already contained in the SXR activity of the Sun.

  ➢ Inclusion of AR dynamics is necessary for prediction capabilities.

Receiver Operating Characteristic (ROC)

— Neural Network prediction
- - - Neural Network prediction (no AR data)
····· SWPC NOAA forecast

Weighted TSS (WTSS) score

There is one more way to approach the "all-clear" forecast:

➢ True Skill Statistics score is defined as:

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} = 1 - \frac{FN}{P} - \frac{FP}{N}$$

➢ Let's apply weights to the missed event rate and the false alarm rate:

$$WTSS(\alpha) = 1 - \frac{2}{\alpha + 1}\left(\alpha \frac{FN}{P} + \frac{FP}{N}\right)$$

➢ The parameter α indicates how much stronger our preference is for making the missed event rate lower with respect to the false alarm rate

➢ The WTSS(α) score has the same properties as the TSS score:

    ➢ It ranges from -1 (totally wrong forecasts) to 1 (fully correct forecasts), where 0 corresponds to random guess forecasts.

    ➢ It is not sensitive to the class-imbalance ratio.

    ➢ $WTSS(1) = TSS$

WTSS results

➤ The SWPC NOAA operational forecast and ML-based prediction are almost the same for $\alpha < 1$.

➤ The ML-based prediction outperforms the SWPC NOAA operational forecast for $\alpha \geq 1$.

➤ Predictions based on AR parameters only have significantly lower scores than the other two predictions (SWPC NOAA and using all parameters) for small α.
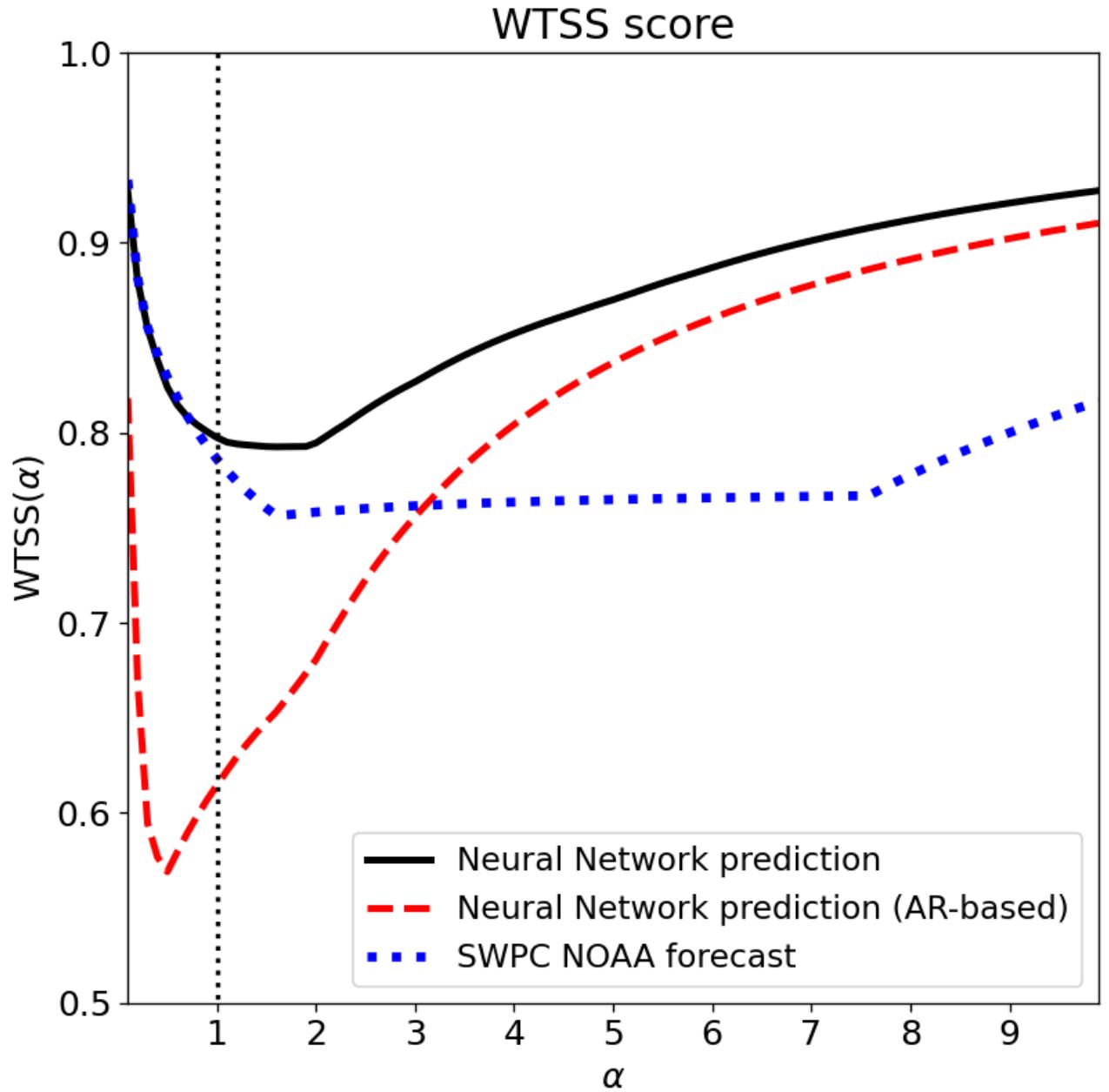
Summary of the results

➢ Even a feature-based binary classification is an interesting problem!

➢ Inclusion of the western limb and far-side ARs is necessary if the AR features are considered in the forecast.

➢ Inclusion of SEP characteristics is the most critical for prediction.

➢ Exclusion of AR characteristics (in the form used in this study) does not seem to affect the predictions.

➢ Machine learning-based forecast seems to be very promising in situations when missed events are very undesirable ($\alpha > 1$ for WTSS). This is a good sign for "all-clear" forecast development!

Proposed Year 2 Milestones (ML-related)

➢ Complete the development and analysis of the parameter-dependent binary and probabilistic "all-clear" SPE forecasts for different timescales, different particle flux and energy thresholds, and subject to availability of different data sources.

➢ Perform initial investigation of the enhancement of "all-clear" forecasts by predicting AR descriptors using Recurrent Neural Networks.

Year 2 Milestone 1: Details

➢ Continue the development and analysis of the parameter-dependent binary and probabilistic "all-clear" SPE forecasts for different timescales, different particle flux and energy thresholds, and subject to availability of different data sources.

  ➢ Investigate how inclusion/exclusion of features from various data sources (soft X-ray properties of solar flares, identified CME records, neutron monitor data, etc) affect the SPE prediction

  ➢ Complete the development and deploy the operational binary and probabilistic SPE forecasts with SWPC NOAA-comparable settings (daily-based prediction of > 10 MeV > 10 pfu proton events) relying on a longer-term data set (since 1996).

  ➢ Investigate the possibility to forecast SPE events for other energy and flux thresholds (with the particular interest on forecasting > 100 MeV > 1 pfu events) and timescales (longer-term all-clear forecasts and shorter-term warnings).

Year 2 Milestone 2: Details

➢ Perform initial investigation of the enhancement of "all-clear" forecasts by predicting AR descriptors using Recurrent Neural Networks and by analyzing time series history of the descriptors.

  ➢ Our preliminary study indicates importance of ARs at and behind the Western limb. The RNN-based approach will allow us to estimate more correctly the parameters of extended ARs.

  ➢ Another way to enhance SPE forecasts (and a possible alternative to RNN-based model for AR extension) is to utilize the time series history of the descriptors. The following models will be considered:

    ➢ RNN-based models *

    ➢ Convolutional Neural Networks (CNNs) *

    ➢ Temporal Attention Models (TAMs)

    The model is tested on the proton flux prediction problem

Thank You for Your Attention!