PhD Thesis defense

Deep Neural Networks for Geomagnetic Indices Forecasting

Armando Collado Villaverde {armando.collado@uah.es}

July 2025

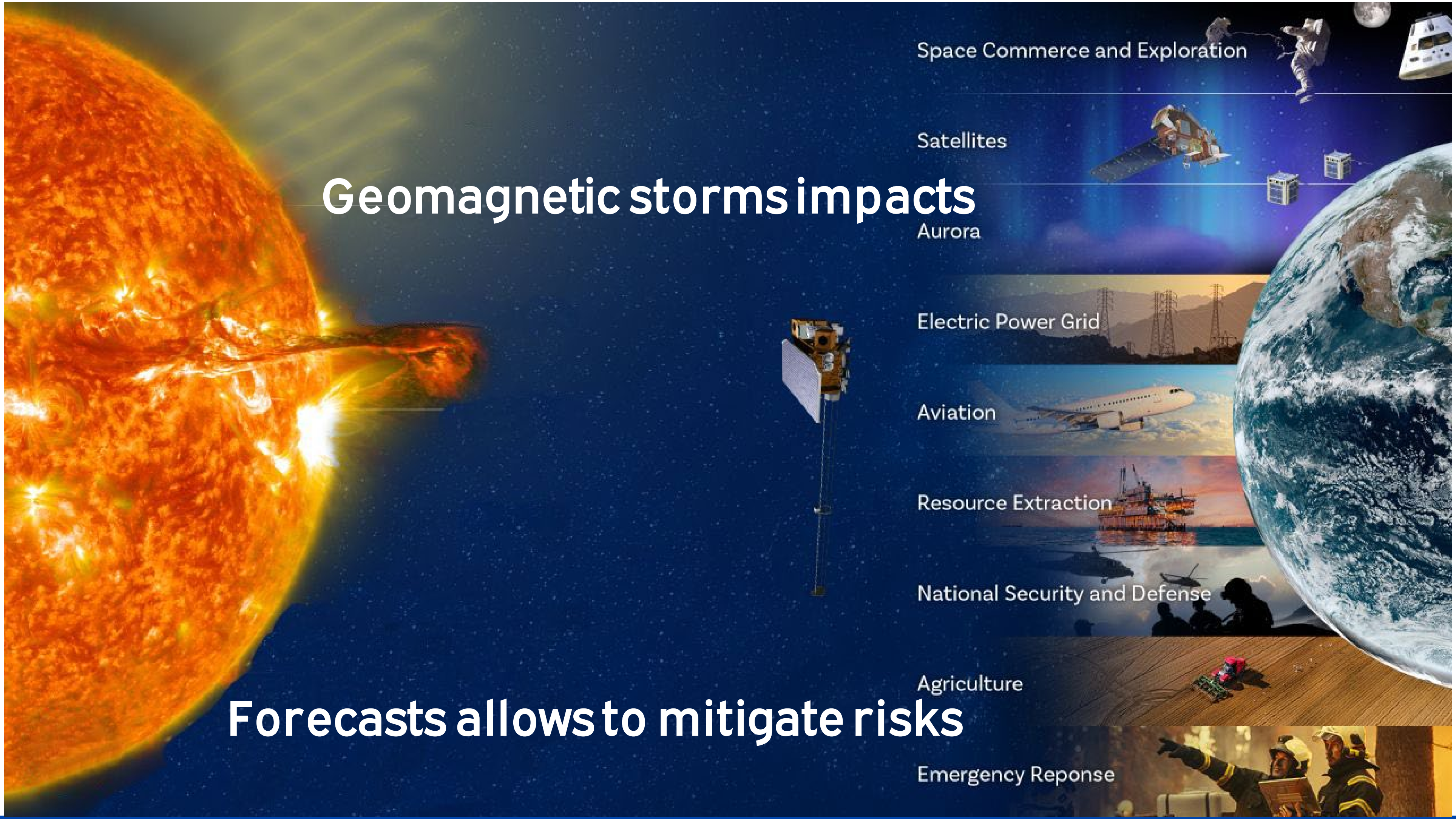
esa SWE Space Weather Group

Agenda

1. Introduction and context
2. Deep Learning fundamentals and its application to Space Weather
3. Training and evaluation framework
4. Model development and architecture
5. Adaptation to local indices
6. Operational case study: storm of May 2024
7. Conclusions and future work

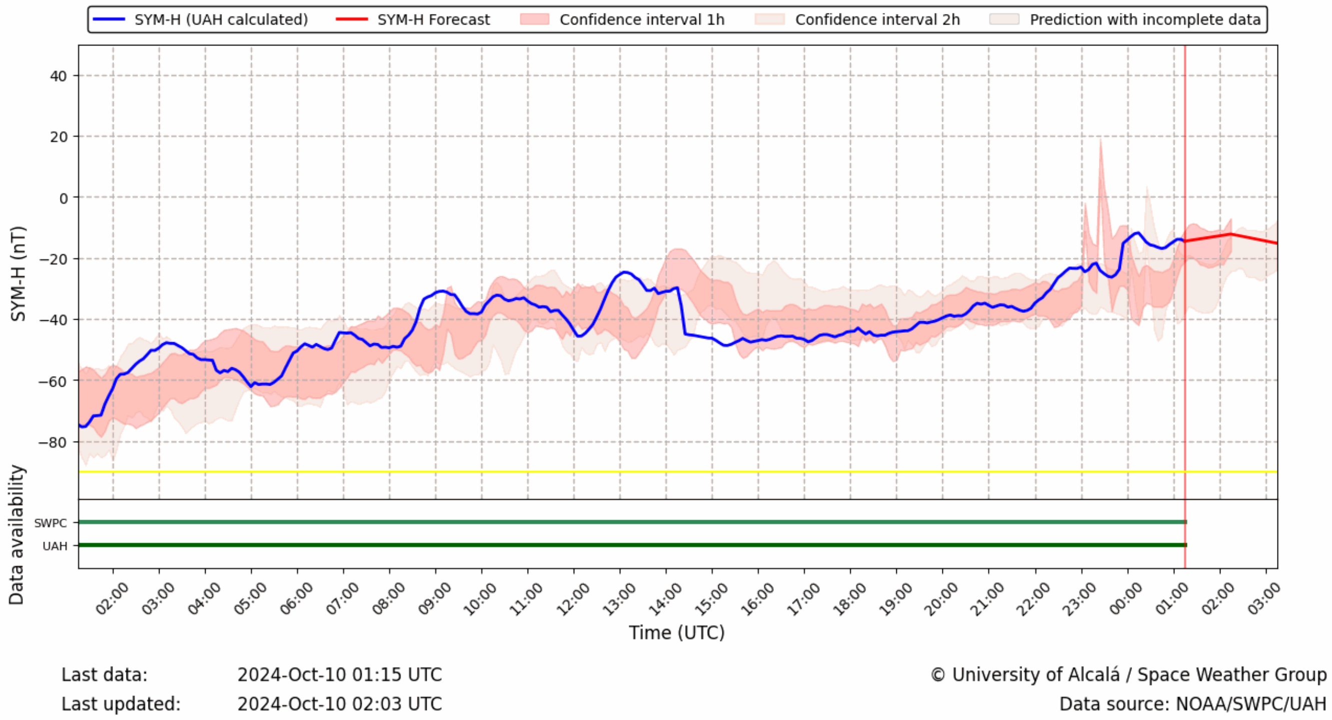
Introduction and context

Research Motivation



Introduction and context
Objectives of the Thesis

- Primary objective:
  - Develop Deep Neural Networks for real-time forecasting of geomagnetic indices
- Supporting objectives:
  - Address challenges in data preprocessing, including missing data
  - Establish a comprehensive training and evaluation framework
  - Quantify uncertainty in the forecasts.
  - Extend the application of DNN models to local geomagnetic indices



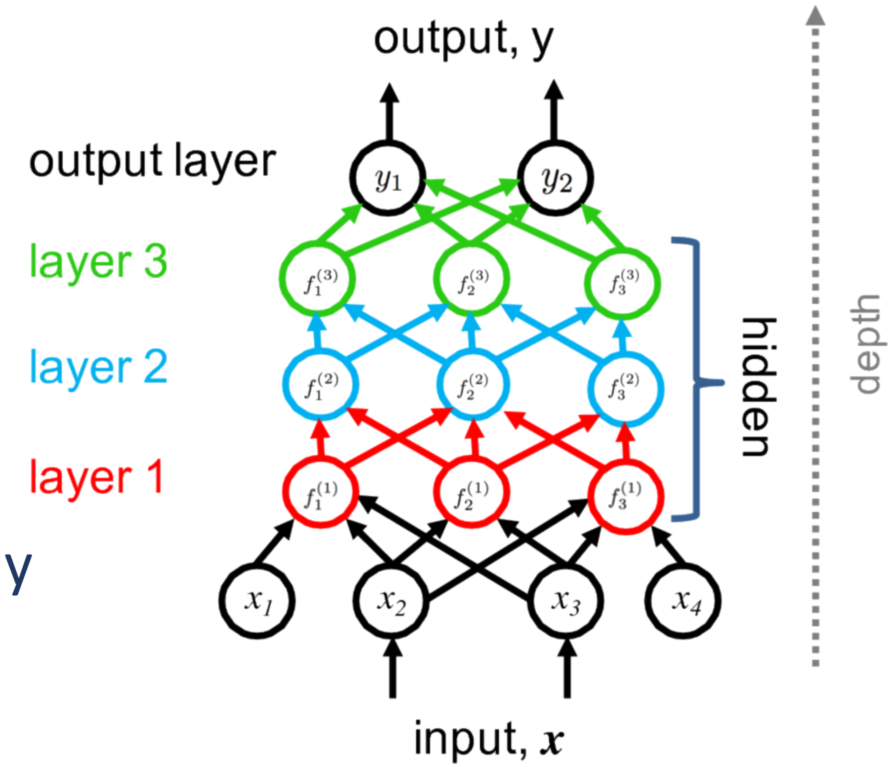
Agenda

1. Introduction and context
2. Deep Learning fundamentals and its application to Space Weather
3. Training and evaluation framework
4. Model development and architecture
5. Adaptation to local indices
6. Operational case study: storm of May 2024
7. Conclusions and future work

Deep Learning and its application to Space Weather

Overview of Machine Learning and Neural Networks

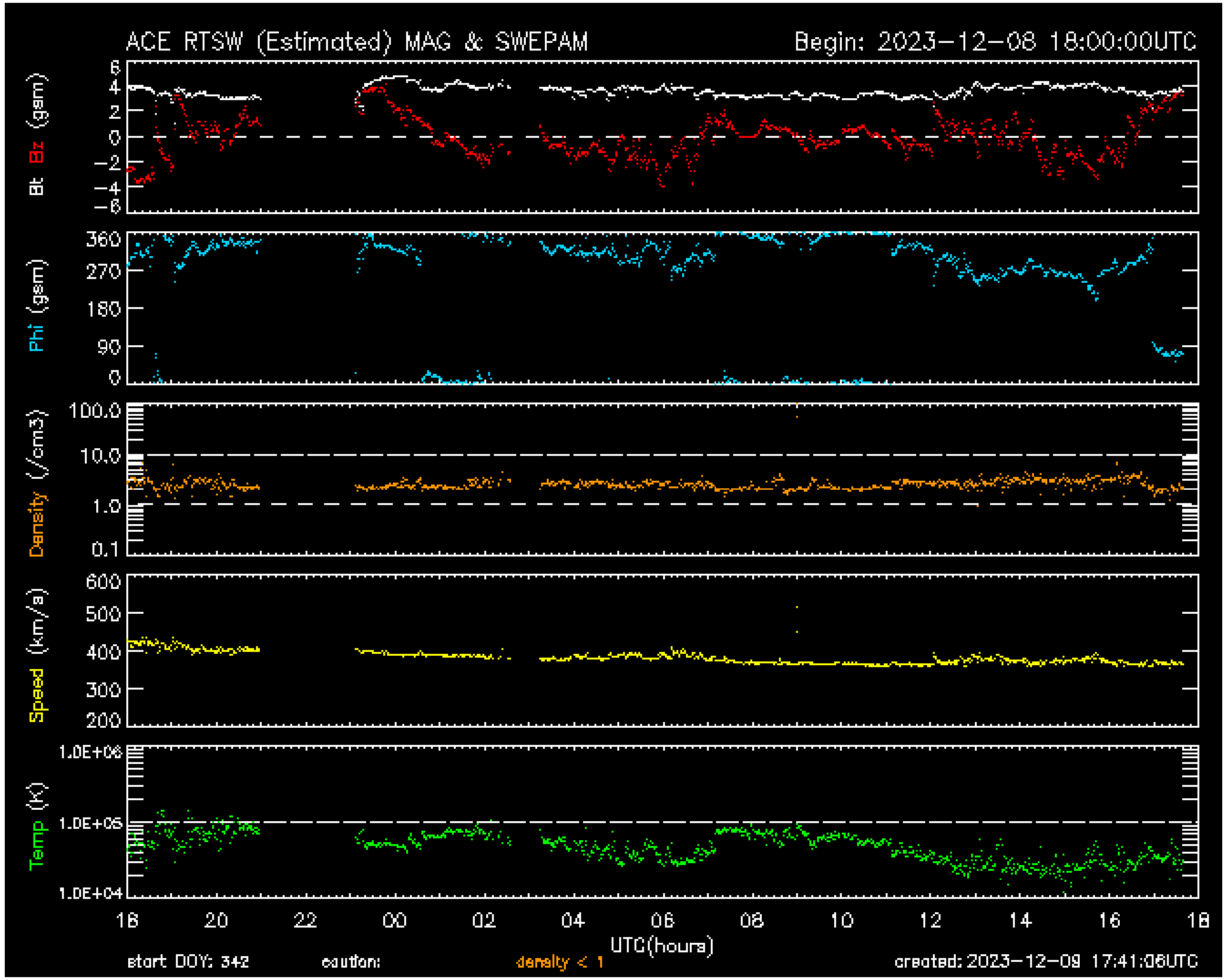
- Machine Learning
  - Subset of Artificial Intelligence, algorithms that map input data to outputs, learning patterns and making predictions without explicit programming
  - The learning process involves training models on historical data, giving examples of the input and expected output
- Neural networks: the core of deep learning
  - Each layer transforms the input data into increasingly abstract representations
  - Flexible and scalable
  - Performance is closely related to the available data



Deep Learning and its application to Space Weather

Challenges

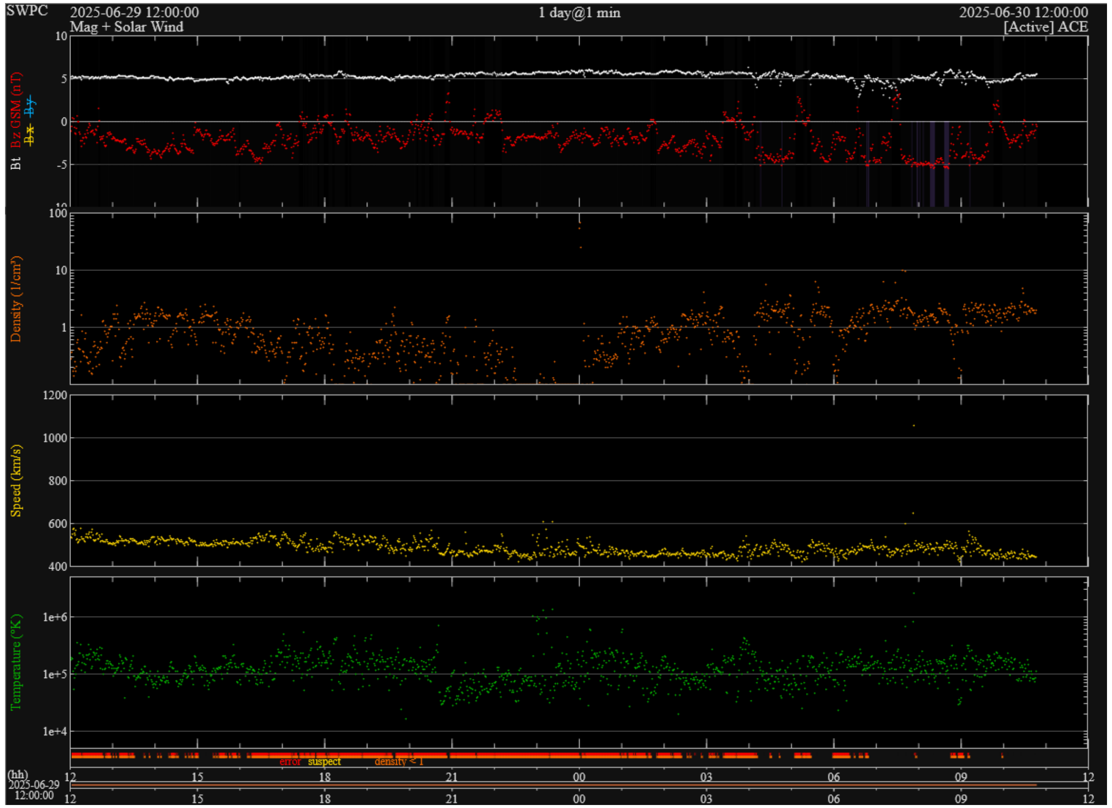
- Dynamic Space Weather:
  - Rapid changes in conditions
  - Need for immediate data processing
  - We can only use features available in real-time
- Data challenges:
  - Difference between historical and real-time data
  - Completeness of input data
- Features from ACE:
  - Magnetometer instrument
  - SWEPAM instrument (plasma)



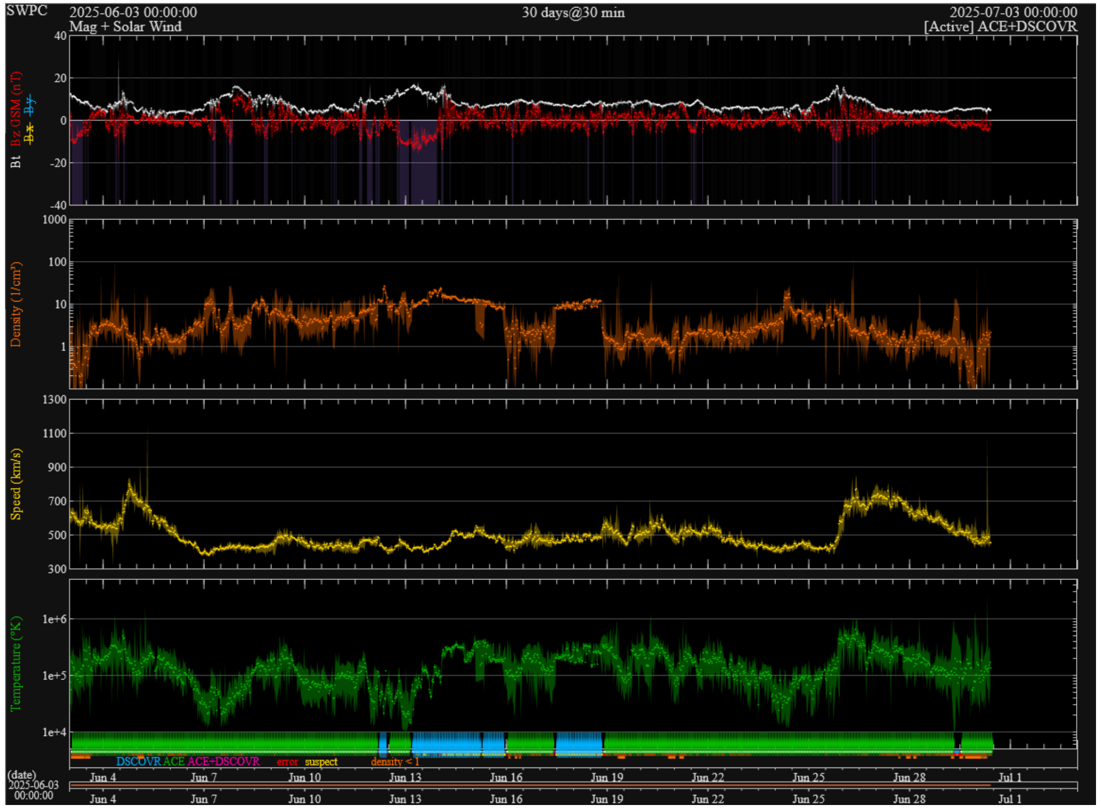
Deep Learning and its application to Space Weather

Challenges

Last day → 100% ACE



Last month → mostly ACE

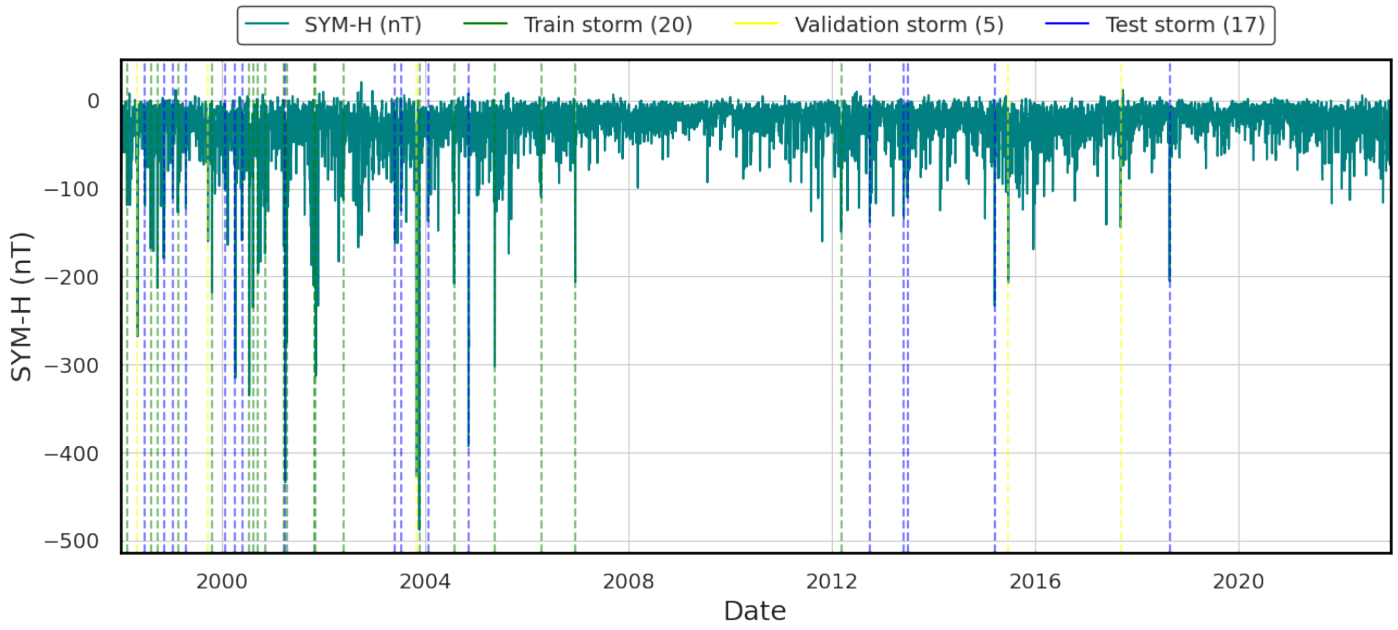


Green -> ACE data | Blue -> DSCOVR

Deep Learning and its application to Space Weather

Challenges

- Intense storms are naturally rare
- Related works have used the same set of storms to allow a fair comparison
- Stablished in 2020 using 42 storms
  - 20 for training
  - 5 for validation
  - 17 for testing



Agenda

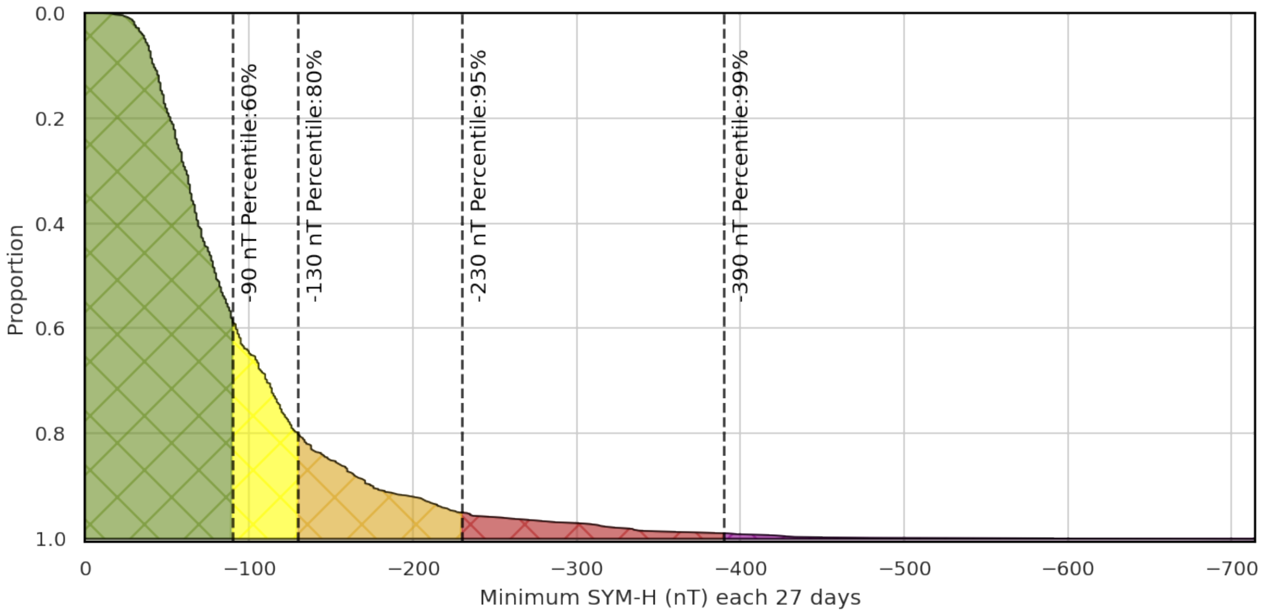
1. Introduction and context
2. Deep Learning fundamentals and its application to Space Weather
3. Training and evaluation framework
4. Model development and architecture
5. Adaptation to local indices
6. Operational case study: storm of May 2024
7. Conclusions and future work

Training and evaluation framework

SYM-H storm classification

- Framework to classify storms based on intensity
- Create statistically backed subsets and expand existing ones
- Based on the Cumulative Distribution function and percentage of occurrence

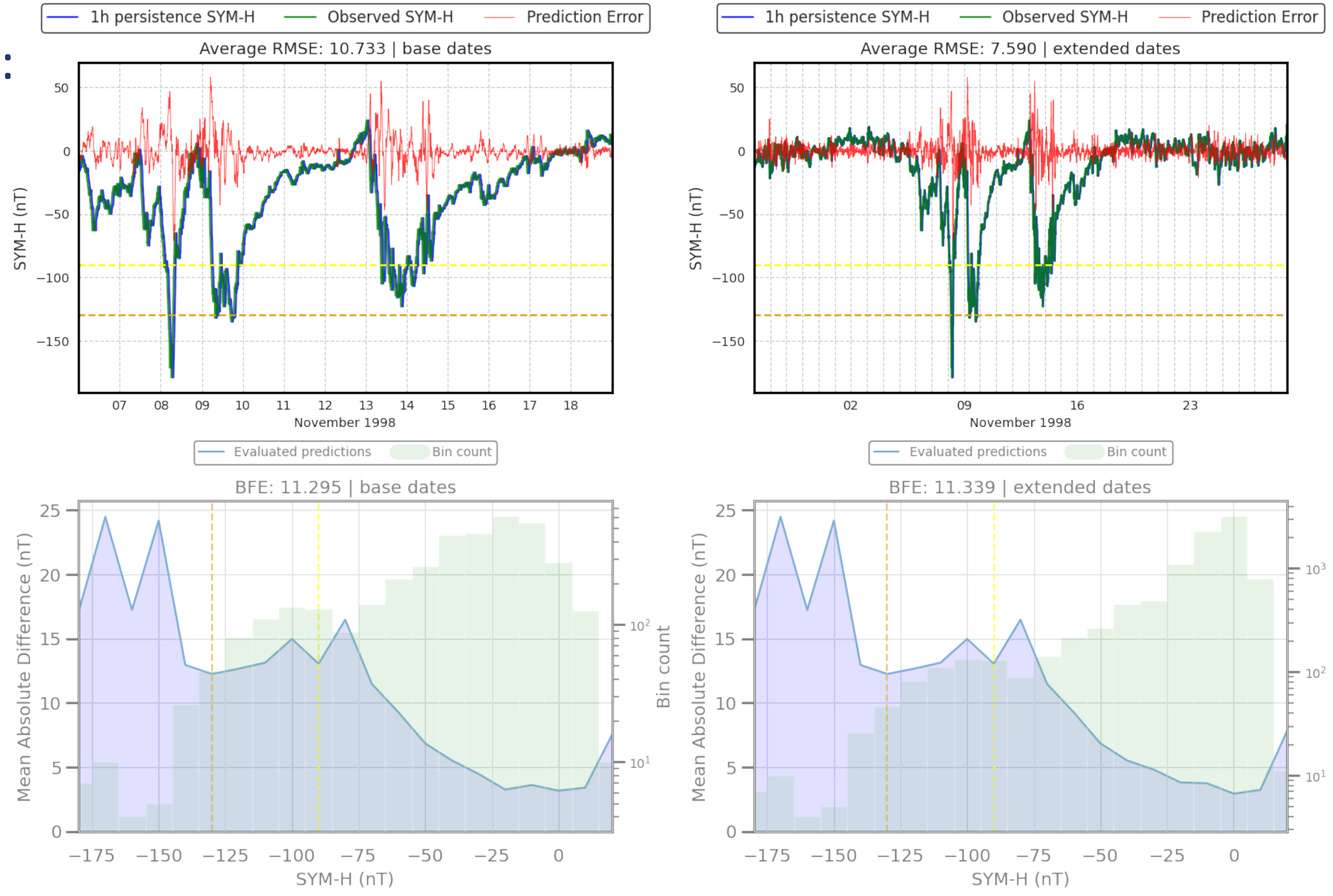
Intensity	SYM-H (nT)	Percentile (%)	# Data points
Inactive	[-90, ∞)	[60, 0)	330
Low	[-130, -90)	[80, 60)	123
Moderate	[-230, -130)	[95, 80)	87
Intense	[-390, -230)	[99, 95)	22
Superintense	[-∞, -390)	(100, 99)	7



Training and evaluation framework

Performance metrics

- Traditional Forecasting Assessment Metrics:
  - R^2 and RMSE for model evaluation.
  - Average error across all the storm duration, require same dates
  - Storm peak errors and inactive period have same importance
  - Good performance during inactive can mask poor performance during critical active times
- Propose a metric that differentiates storm activity levels: BFE
  - Resilient to differences in storm duration

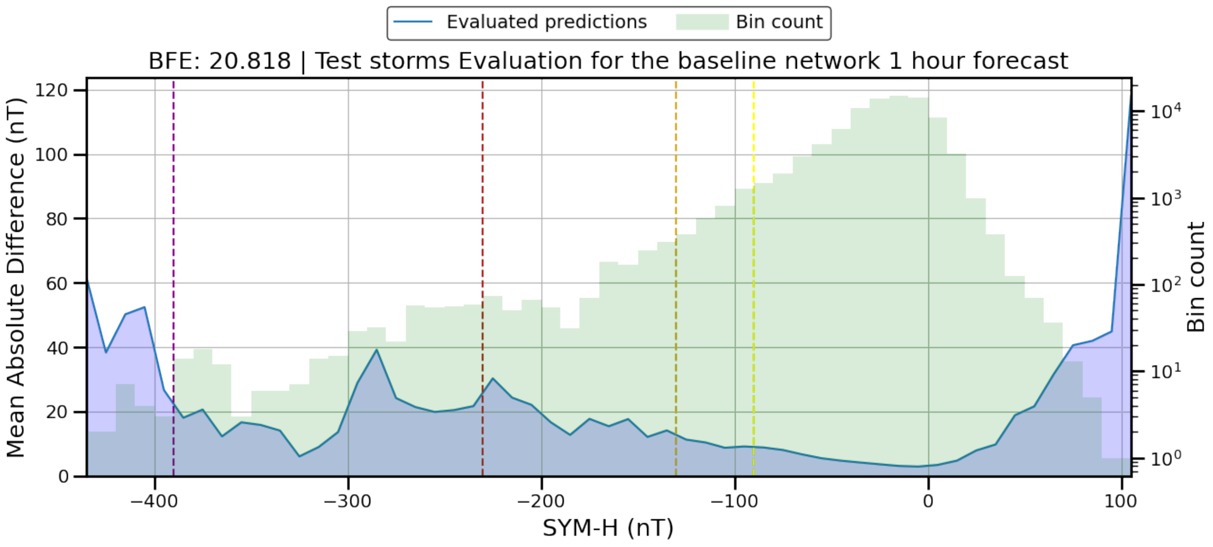


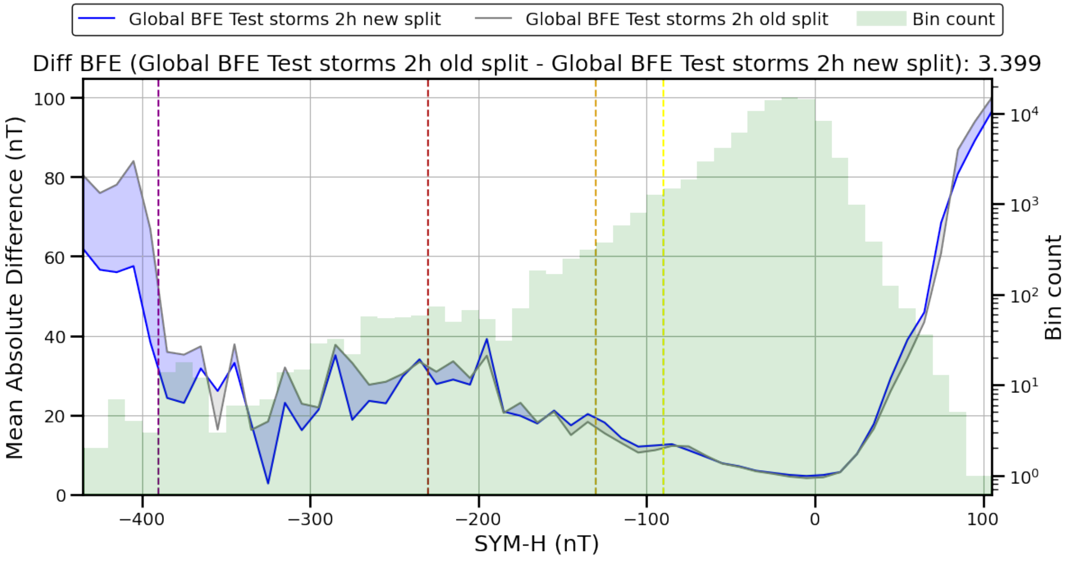
Training and evaluation framework
Performance metrics

- Traditional Forecasting Assessment Metrics:
  - R^2 and RMSE for model evaluation.
  - Average error across all the storm duration, require same dates
  - Storm peak errors and inactive period have same importance
  - Good performance during inactive can mask poor performance during critical active times
- Propose a metric that differentiates storm activity levels: BFE
  - Resilient to differences in storm duration

Training and evaluation framework
Performance metrics

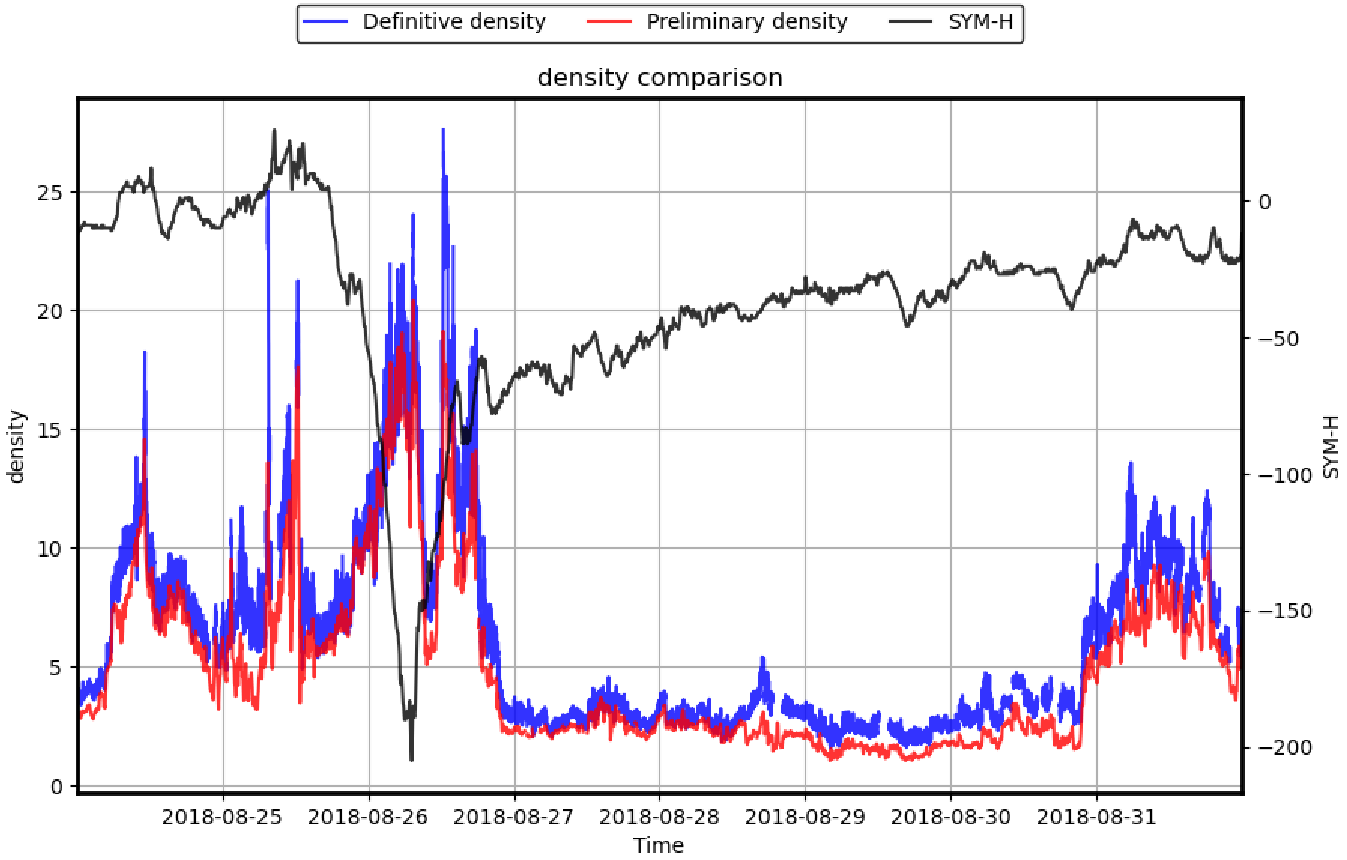
- Provides user-friendly model performance comparisons
- Validates model precision across storm intensities
- Better comparison between models





Training and evaluation framework
Testing on key parameters

- Historical definitive data is different to real-time data
- Test on preliminary parameters
  - Closer to the data available in real time
  - Only available for a limited time (2017 onward)
  - Lower resolution
  - Contains more errors

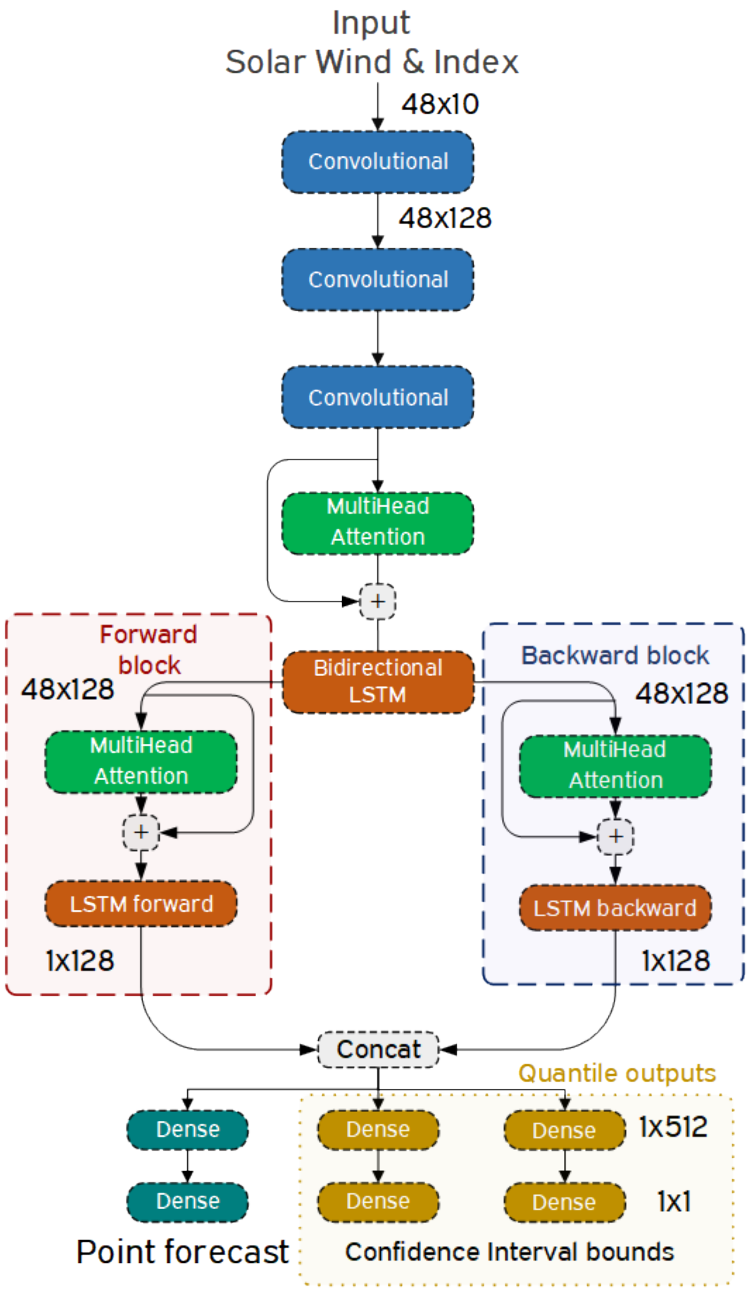


Agenda

Model development and architecture

Neural Network

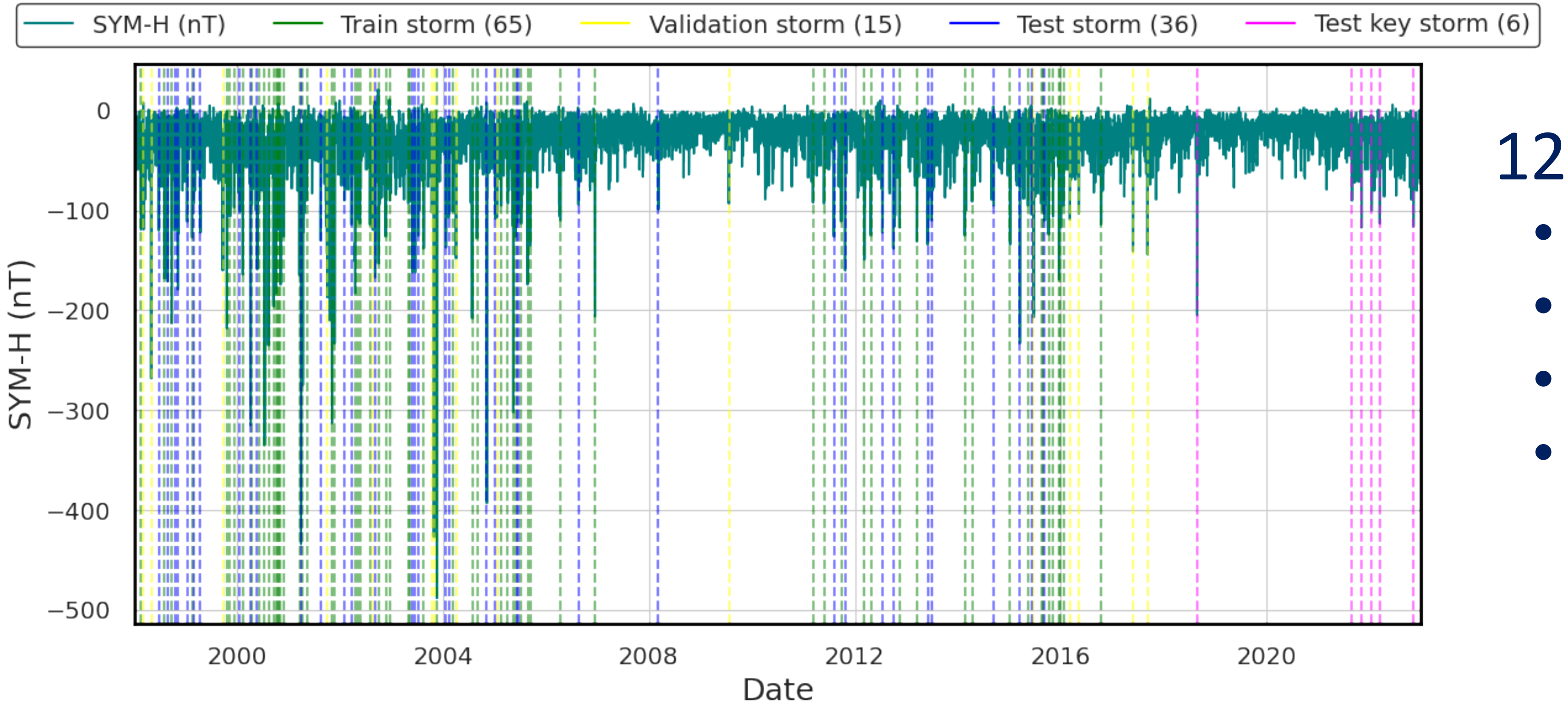
- Input: IMF, solar wind plasma and previous values of the index for the last 4 hours in 5-min averages
- Multiple outputs:
  - Point forecast in the next 1 and 2 hours (MSE)
  - Prediction interval with 90% confidence (Quantile)



Model development and architecture

Training

- We have expanded storms selection with recent storms
- BFE as validation metric

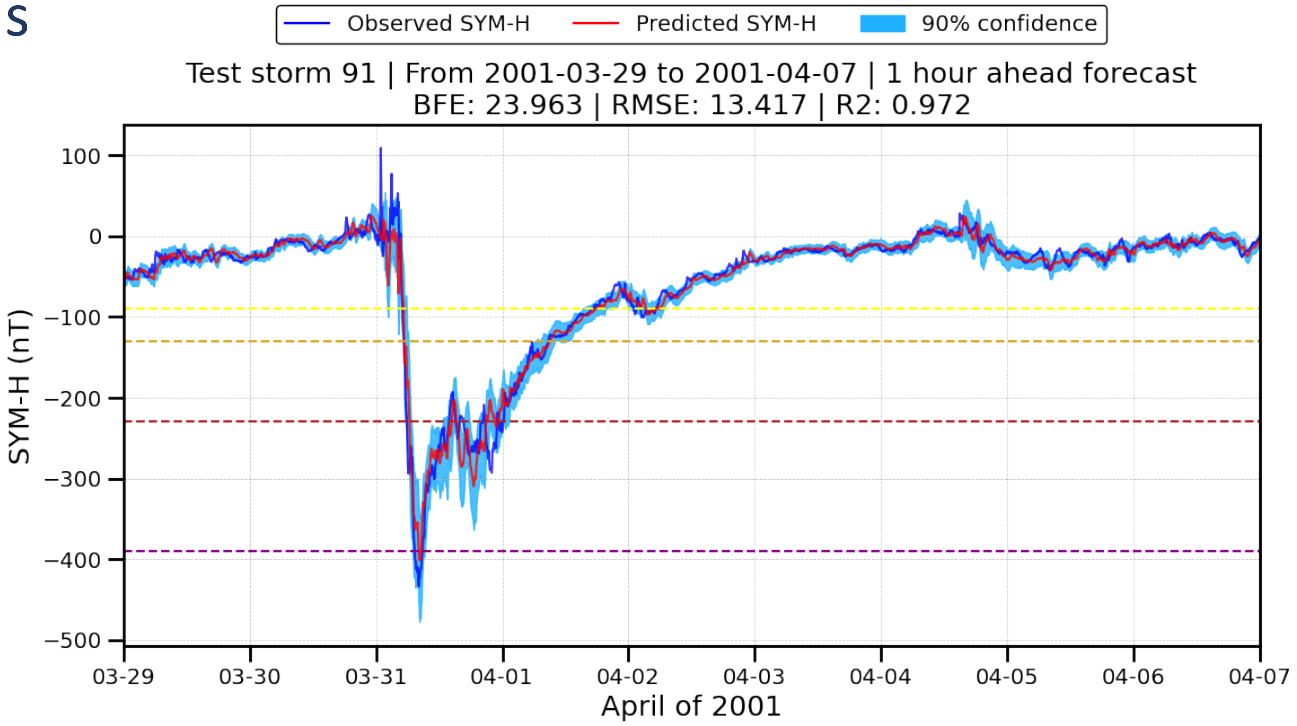


122 storms:

- 65 training
- 15 validation
- 36 test
- 6 test with key parameters

Model development and architecture
Results

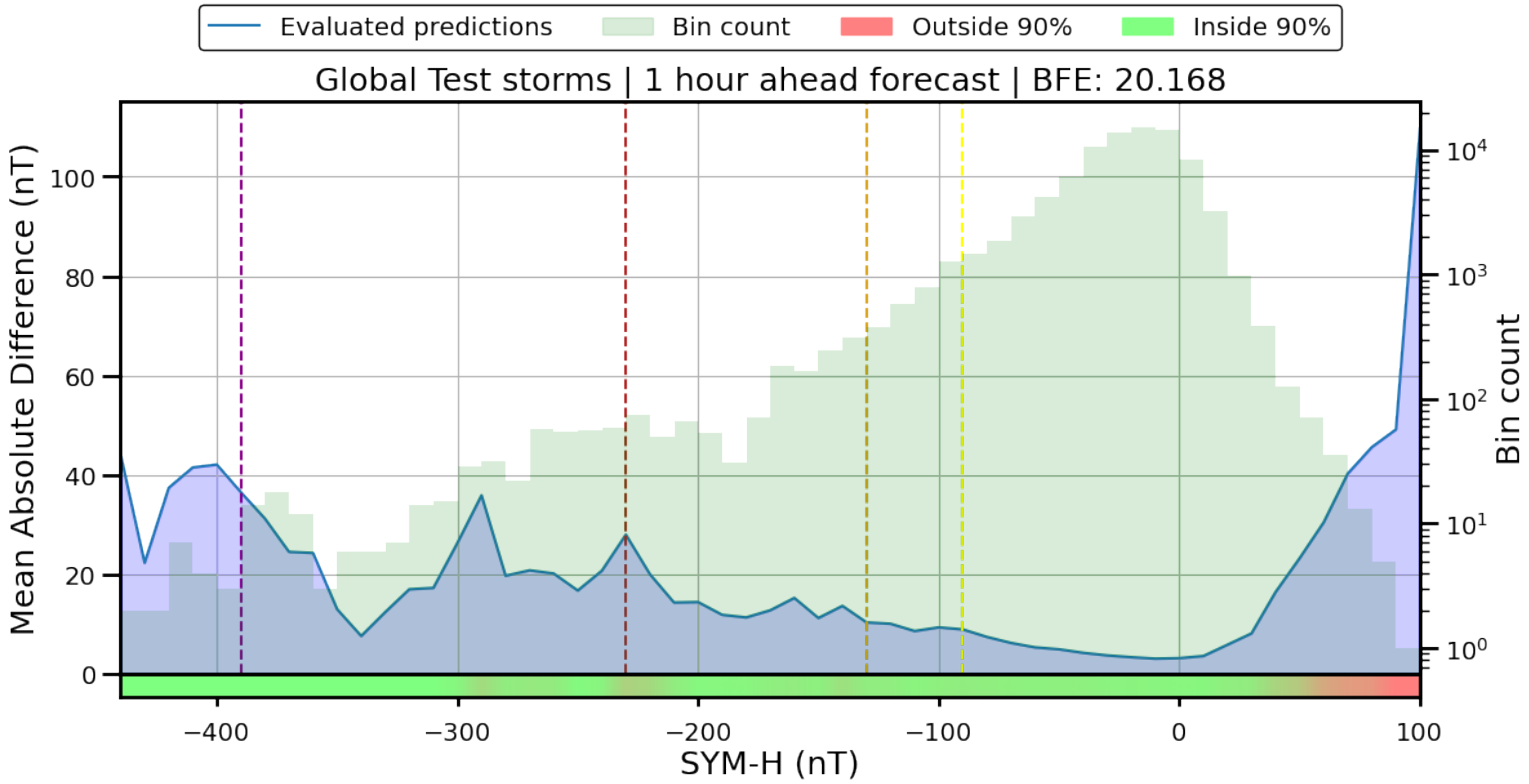
- Evaluate both the global concatenation of all the storms and the average of all the test storms
  - Average of the storms can be skewed due to class unbalance
  - Point forecast metrics: BFE, RMSE & R2
- Extra metrics for the prediction intervals:
  - PICP: Percentage of values inside the interval
  - PIAW: Average width of the prediction interval
  - PIBW: Binned width of the prediction interval



Model development and architecture

Results test 1h – BFE

- Bottom heatmap shows the percentage of forecast inside the interval



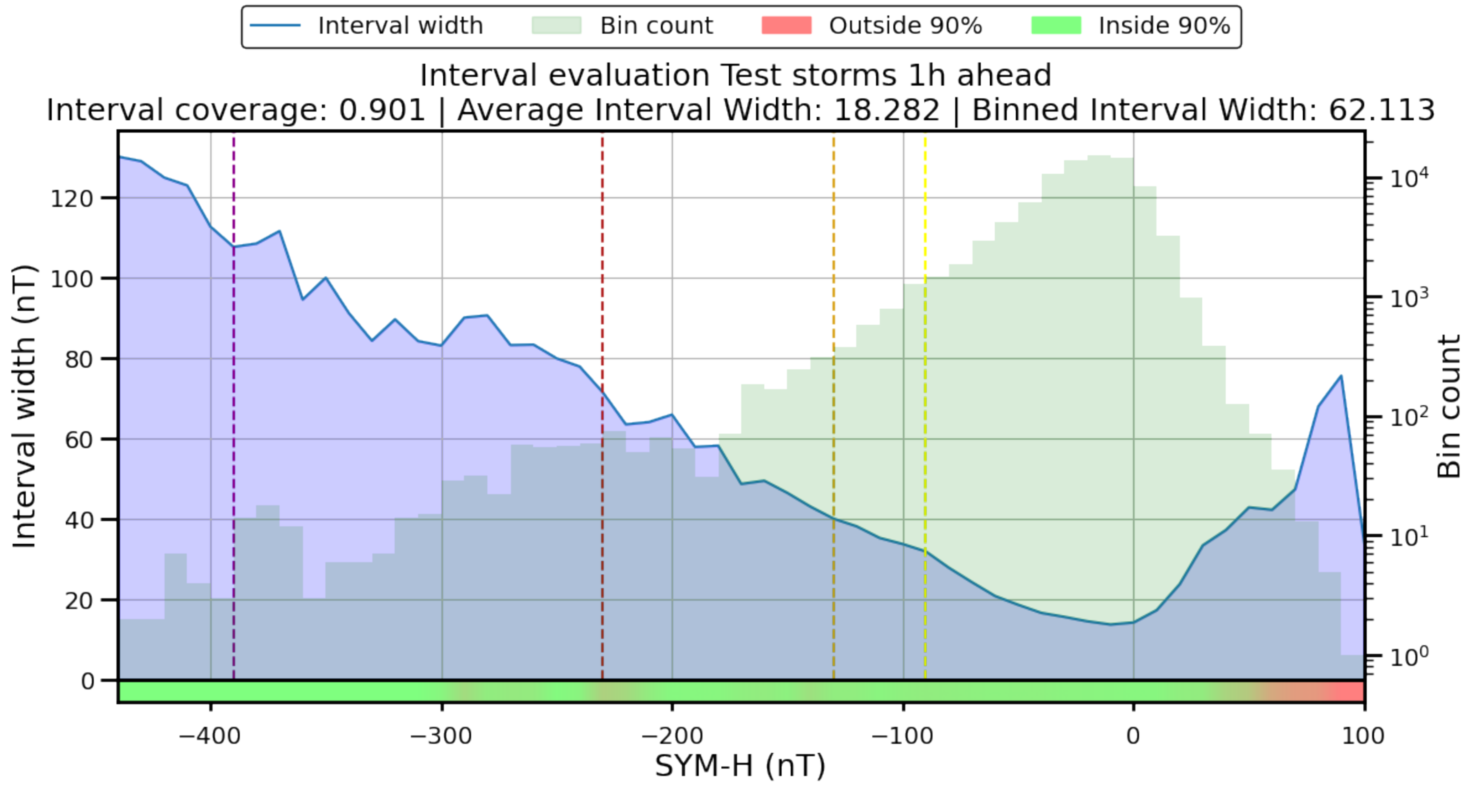
Mode	BFE	RMSE	R2
Mean	9.612	6.818	0.935
Global	20.168	7.263	0.961

Mode	PICP	PIAW	PIBW
Mean	0.903	18.086	29.889
Global	0.901	18.282	62.113

Model development and architecture

Results test 1h – Interval width

- Bottom heatmap shows the percentage of forecast inside the interval



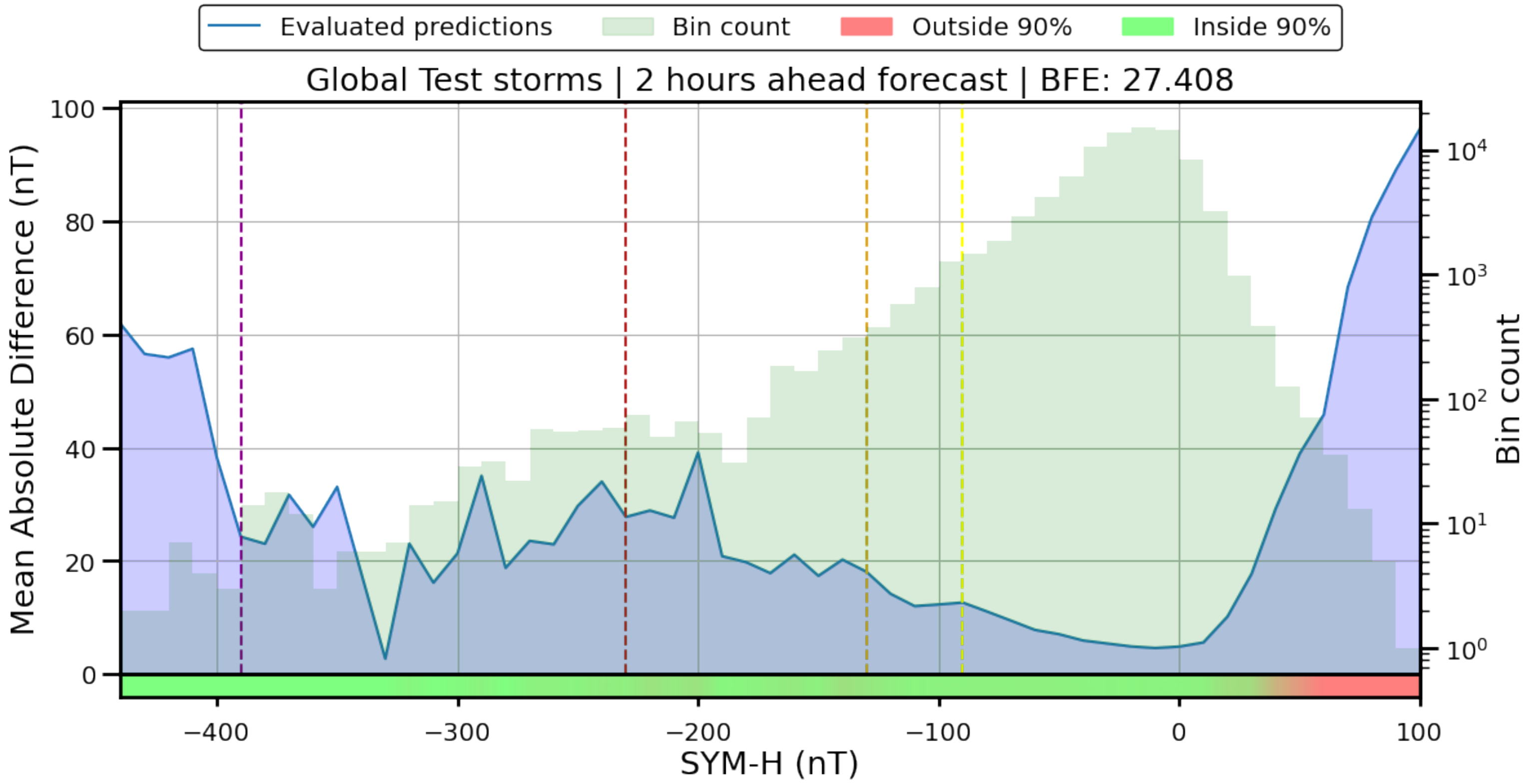
Mode	BFE	RMSE	R2
Mean	9.612	6.818	0.935
Global	20.168	7.263	0.961

Mode	PICP	PIAW	PIBW
Mean	0.903	18.086	29.889
Global	0.901	18.282	62.113

Model development and architecture

Results test 2h – BFE

- Bottom heatmap shows the percentage of forecast inside the interval



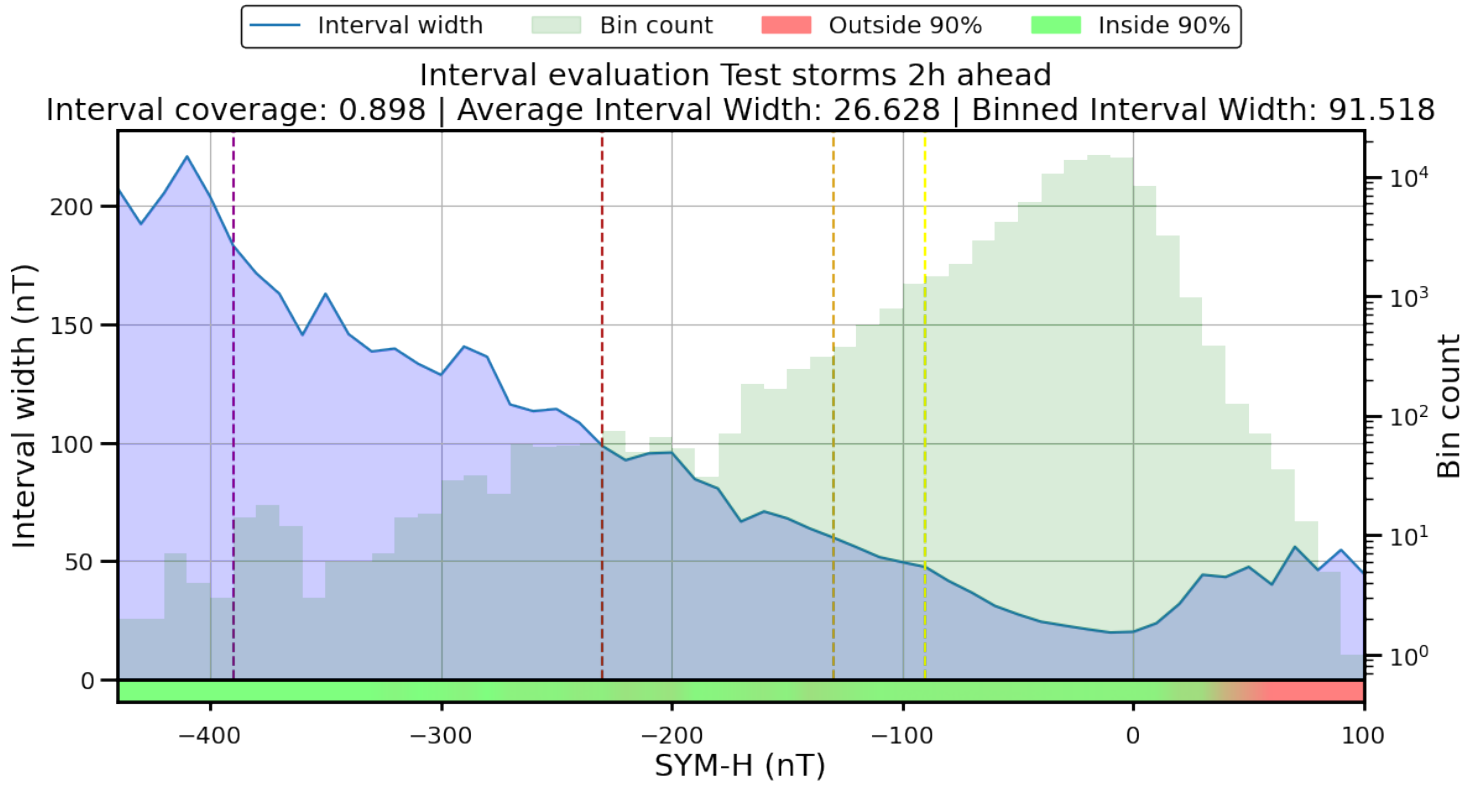
Mode	BFE	RMSE	R2
Mean	14.014	9.817	0.857
Global	27.408	10.559	0.917

Mode	PICP	PIAW	PIBW
Mean	0.900	26.366	44.196
Global	0.898	26.628	91.518

Model development and architecture

Results test 2h – Interval width

- Bottom heatmap shows the percentage of forecast inside the interval



Mode	BFE	RMSE	R2
Mean	14.014	9.817	0.857
Global	27.408	10.559	0.917

Mode	PICP	PIAW	PIBW
Mean	0.900	26.366	44.196
Global	0.898	26.628	91.518

Agenda

Adaptation to local indices
Introduction

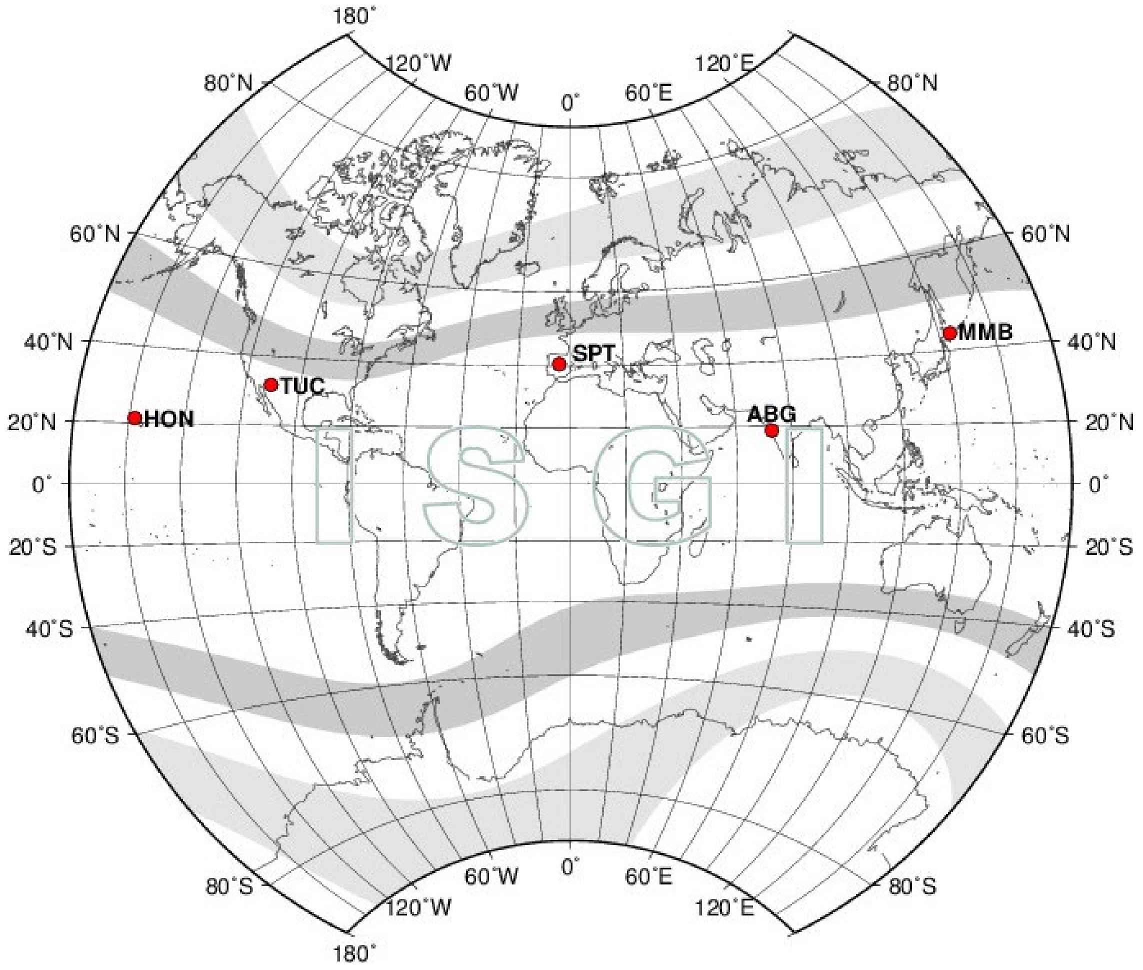
- Local Indices:
  - Measure geomagnetic activity at specific geographic locations, regional space weather effects
  - Derived from magnetometer data at individual ground-based stations
- Relevance of Local Indices:
  - Geomagnetic disturbances vary significantly across geographic regions due to local factors such as latitude, time of day, and Earth's magnetic field configuration
  - Local indices help assess space weather impacts on local infrastructure

Adaptation to local indices

LDi

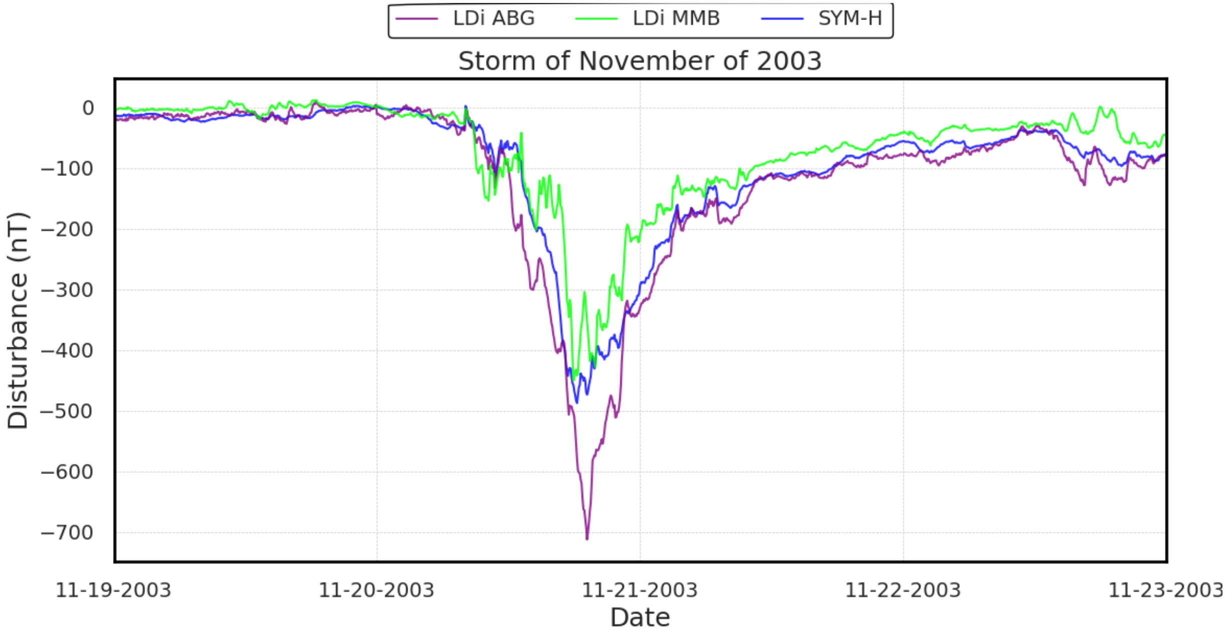
- We use the Local Disturbance Index
  - Real-time 1-minute resolution
  - Localized counterpart of the SYM-H
  - Derived from ground-based magnetometers

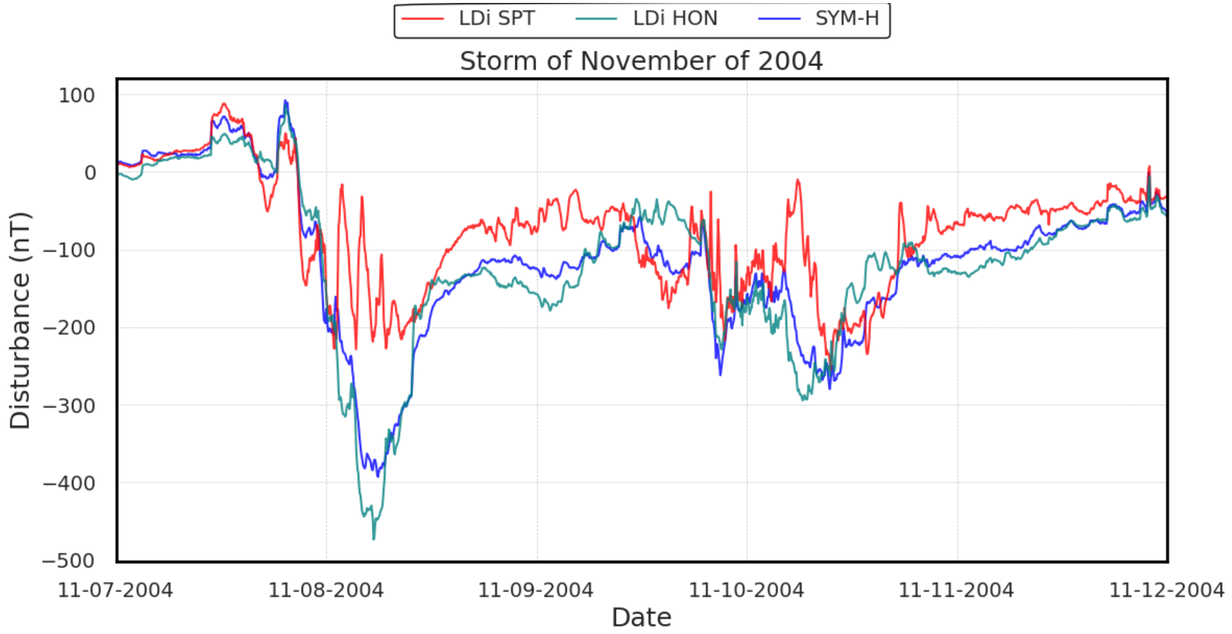
Observatory	Country	Lat	Long
SPT	Spain	39.55	-4.35
TUC	USA	32.17	-110.729
HON	USA	21.32	-158
ABG	India	18.638	72.872
MMB	Japan	43.91	144.19



Adaptation to local indices
LDi

- Comparison of the LDi of two stations in different longitudes to the SYM-H (blue) in two super intense storms

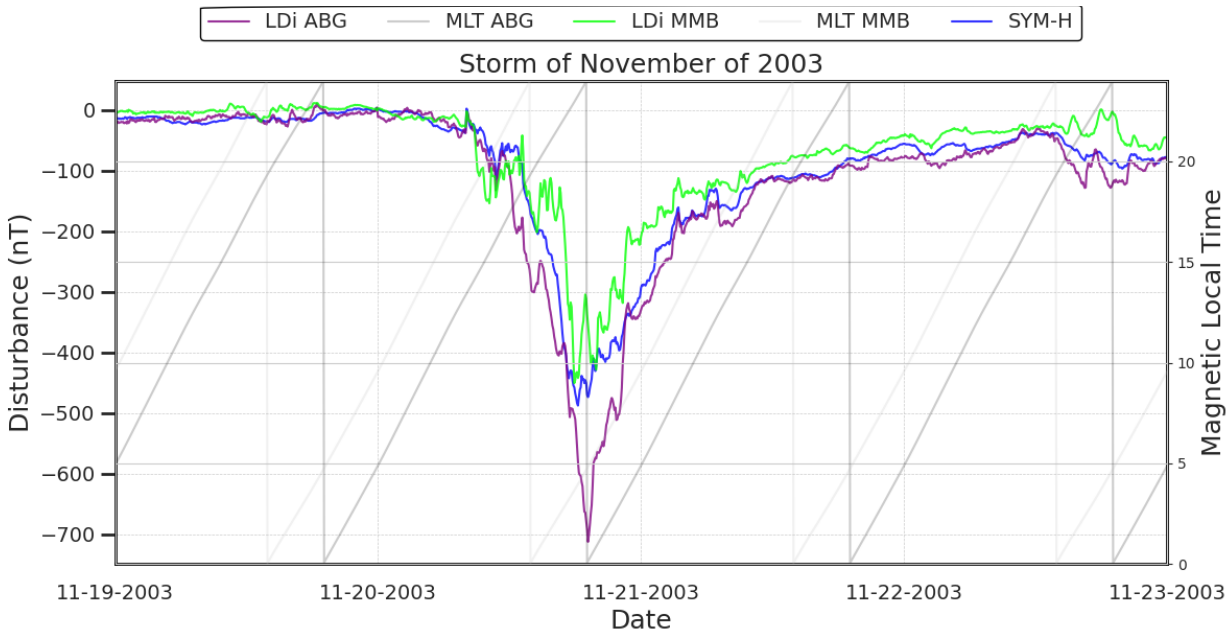


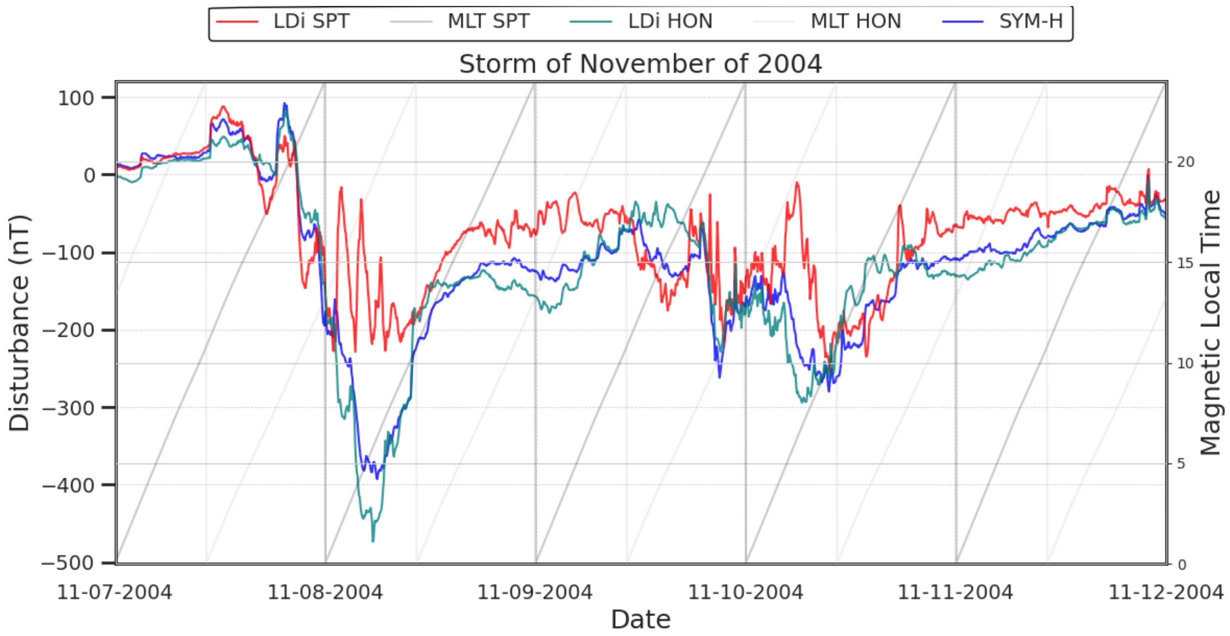


Adaptation to local indices

LDi

- We know the rotation plays an important role
- Add Magnetic Local Time (MLT)

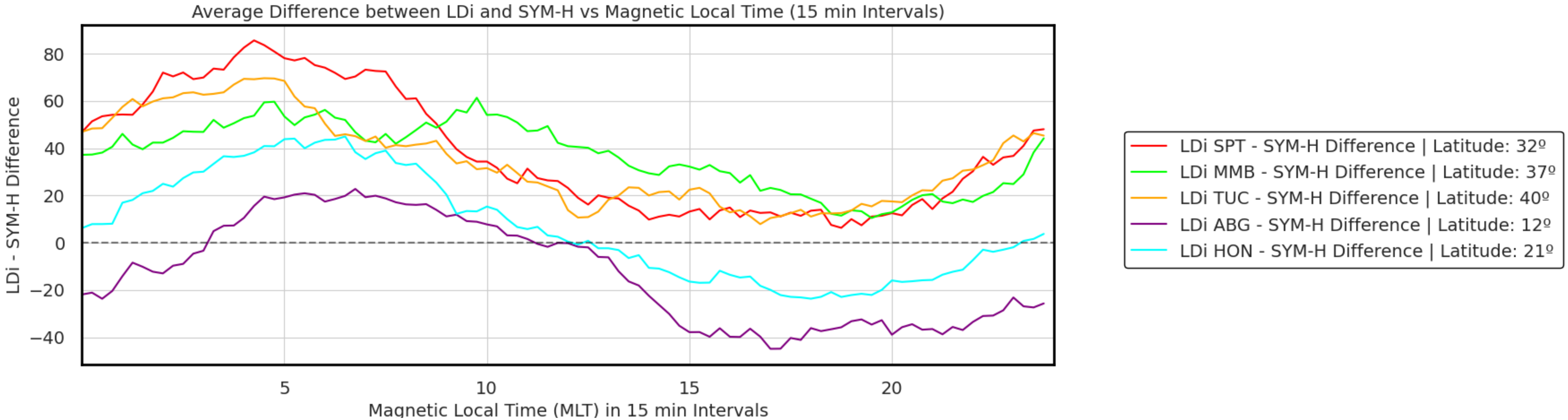




Adaptation to local indices

LDi

- Align the difference between the LDi and the SYM-H of each station against the MLT
- Positive difference at dawn for high latitude, negative at dusk for lower
- Latitude also influences in the LDi



Adaptation to local indices

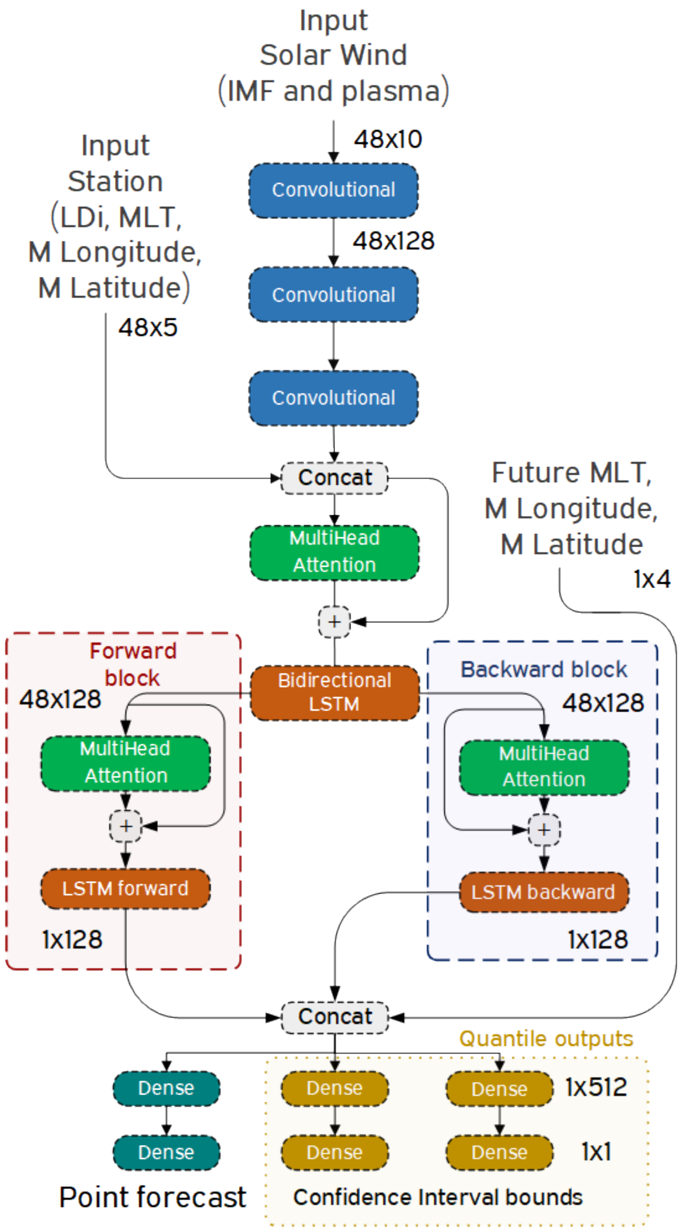
Neural Network

- We can either do a different model for each station or a single model for all of them
- Multiple models faces the problem of data dilution:
  - Further division on whether the storm reached at different MLTs
  - Harder to generalize
- Train a single model with the five stations:
  - More complex
  - Same disturbance measured by ACE results in different LDi behavior
  - Reuse sets for the SYM-H

Adaptation to local indices

Neural Network

- Separate the Solar Wind data from the LDi and station information
- Consider MLT, LDi, Magnetic Longitude and Latitude for each station
- Add as extra inputs the MLT, Longitude and Latitude of the forecast
- Incorporate Quantile forecasts
- Forecast 2 hours ahead



Adaptation to local indices

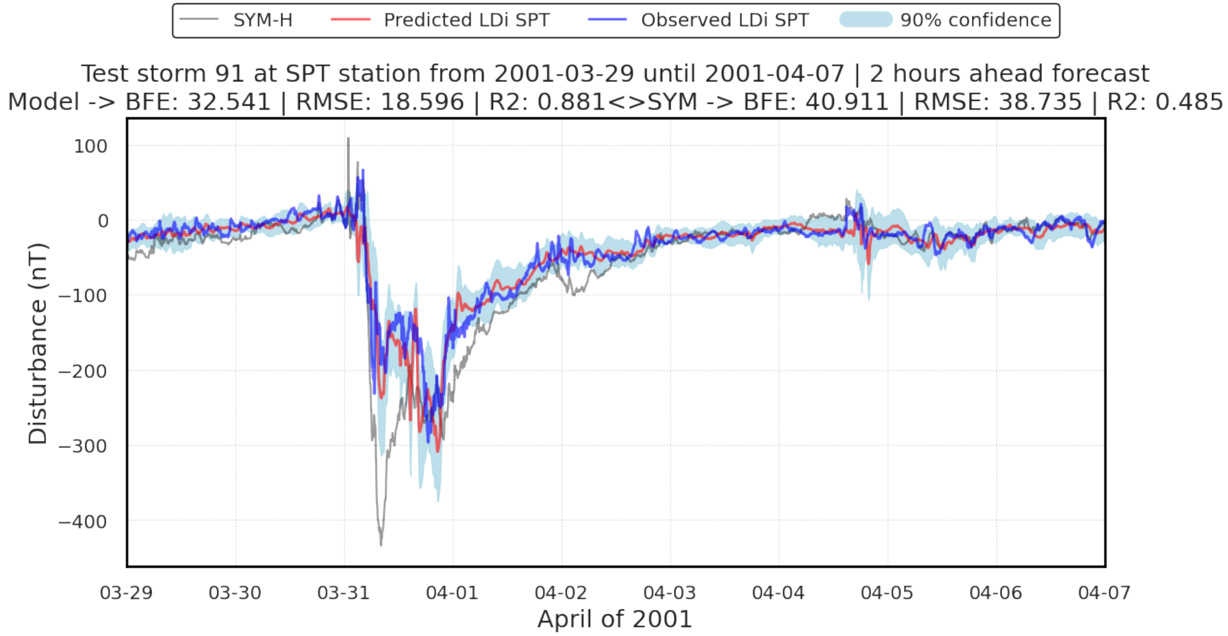
Results San-Pablo Toledo – Test storms 2 hours

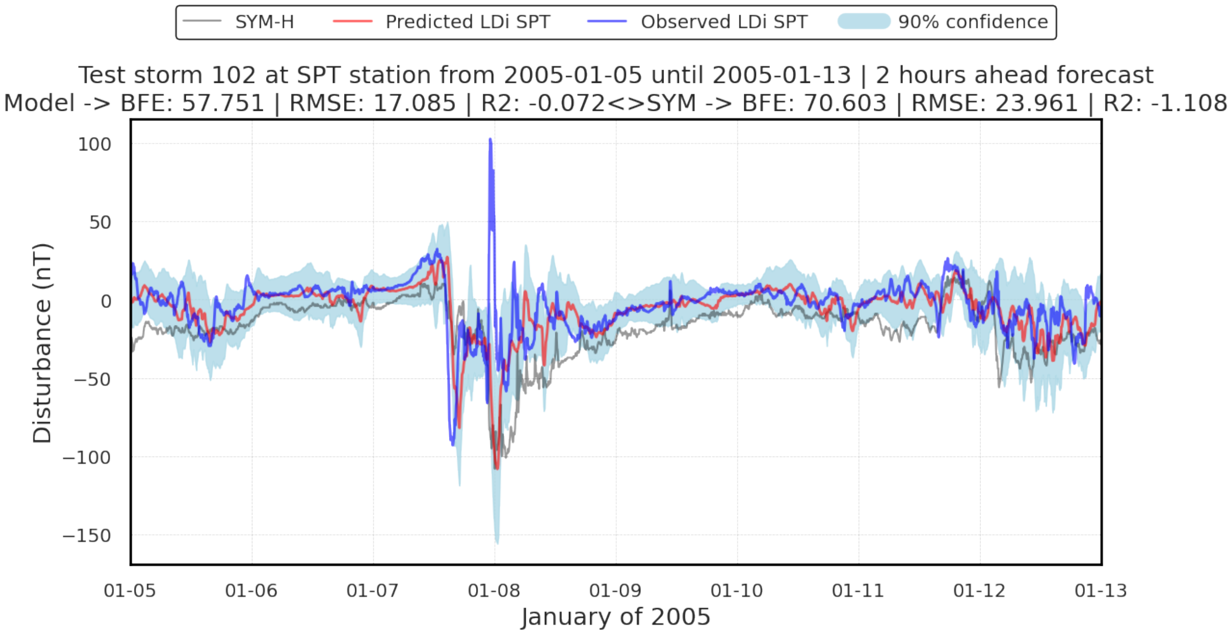
Globally superintense storm of 2 CMEs:

- First one is the most intense globally
- Second produces a higher disturbance only on SPT

Sometimes the spikes are positive in the local environment

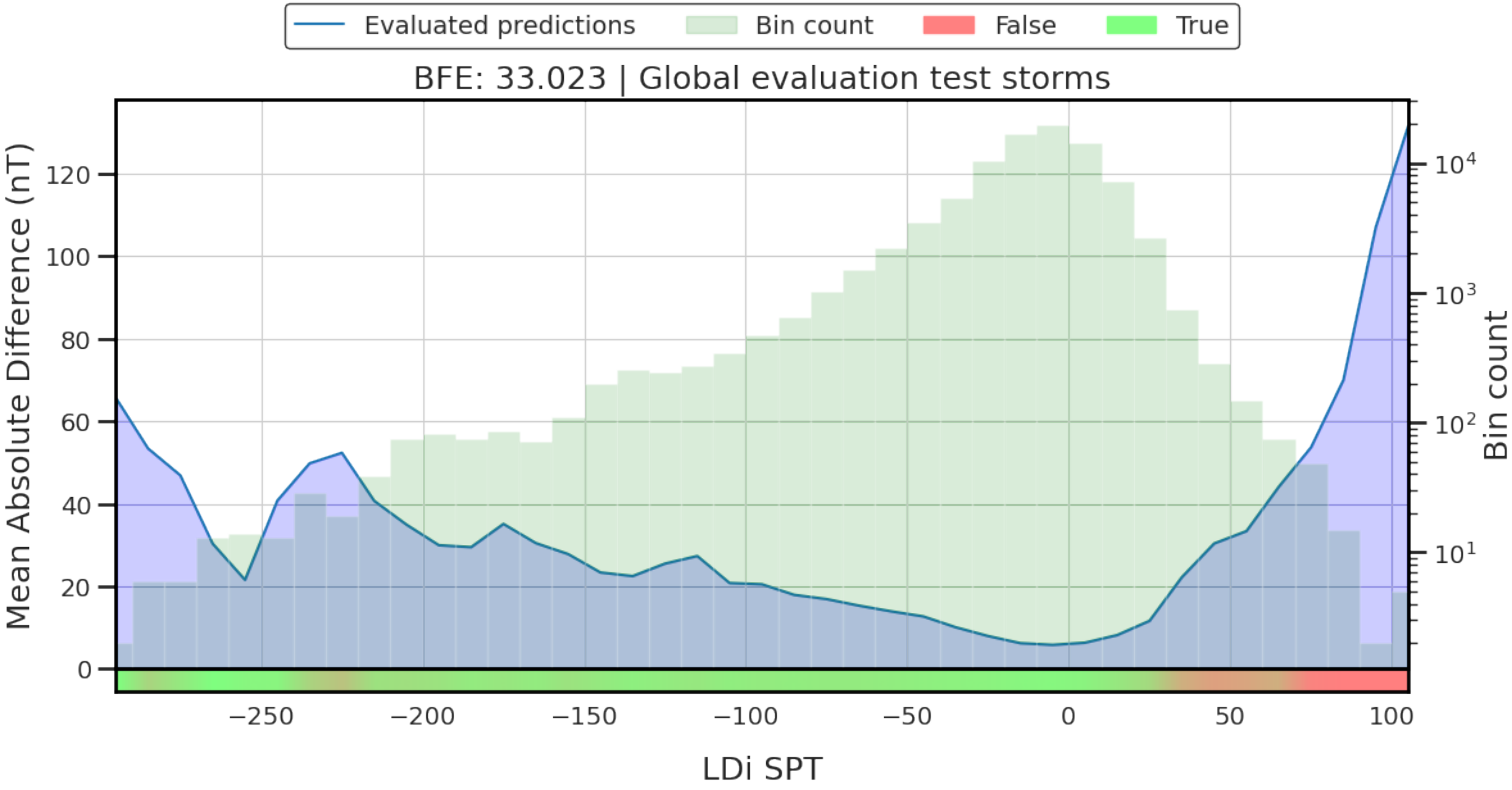
Network fails to forecast that case





Adaptation to local indices

Results San-Pablo Toledo – BFE Test storms 2 hours



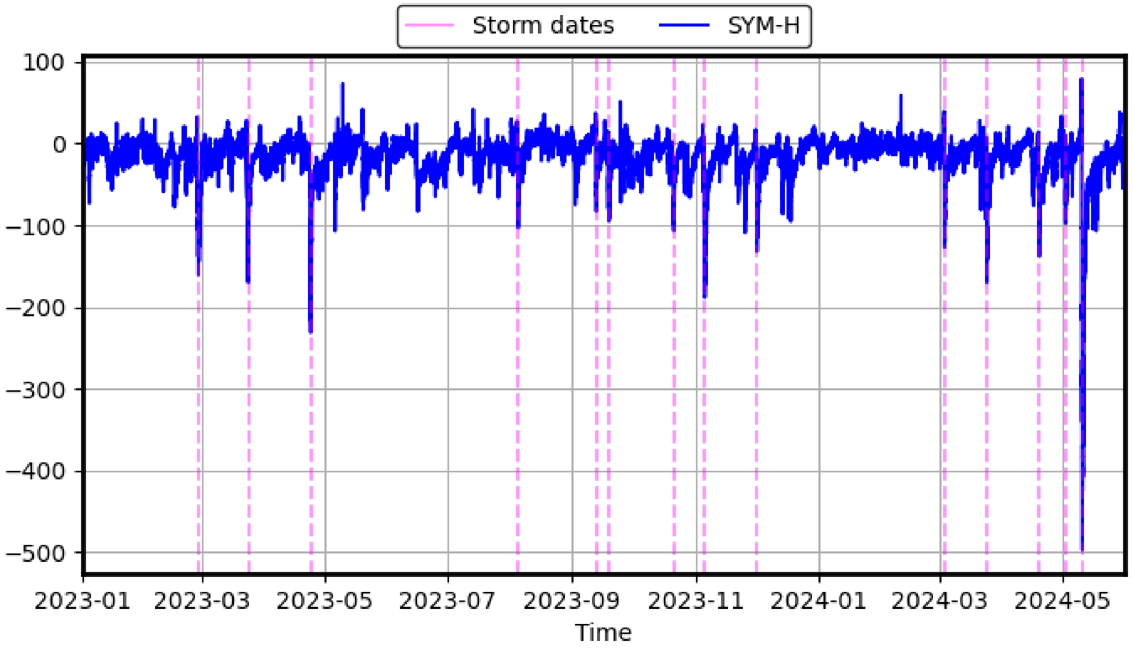
Model	BFE	RMSE	R2
Network	33.023	13.666	0.798
Persistence	42.545	17.295	0.677

PICP	PIAW	PIBW
0.861	36.262	45.755

Adaptation to local indices

Test on unseen stations – Coimbra

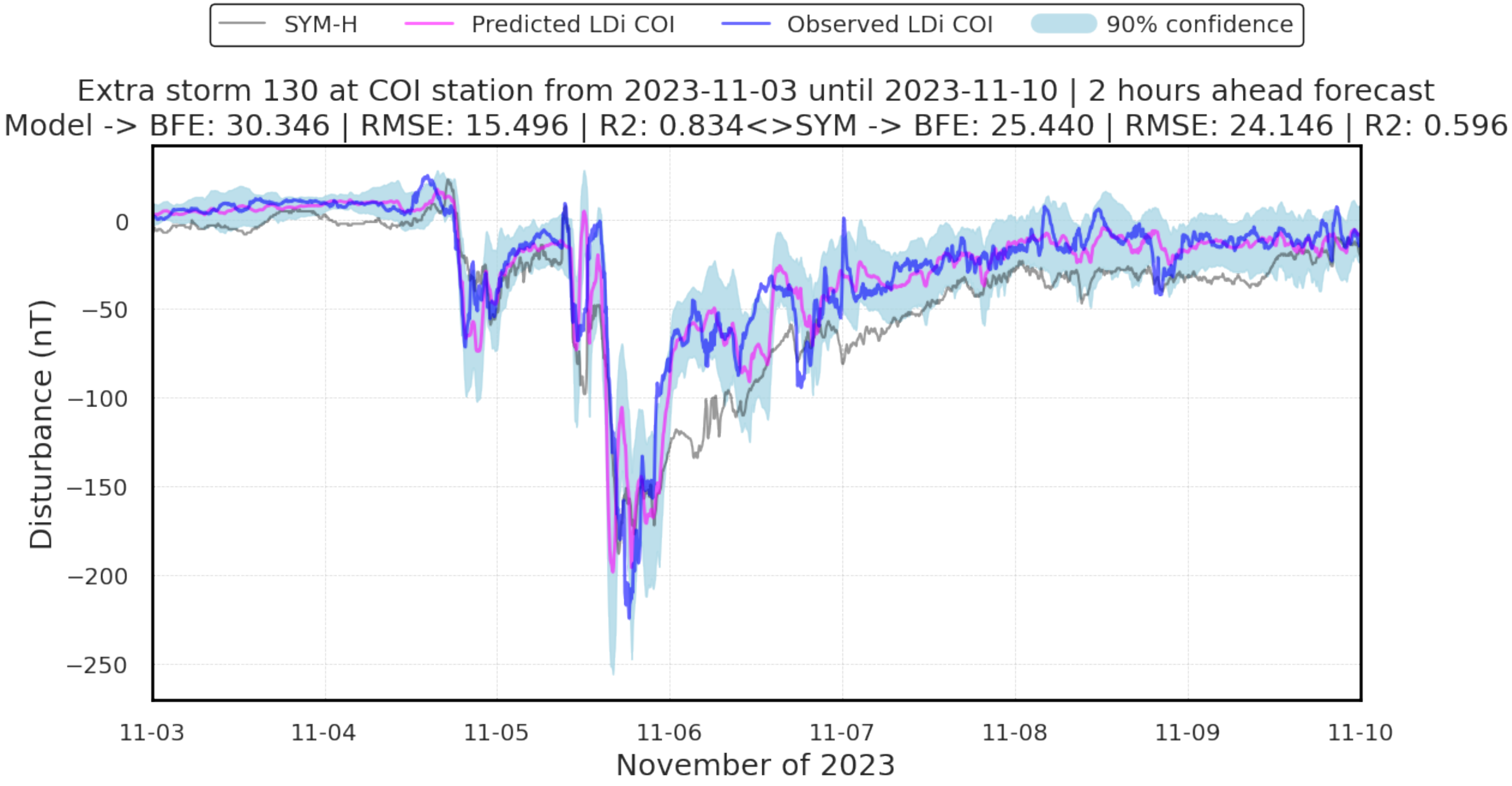
- Considering the input data, we can use the model on unseen stations
- We chose Coimbra, recent magnetometer, LDi from January 2023
- Evaluate on 13 storms using preliminary parameters
  - 1 Superintense
  - 1 Intense
  - 6 Moderate
  - 5 Low



Adaptation to local indices

Results Coimbra – Extra storms 2 hours

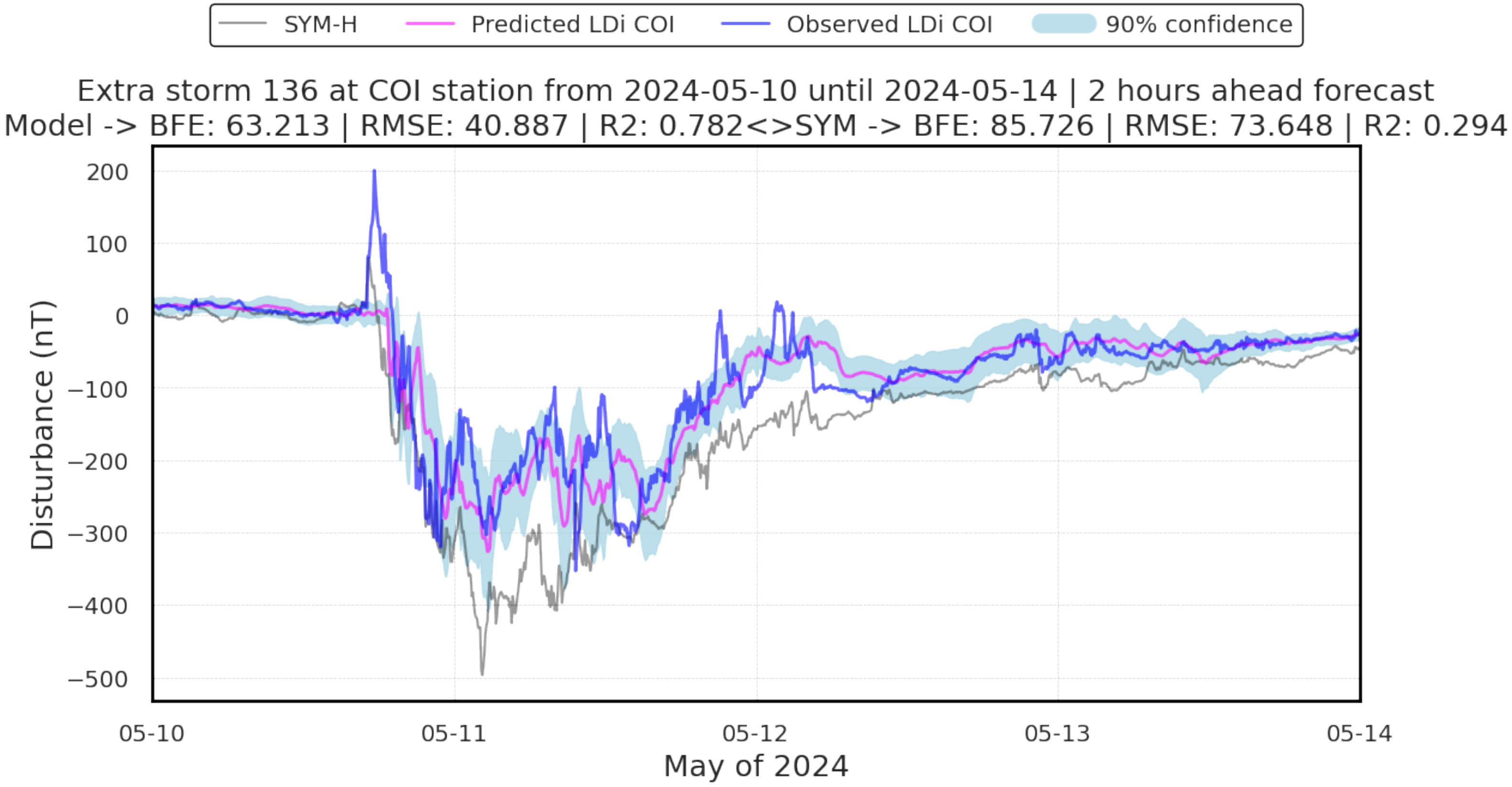
- For the intense storm of November 2023, the forecast is quite accurate



Adaptation to local indices

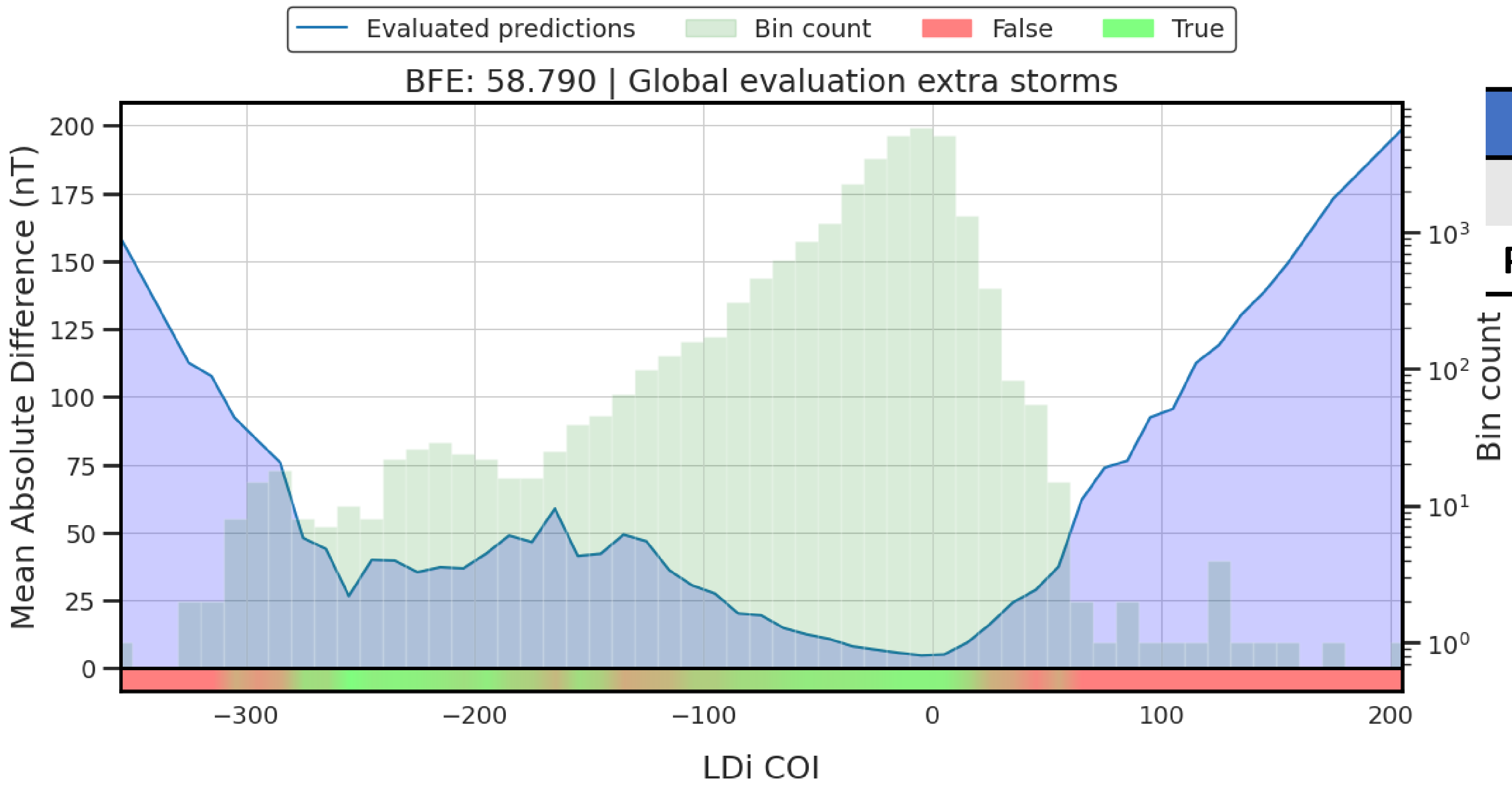
Results Coimbra – Extra storms 2 hours

- For the super intense storm of May 2024, the positive peak is not forecasted
- The negative peaks are again at different times than the global ones, but they are forecasted successfully



Adaptation to local indices

Results Coimbra – Evaluated storms 2 hours



Model	BFE	RMSE	R2
Network	58.790	14.718	0.812
Persistence	73.540	20.061	0.650

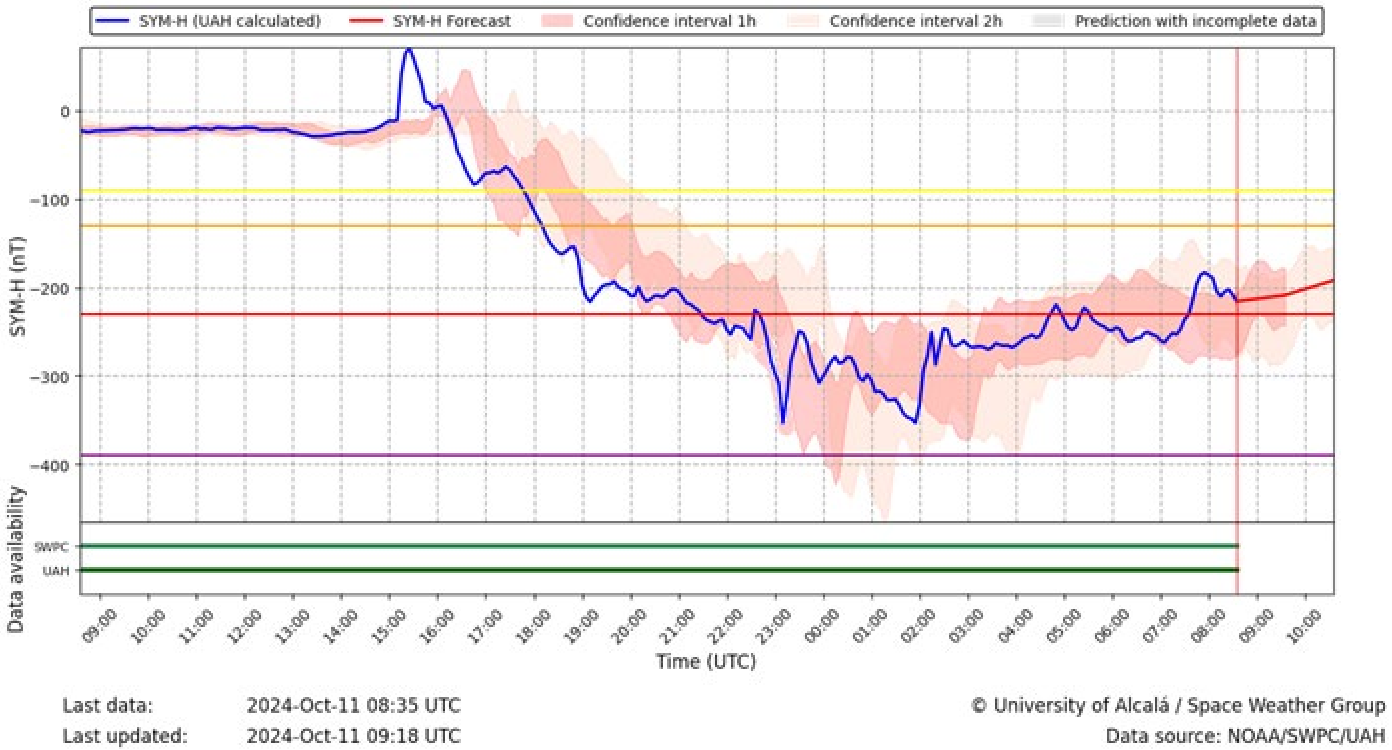
PICP	PIAW	PIBW
0.804	51.668	79.545

May storm skews the metric

Agenda

Operational case study: storm of May 2024

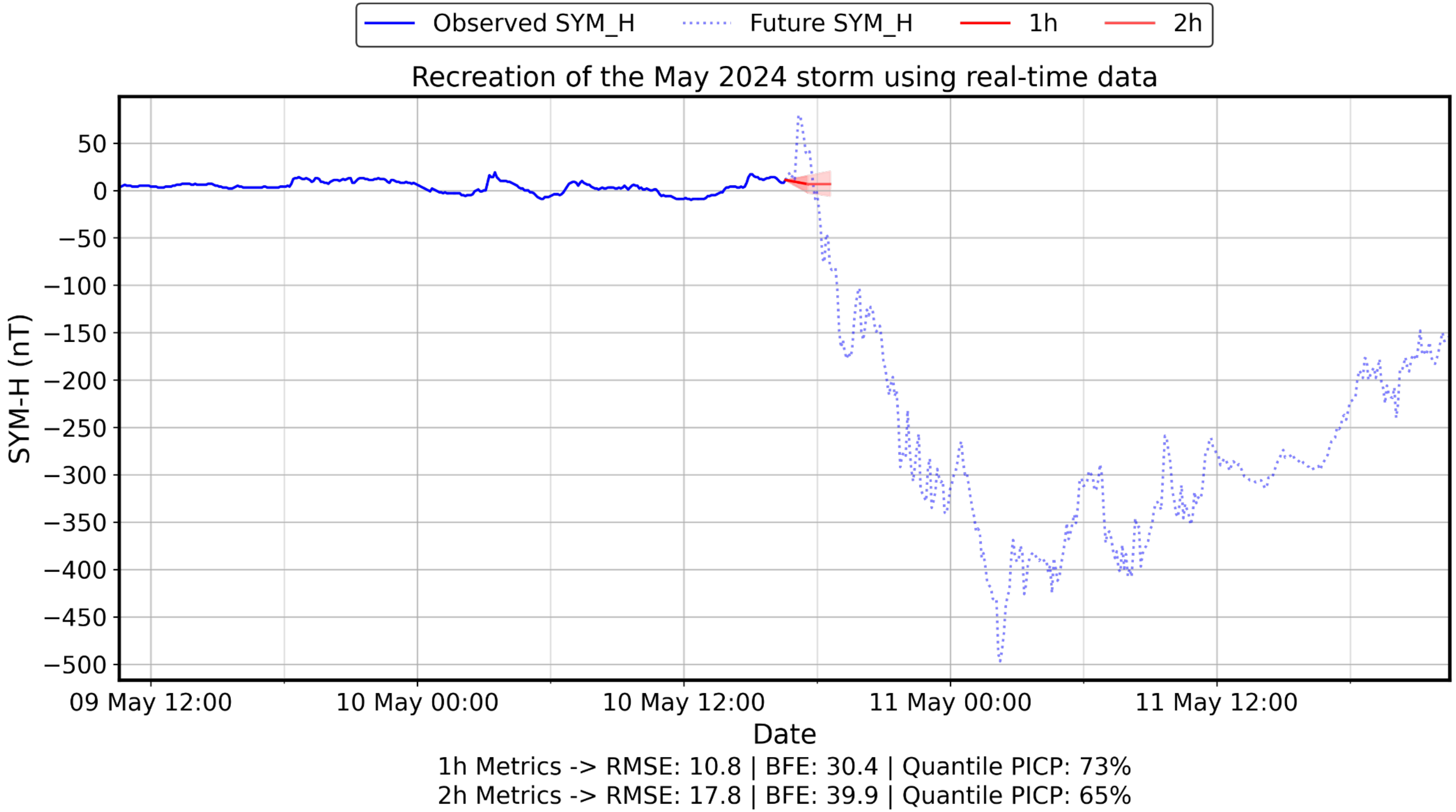
- Presented SYM-H forecast model has been running since March 2024
- Available at https://SeNMEs.es



May storm case study

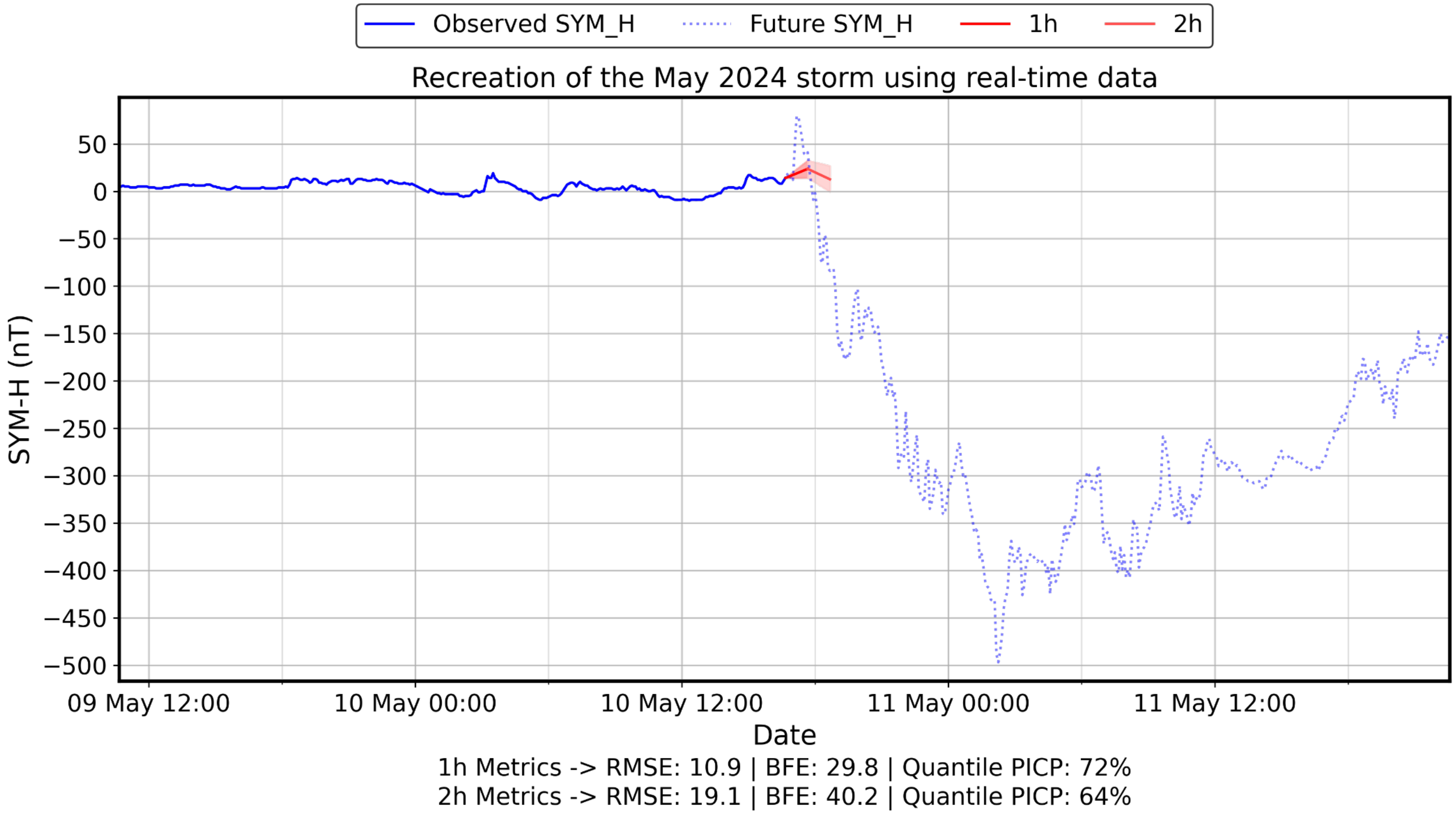
May storm case study

Relevant frames



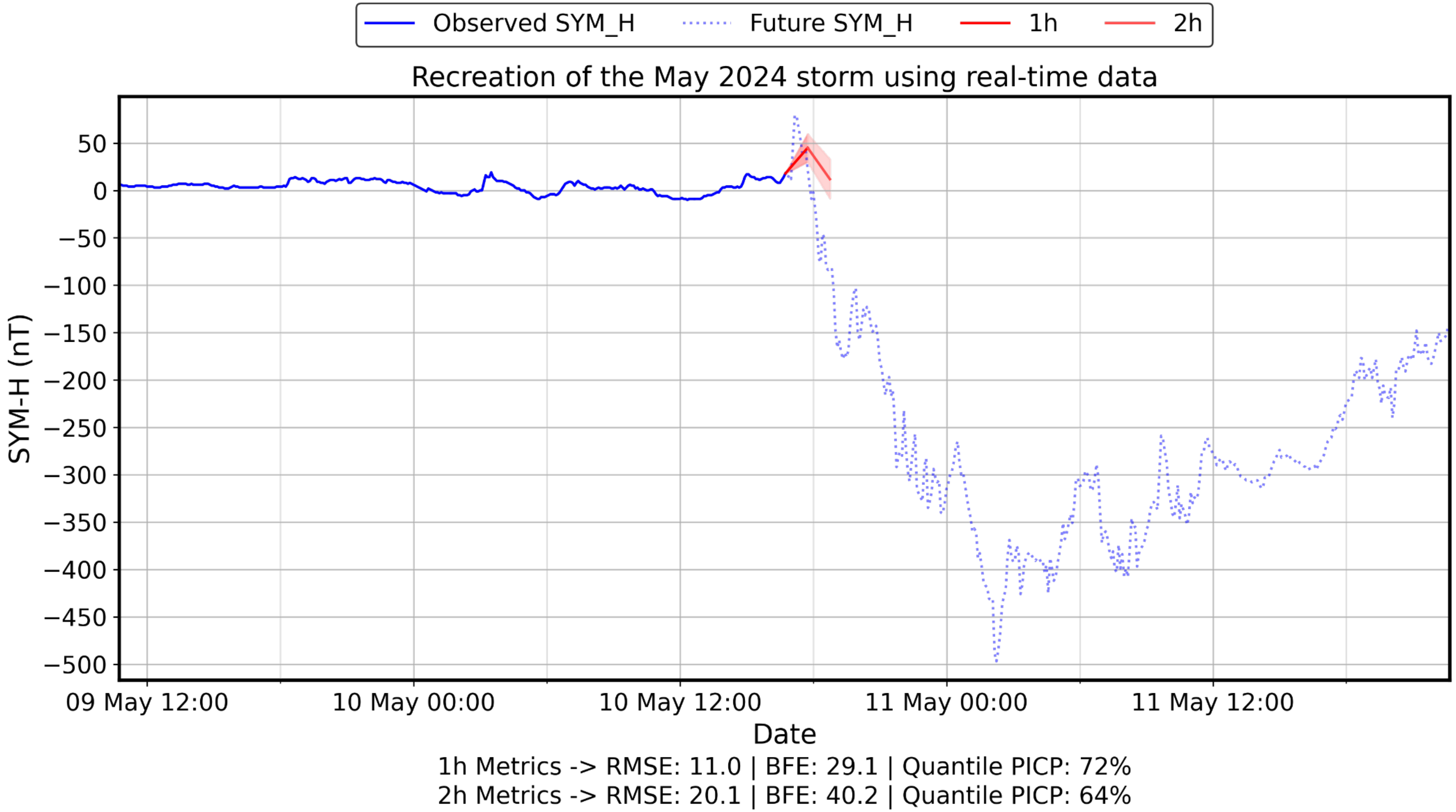
May storm case study

Relevant frames



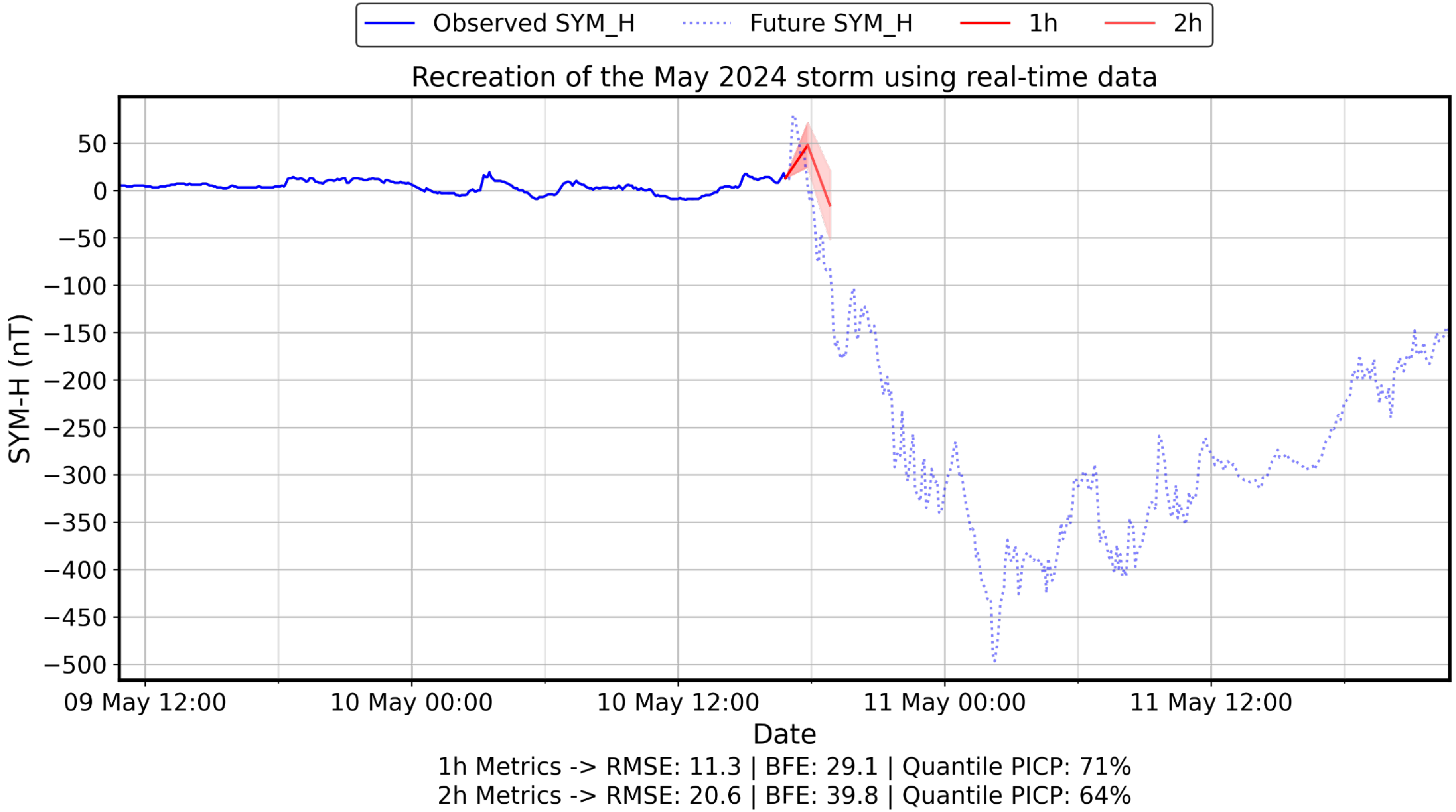
May storm case study

Relevant frames



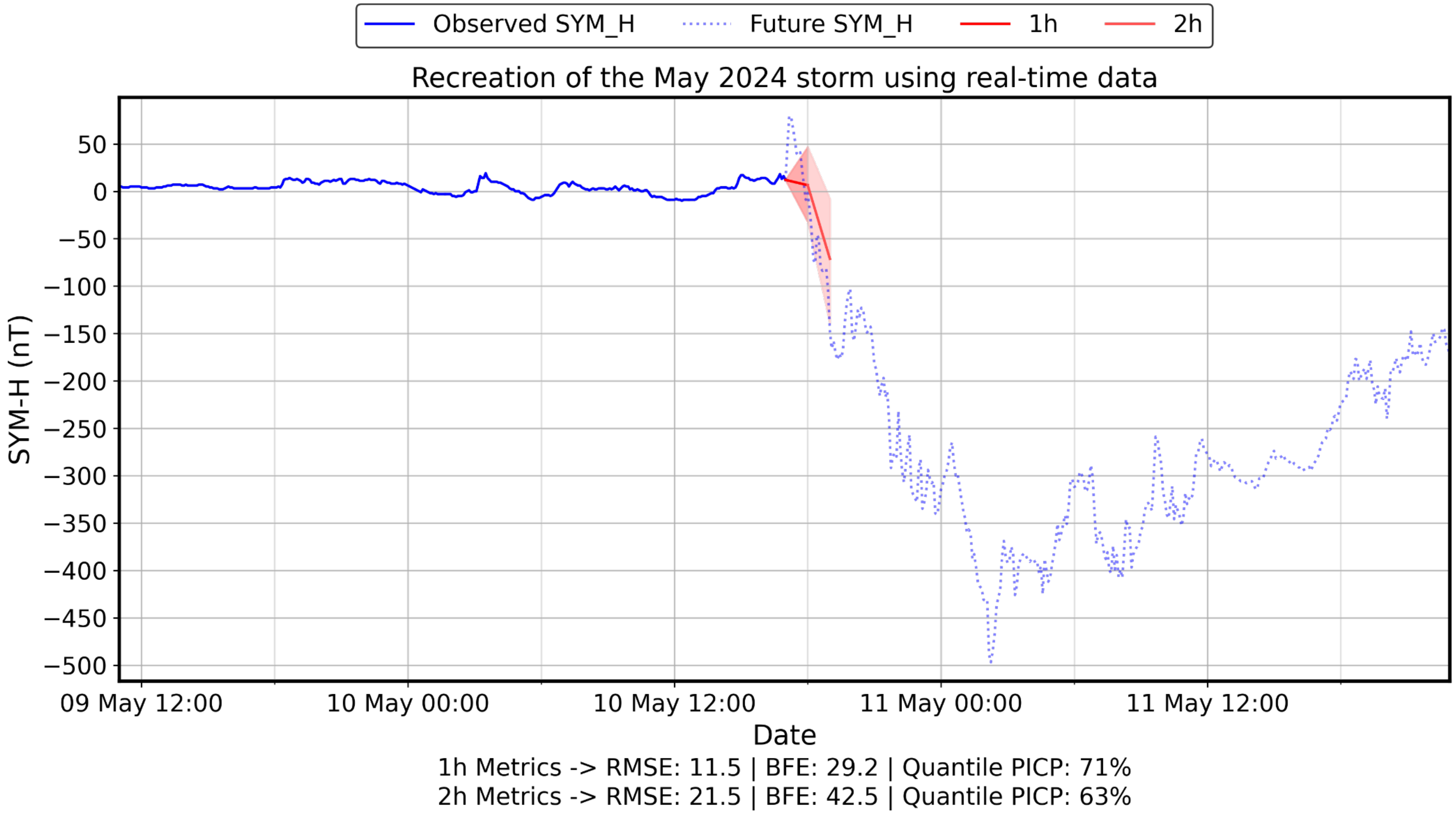
May storm case study

Relevant frames

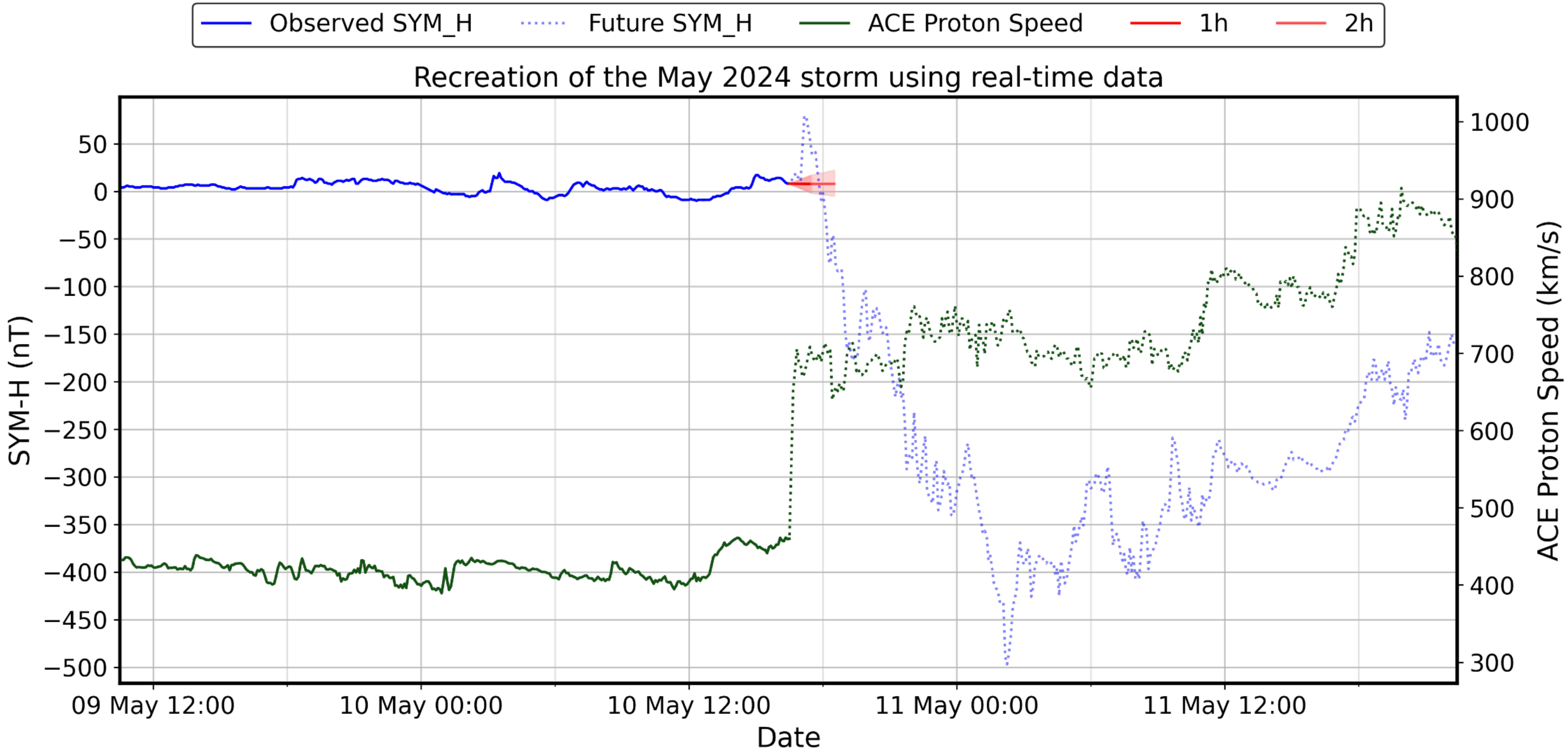


May storm case study

Relevant frames



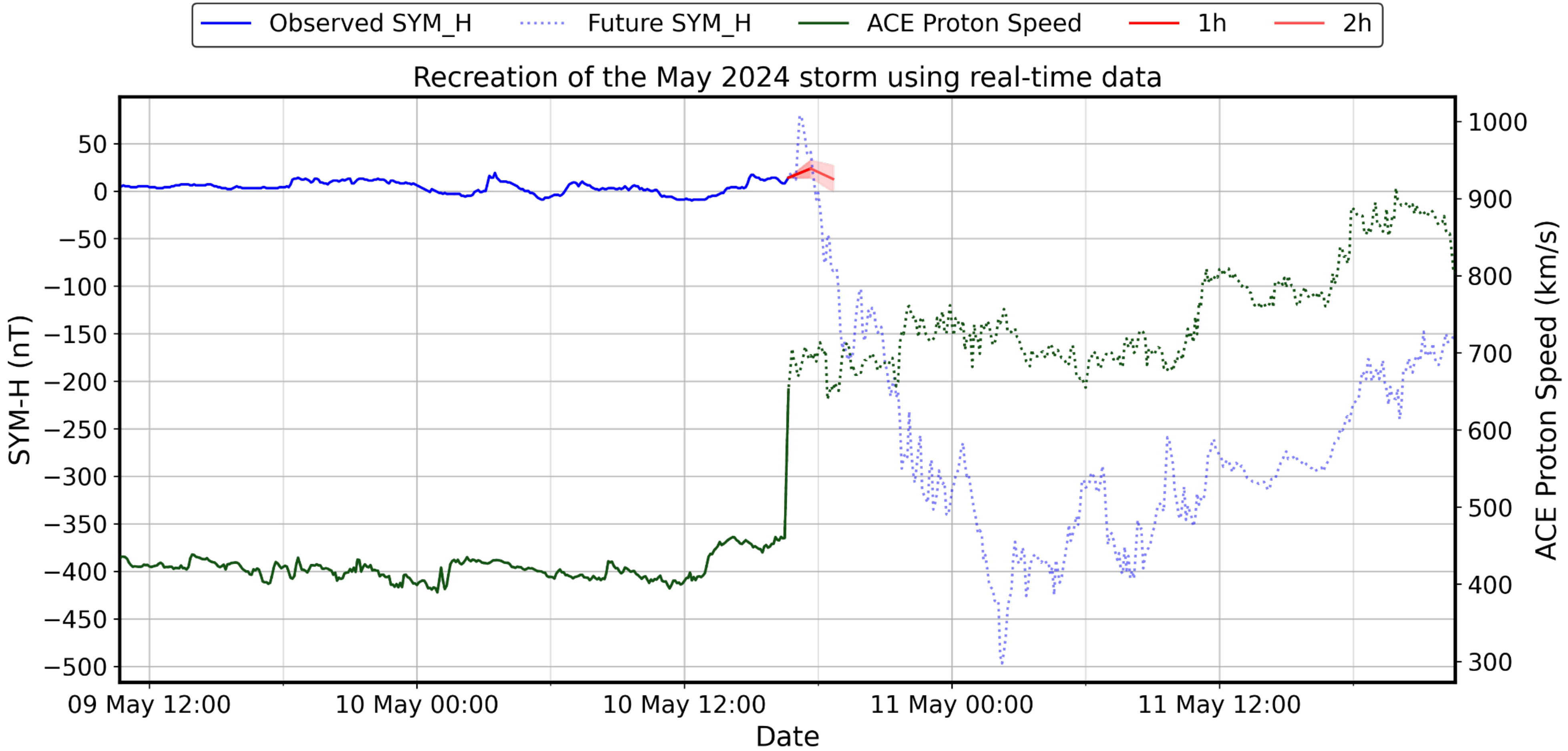
May storm case study
Relevant frames



1h Metrics -> RMSE: 10.6 | BFE: 30.2 | Quantile PICP: 73%
2h Metrics -> RMSE: 16.8 | BFE: 36.2 | Quantile PICP: 65%

May storm case study

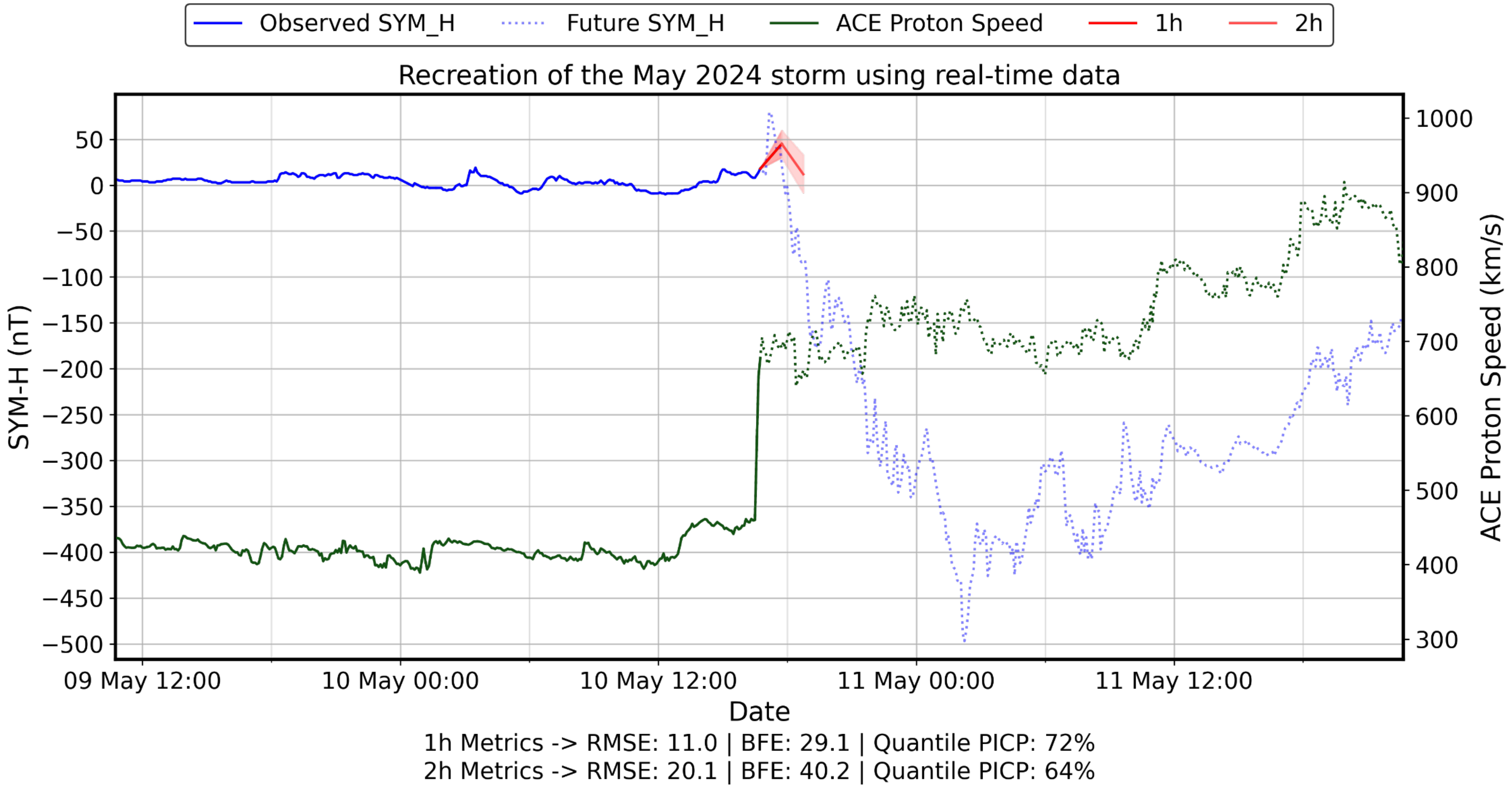
Relevant frames



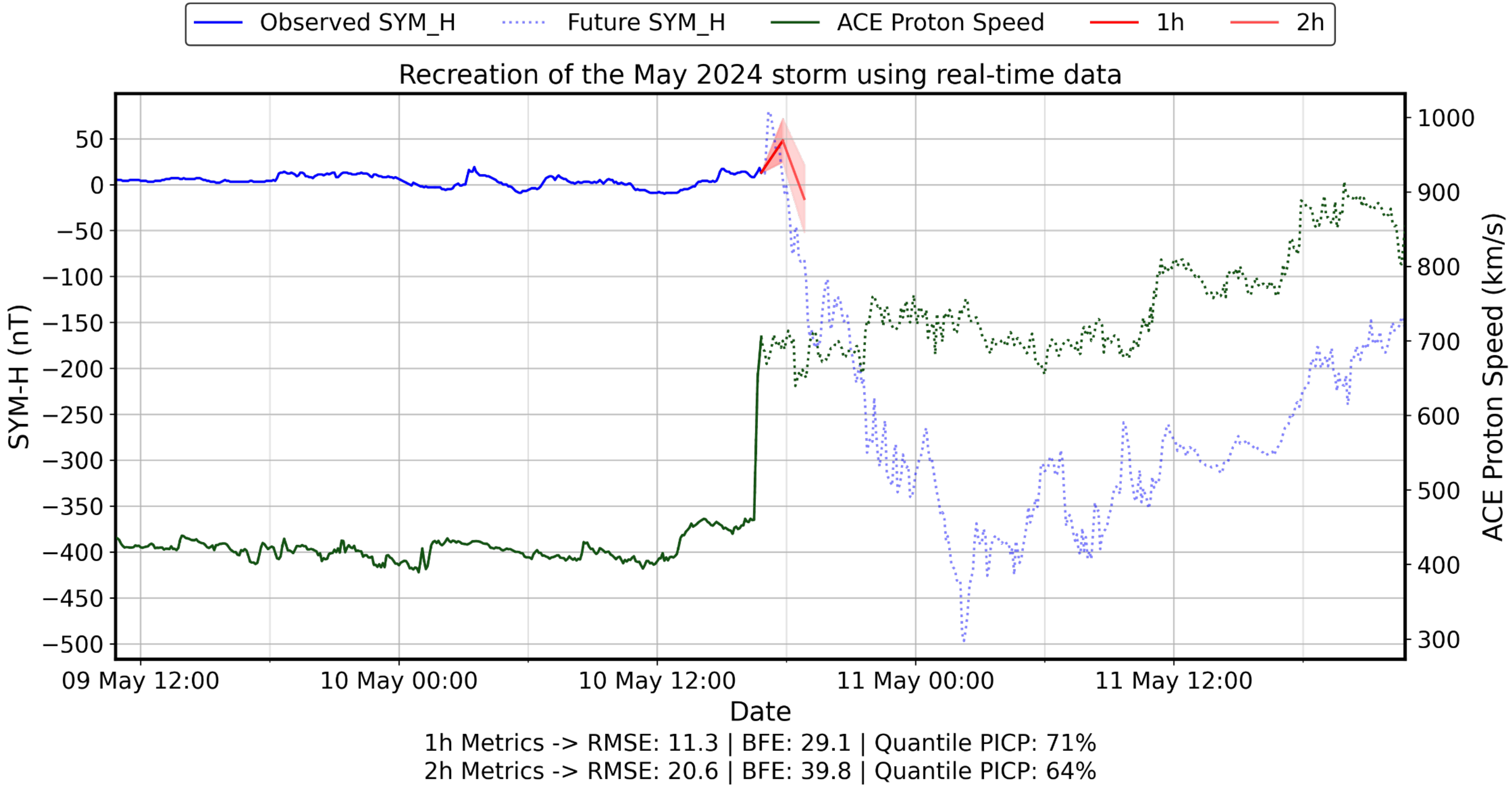
1h Metrics -> RMSE: 10.9 | BFE: 29.8 | Quantile PICP: 72%

2h Metrics -> RMSE: 19.1 | BFE: 40.2 | Quantile PICP: 64%

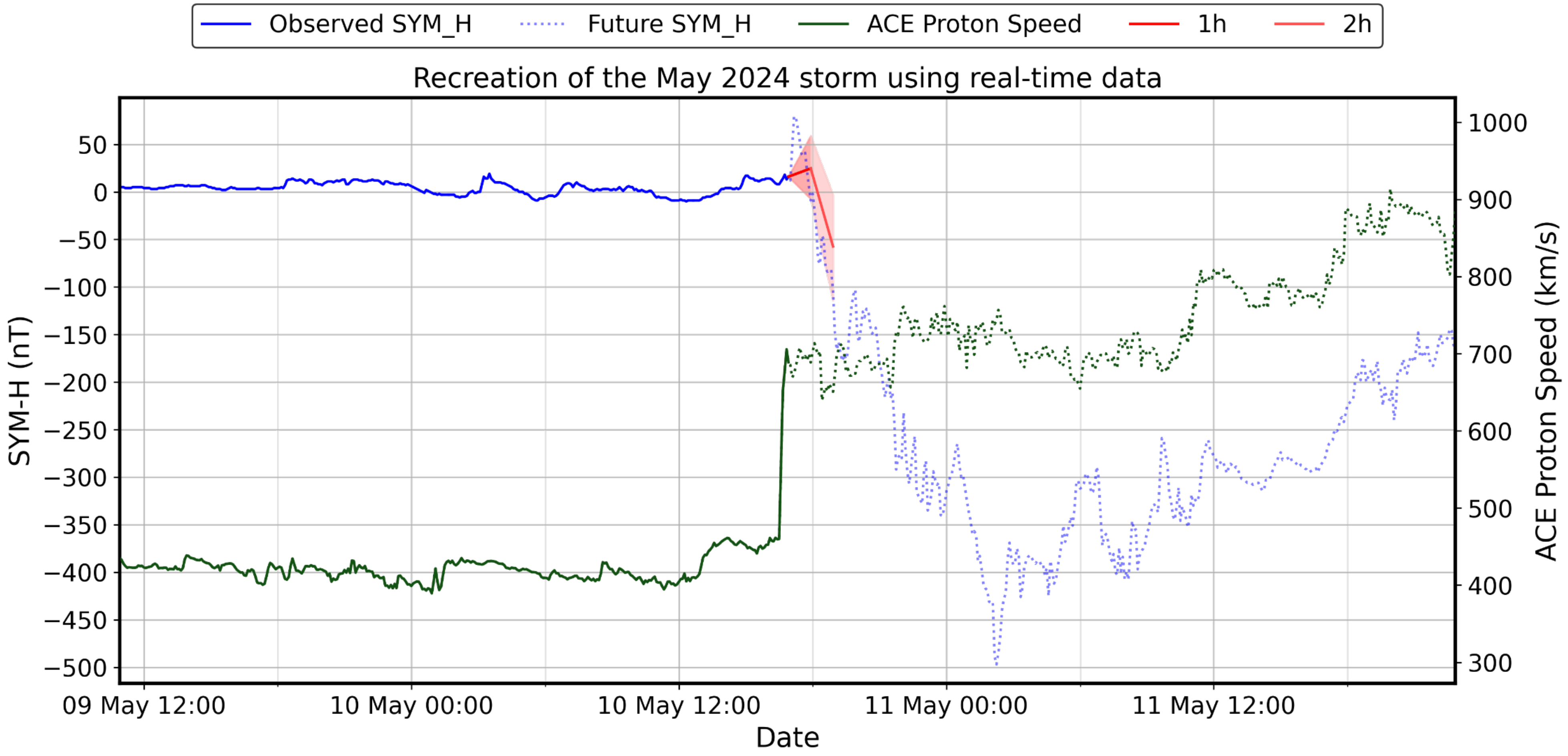
May storm case study
Relevant frames



May storm case study
Relevant frames

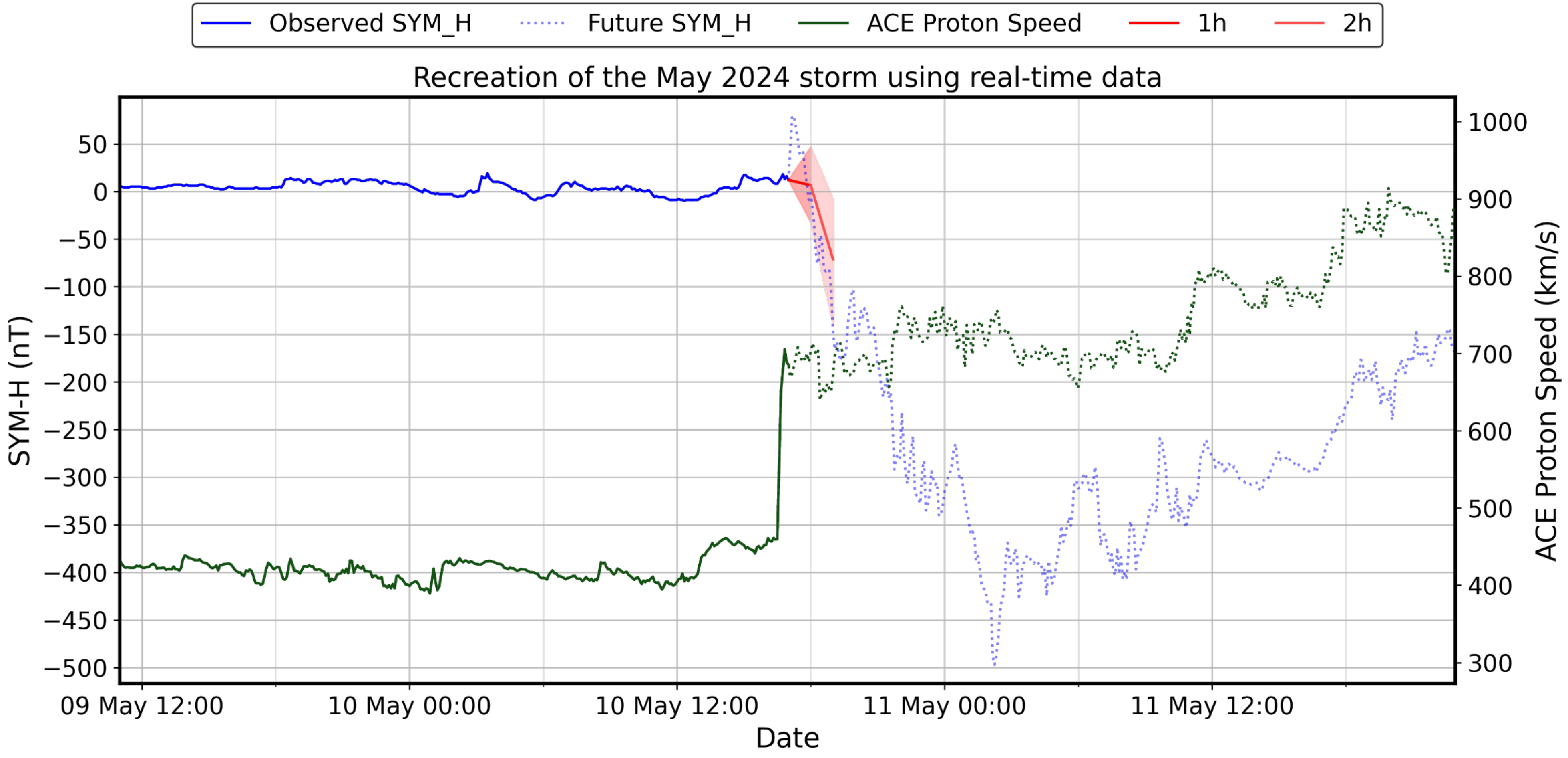


May storm case study
Relevant frames



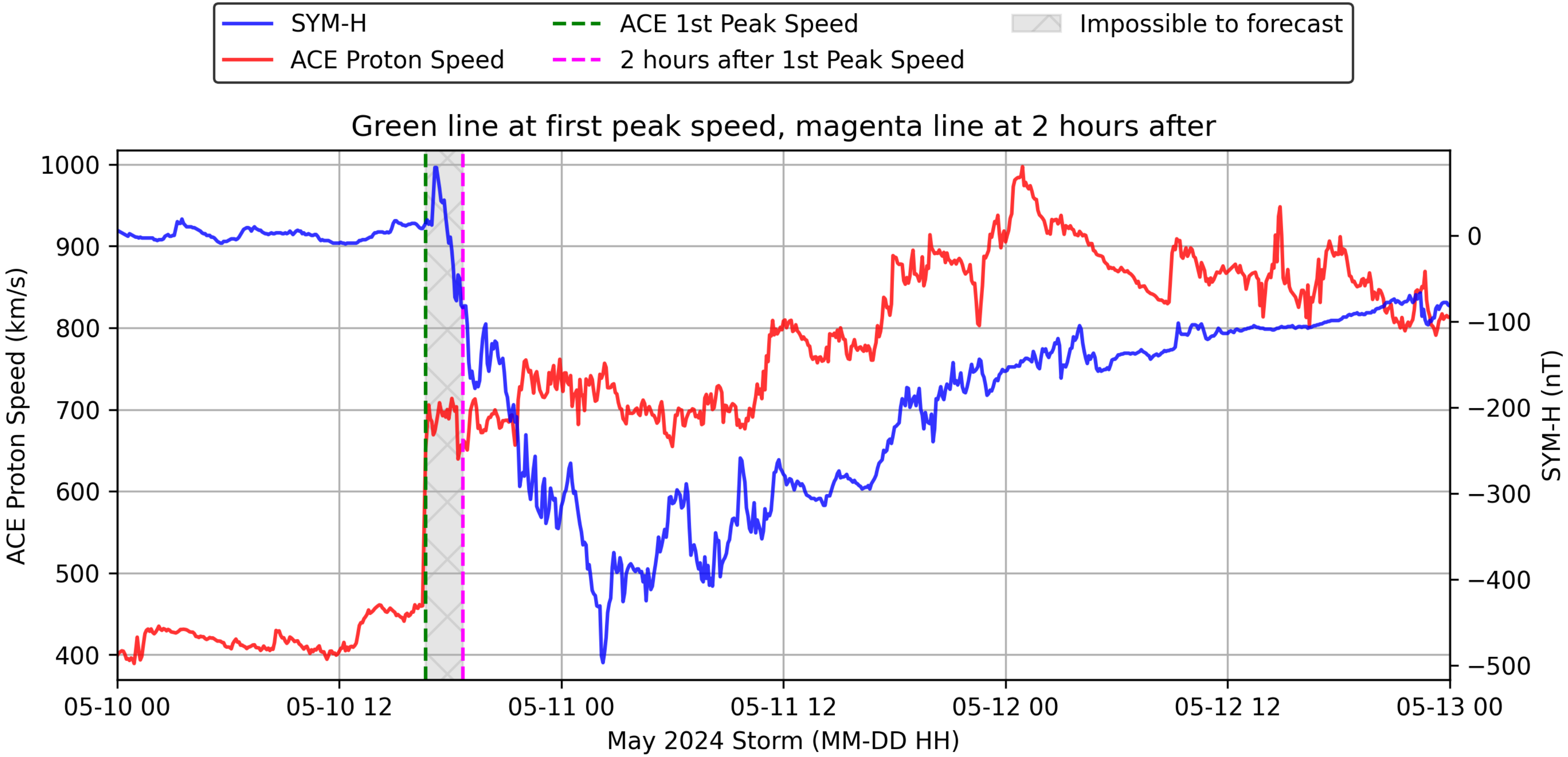
1h Metrics -> RMSE: 11.6 | BFE: 29.2 | Quantile PICP: 71%
2h Metrics -> RMSE: 20.8 | BFE: 40.2 | Quantile PICP: 64%

May storm case study
Relevant frames



1h Metrics -> RMSE: 11.5 | BFE: 29.2 | Quantile PICP: 71%
2h Metrics -> RMSE: 21.5 | BFE: 42.5 | Quantile PICP: 63%

May storm case study
Limits



Agenda

1. Introduction and context
2. Deep Learning fundamentals and its application to Space Weather
3. Training and evaluation framework
4. Model development and architecture
5. Adaptation to local indices
6. Operational case study: storm of May 2024
7. Conclusions and future work

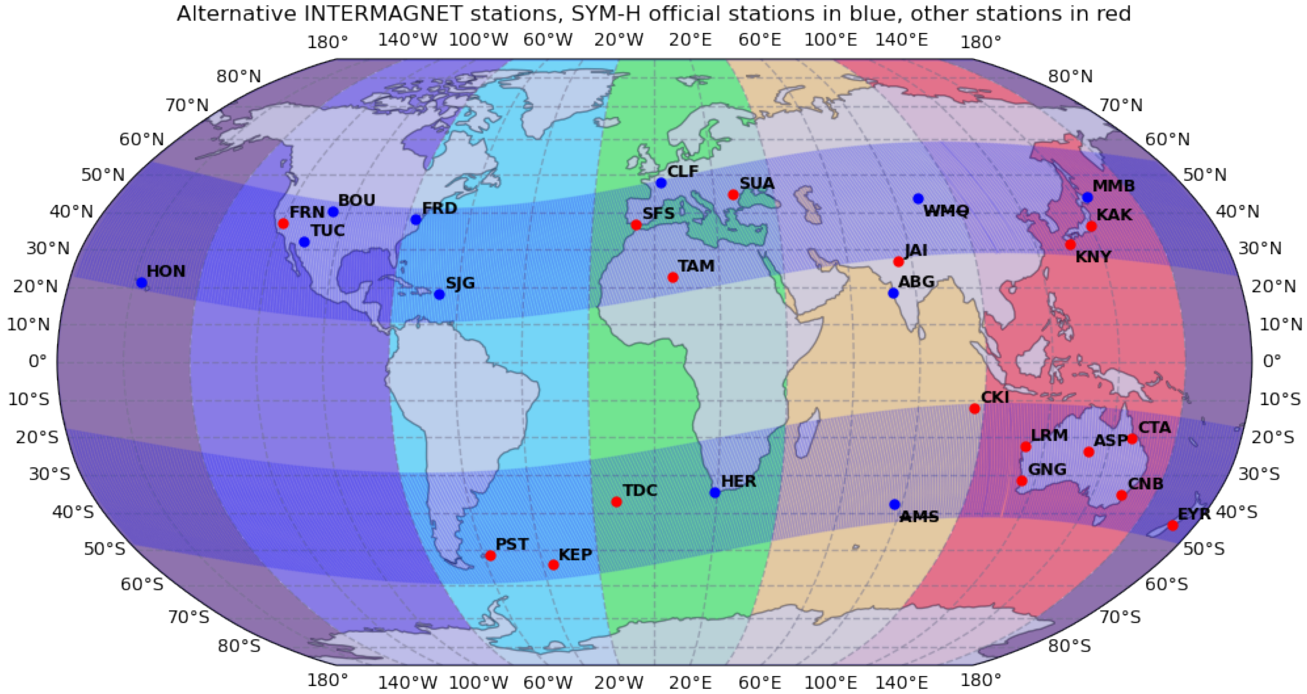
Conclusions

Global indices

- Expanding the dataset has a positive influence on the model
- Real-time data differs heavily from historical data, we lacks an established database, which is crucial for training models to ensure reliability under real-time conditions
- Adding quantile forecasts provides relevant information
- Uncertainty of the forecasts is useful in a real-world environment
  - Better approach over Monte-Carlo dropout, as is deterministic and only needs to be run once
  - Interval width is still large for the most extreme events

Conclusions
Local indices

- Local indices can be forecasted; we expect an improvement if we train with more stations
- Can be applied to new stations with reasonable success



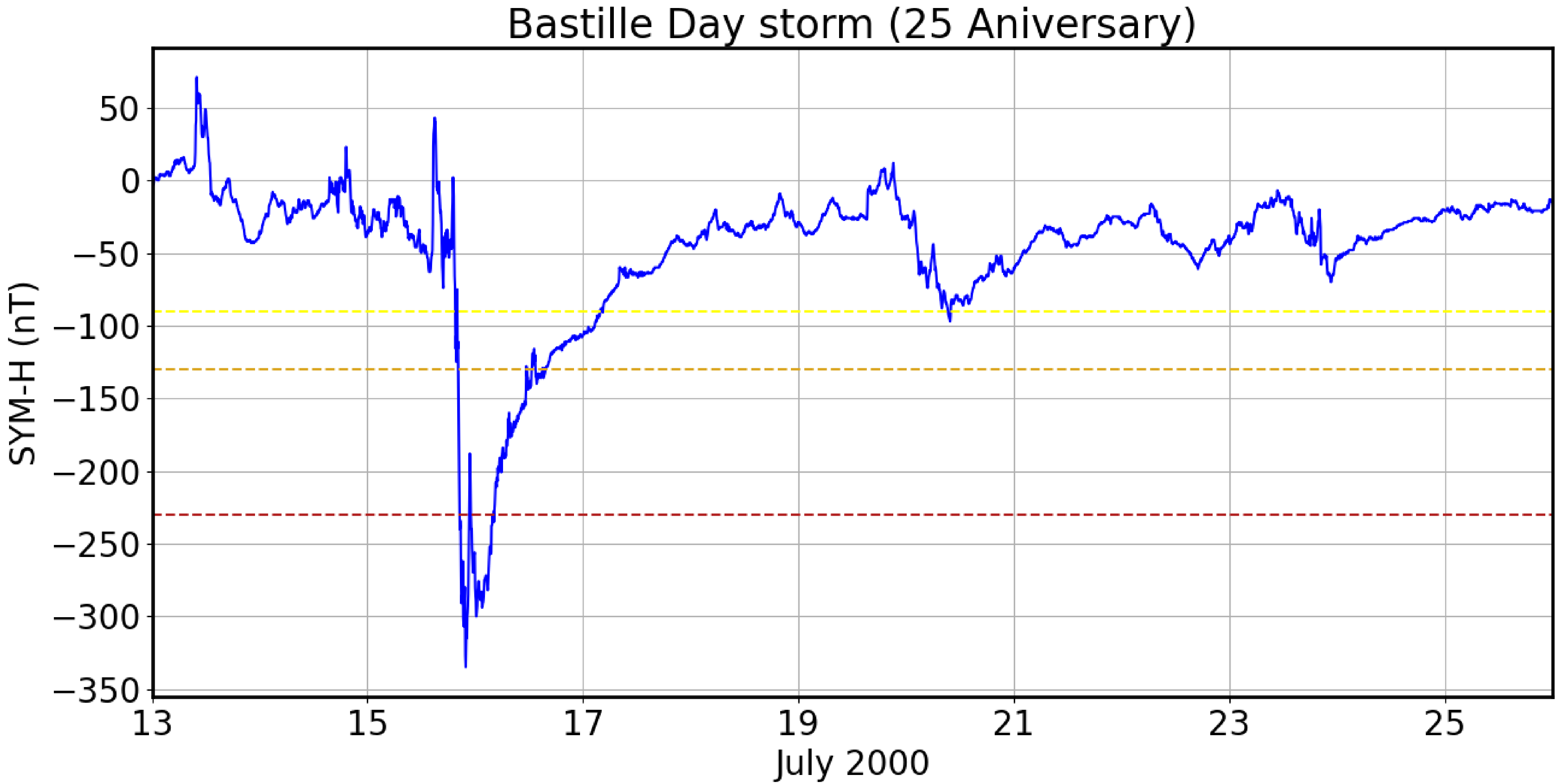
Future work

- Extend the forecast horizon
  - Limited by the time it takes for the disturbances to reach Earth from ACE
  - Accuracy reduces rapidly when increasing lead time
  - We need to know the acceptable forecasting accuracy
- Expand to other indices
  - Hp30 or Hp60
- Training the network with more local stations
  - Increase performance by training with stations across longitudes
- Using solar images to increase lead time
  - Historical data from SOHO, prepare for extra information from Vigil

Publications

High impact journals

1. Collado-Villaverde et al. Deep Neural Networks with Convolutional and LSTM layers for SYM-H and ASY-H forecasting. *Space Weather.* June 2021. DOI: 10.1029/2021SW002748
2. Collado-Villaverde et al. Neural Networks for operational SYM-H forecasting using attention and SWICS plasma features. *Space Weather.* August 2023. DOI: 10.1029/2023SW003485
3. Collado-Villaverde et al. Classifying and bounding geomagnetic storms based on the SYM-H and ASY-H indices. *Natural Hazards.* October 2023. DOI: 10.1007/s11069-023-06241-1
4. Collado-Villaverde et al. A Framework for Evaluating Geomagnetic Indices Forecasting Models. *Space Weather.* March 2024. DOI: 10.1029/2024SW003868
5. Collado-Villaverde et al. Comment on ``Prediction of the SYM-H Index Using a Bayesian Deep Learning Method With Uncertainty Quantification'' by Abduallah et al. (2024). *Space Weather.* March 2024. DOI: 10.1029/2024SW003909
6. Collado-Villaverde et al. Operational SYM-H forecasting with confidence intervals using Deep Neural Networks. *Space Weather.* October 2024. DOI: 10.1029/2024SW004039



Bastille Day storm (25 Aniversary)
50
0
−50
−100
−150
−200
−250
−300
−350
SYM-H (nT)
13
15
17
19
21
23
25
July 2000

Thanks for your attention

Armando Collado Villaverde {armando.collado@uah.es}
July, 2025