V Meeting on Astrophysical Spectroscopy-
September 12th – 15th, Palić, Republic of Serbia

Alternative Evaluation Metrics for Machine Learning Model Selection in Ionospheric VLF Amplitude Data Exclusion

Filip Arnaut, Aleksandra Kolarski
Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
E-mail: filip.arnaut@ipb.ac.rs

Introduction

- When applying machine learning (ML) methods to classify ionospheric VLF amplitude data for data exclusion, there is a significant imbalance in the ML task (85- 15% normal vs. anomalous data class).

- **Research problem:** Standard evaluation metrics for imbalanced ML tasks can yield subpar results, requiring careful interpretation in relation to the distribution of the test dataset.

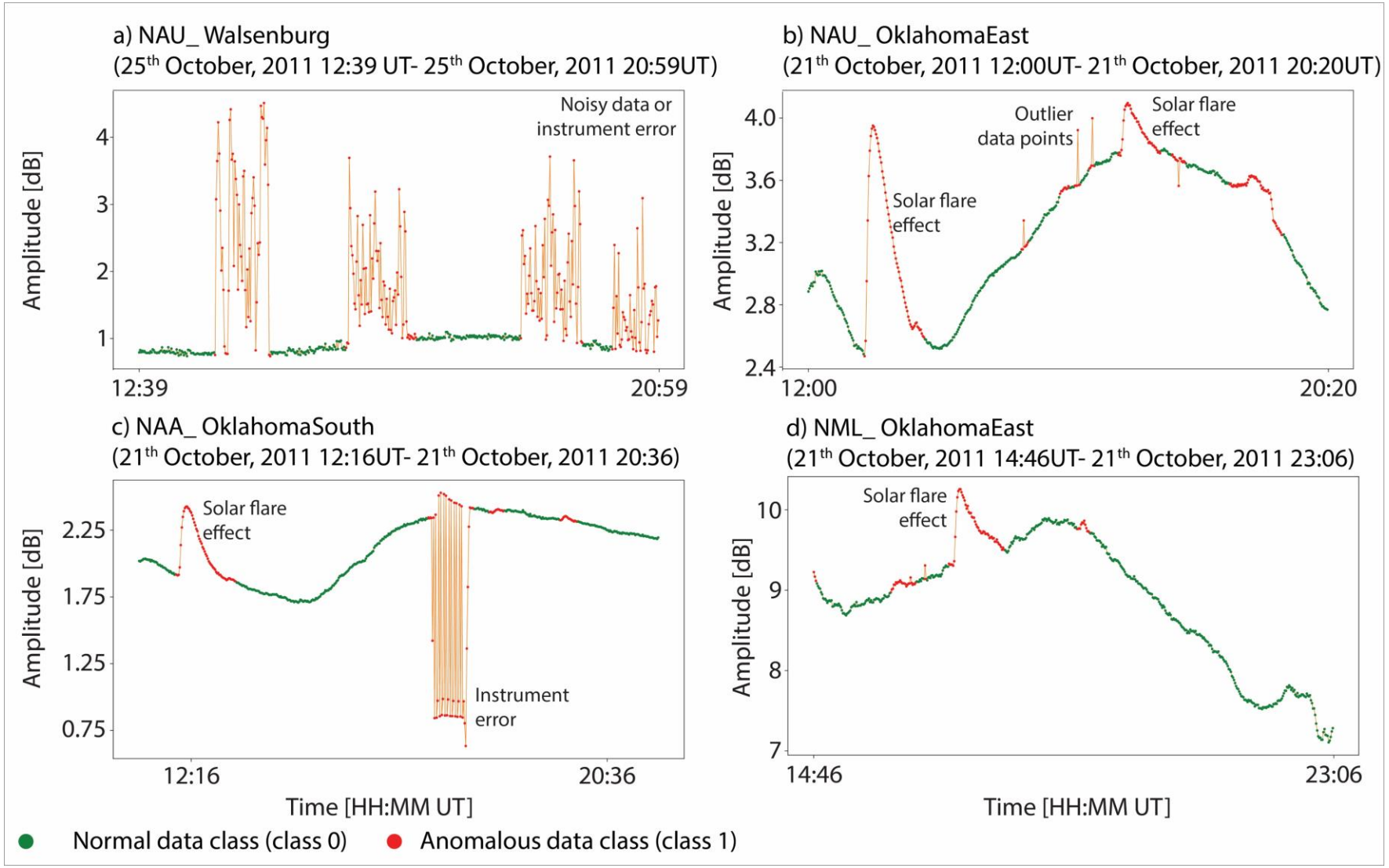
- **Solution:** The utilization of Youden's J statistic, Markedness, General Performance Score (GPS) and Unified Performance Measure (UPM) to strengthen the selection of the employed model (hyper)parameters. Furthermore, the inclusion of a domain-specific evaluation metric can enhance the assessment process.

- **Rationale:** The utilization of GPS and UPM has the potential to provide supplementary insights into the developed models as a single-number metric. Furthermore, the development of a domain-specific evaluation metric may be considered as the most optimal choice among all available evaluation metrics.

Methods and data

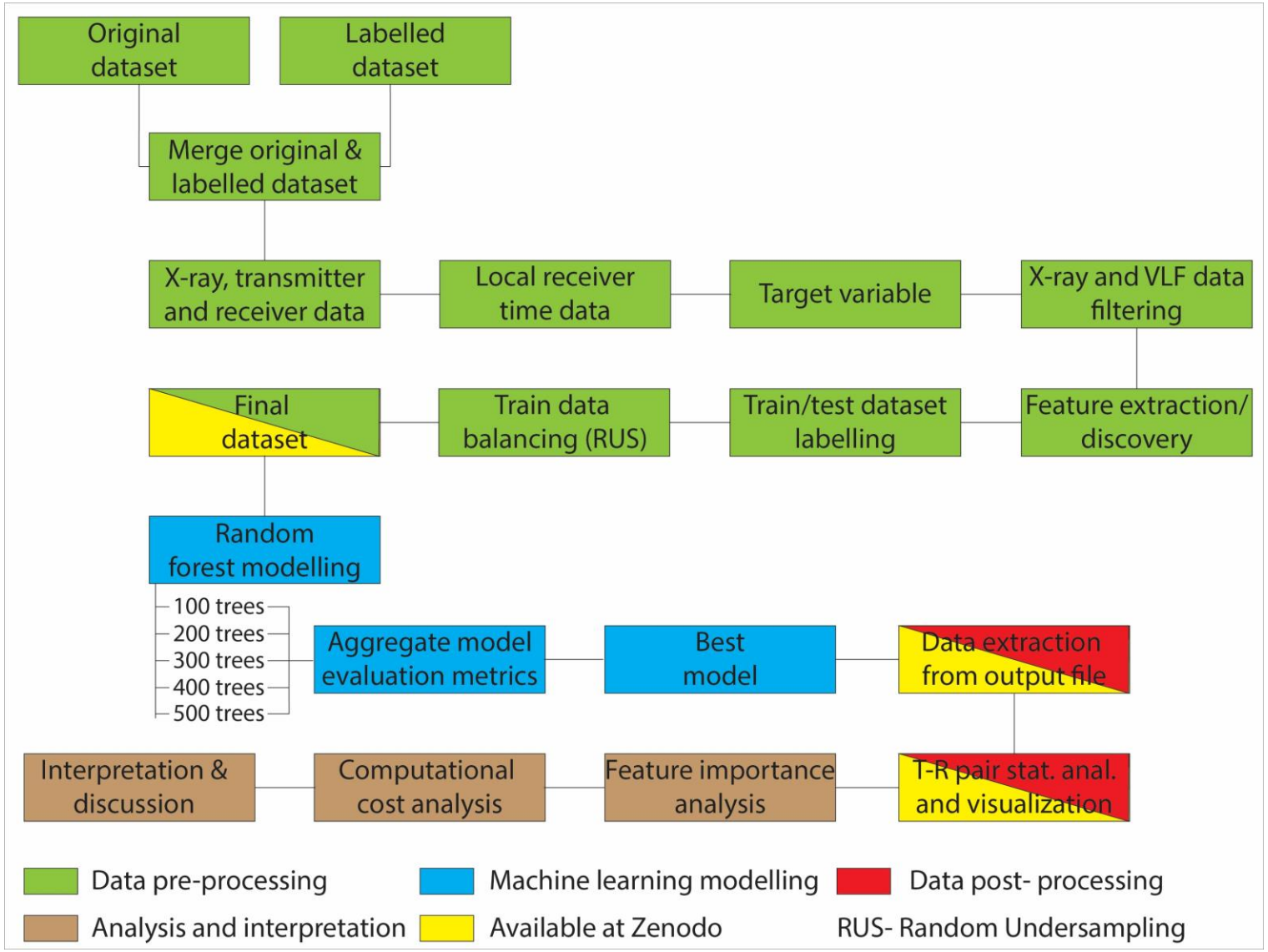
- Why data exclusion?

  Manual data exclusion is a time-consuming and tedious process.



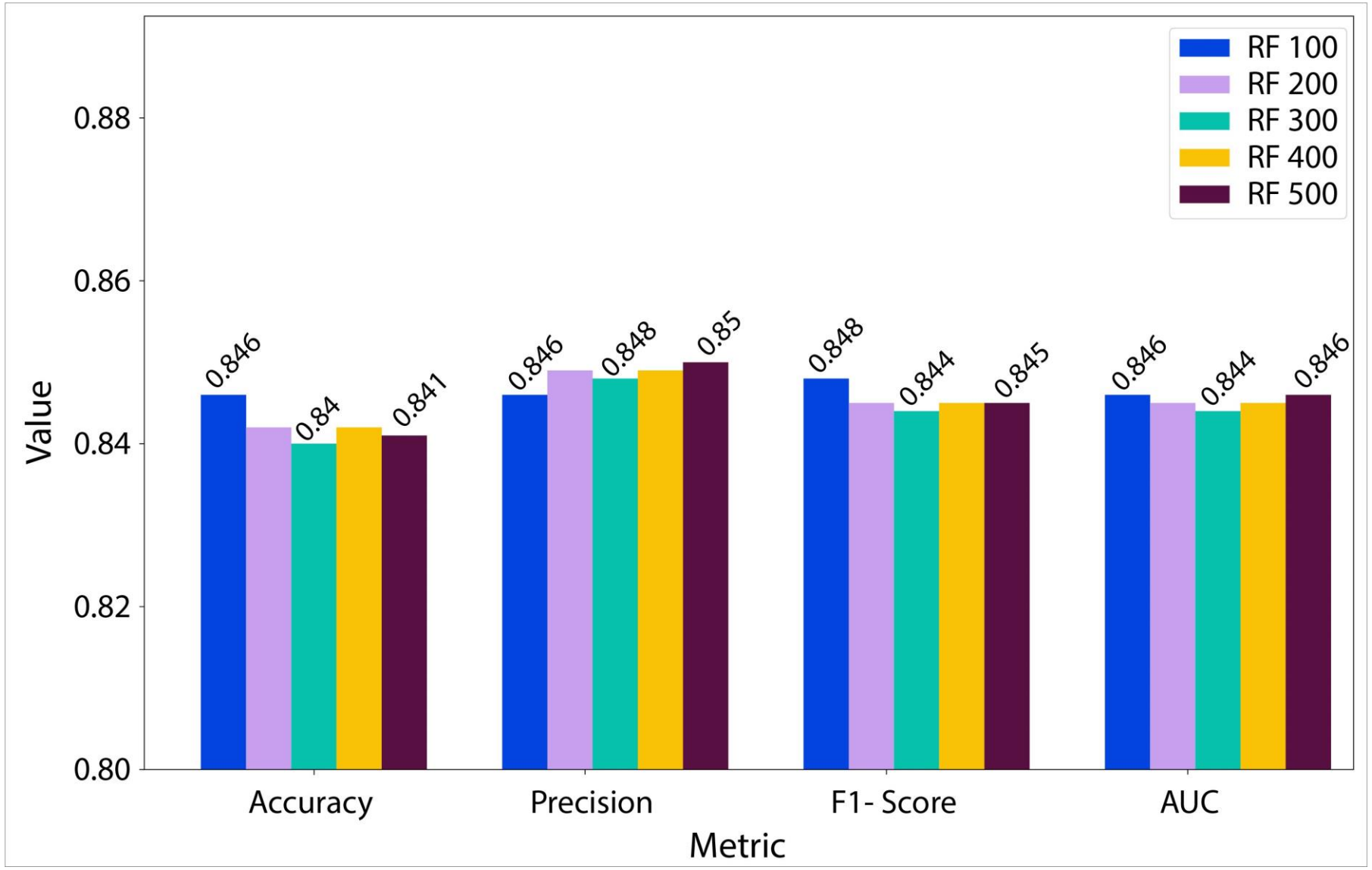
Methods and data

- Data exclusion workflow



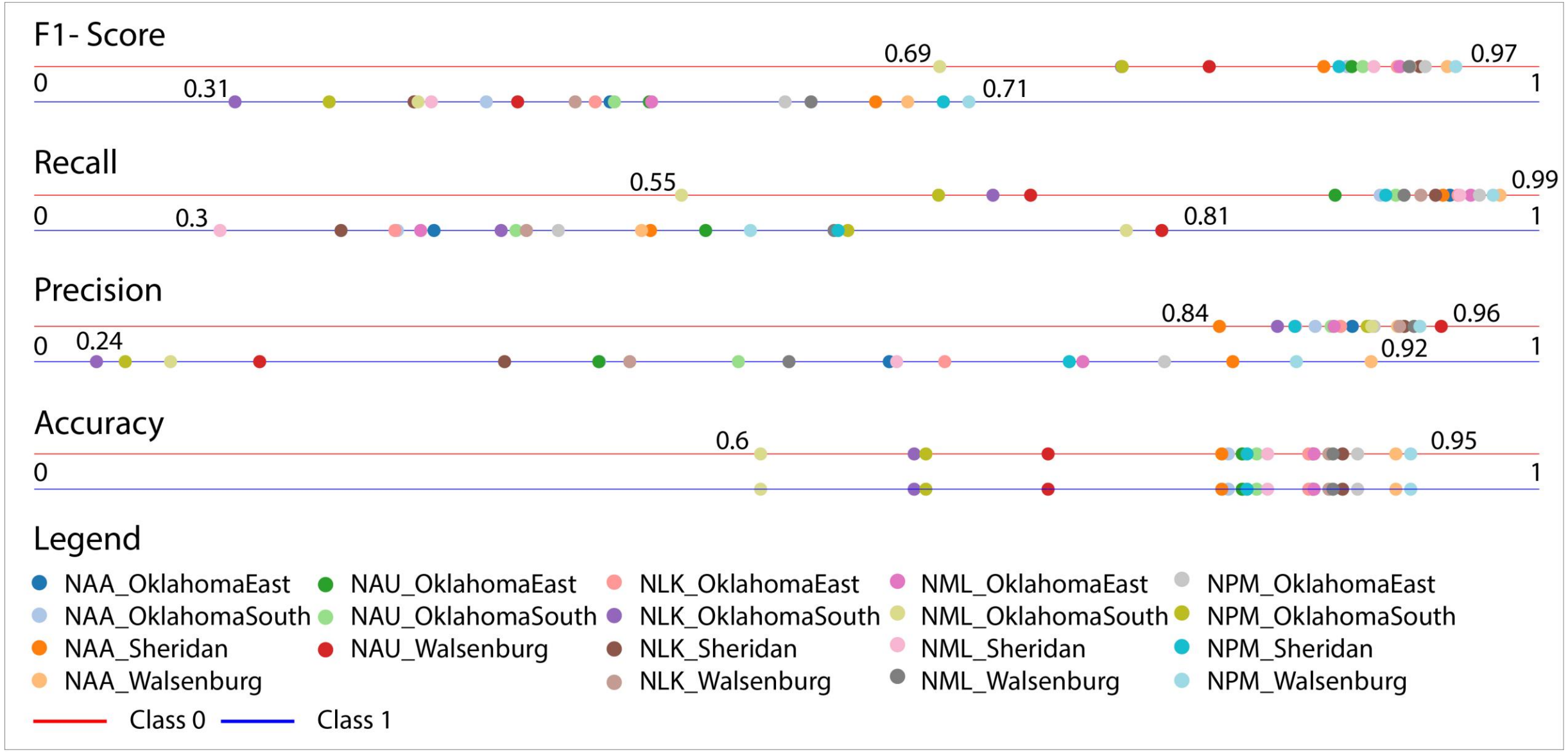
Results and discussion

Random forest modelling



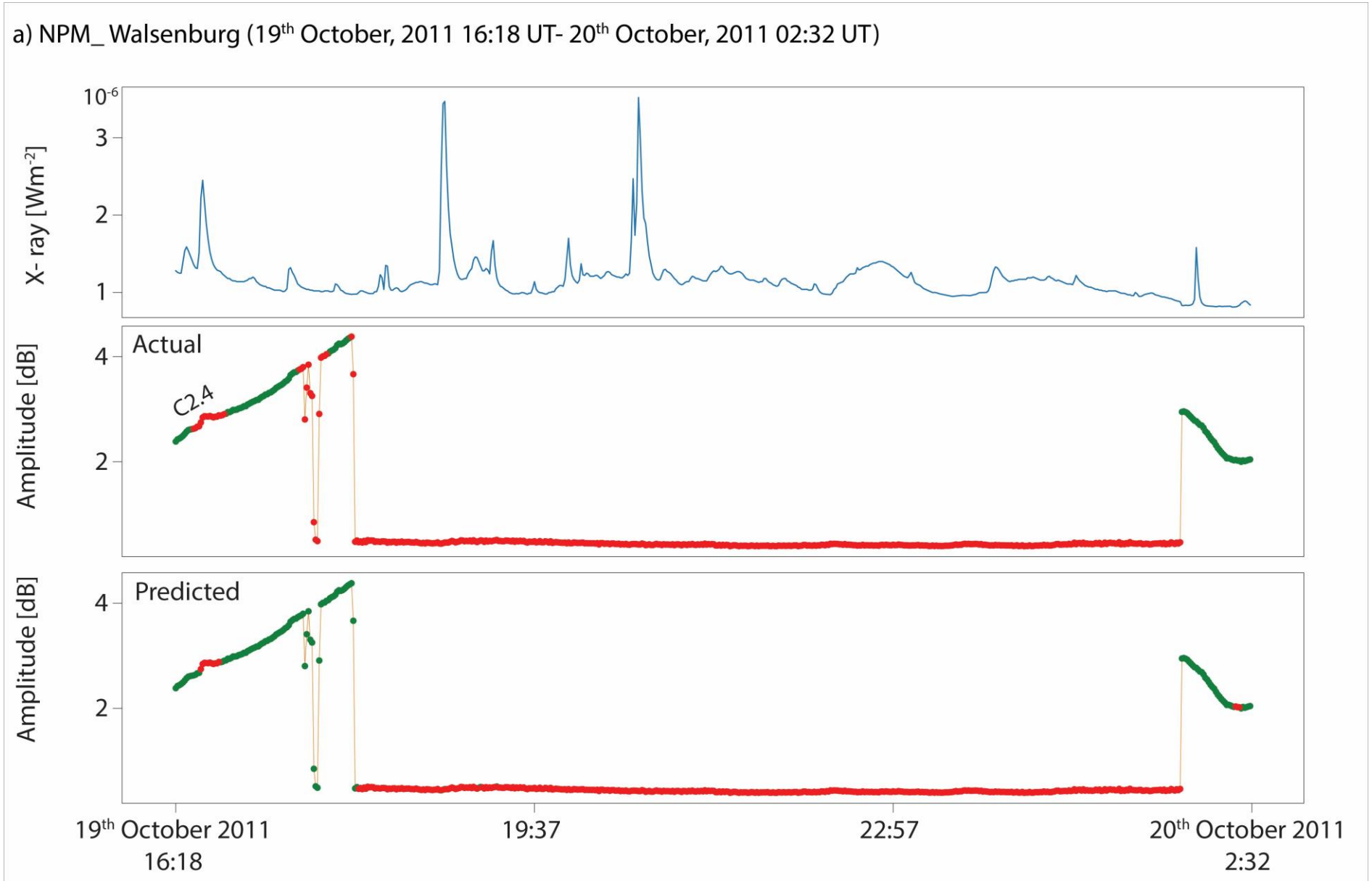
Results and discussion

Random forest modelling



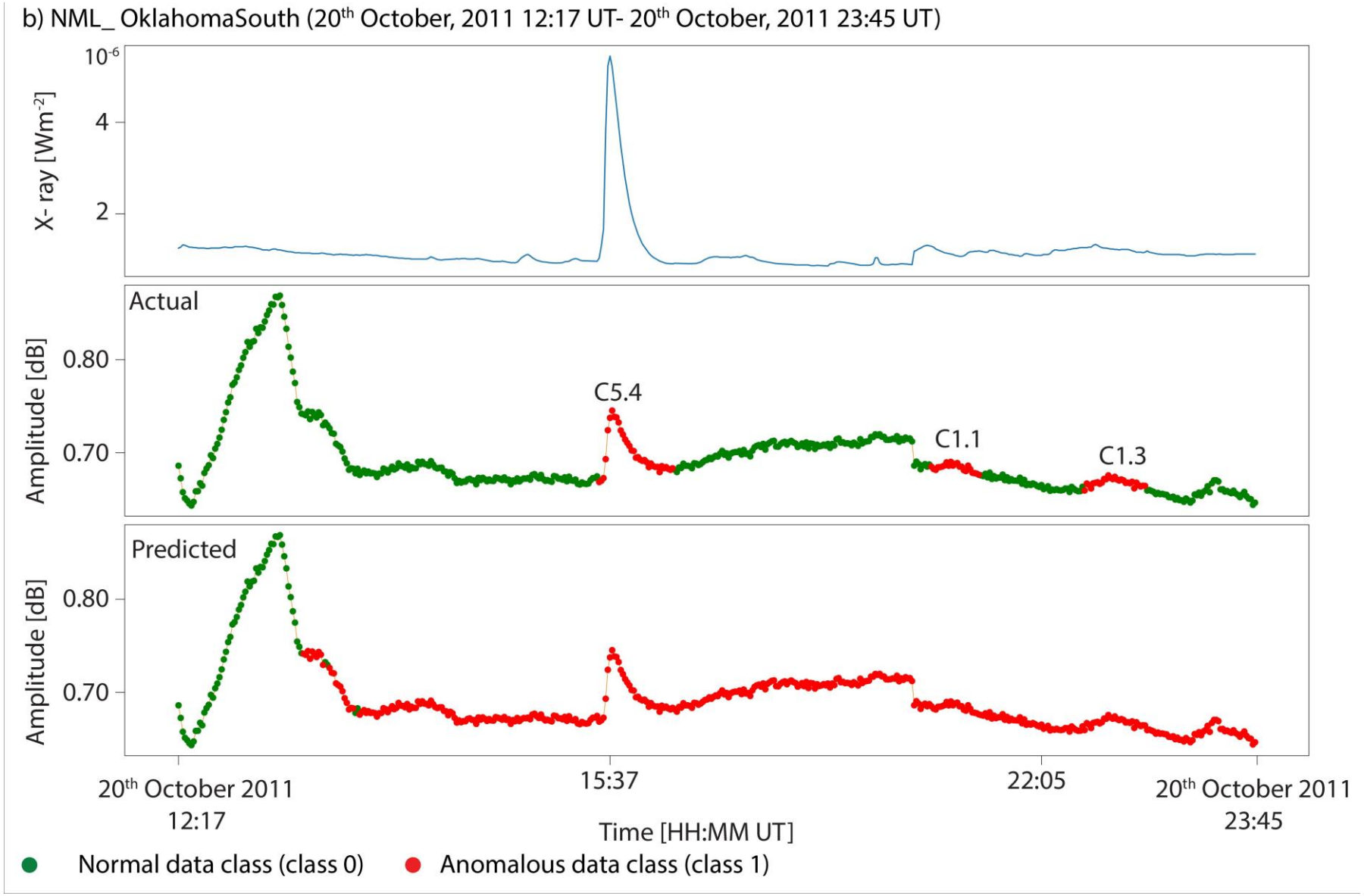
Results and discussion

Random forest modelling



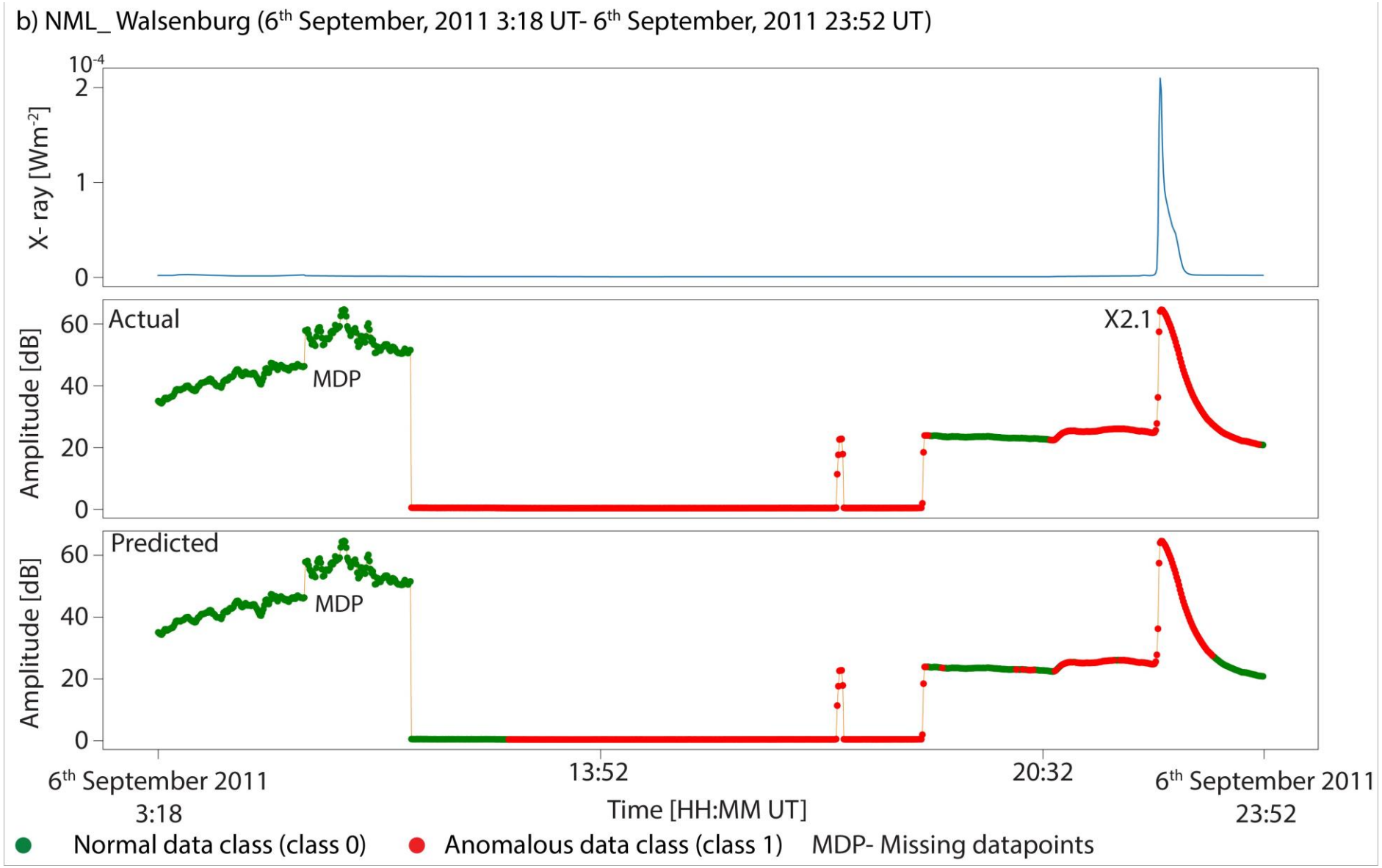
Results and discussion

Random forest modelling



Results and discussion

Random forest modelling



Results and discussion

Alternative evaluation metrics

Number of trees	Youden's J statistic	Markedness	GPS	UPS
100	0.692	0.673	0.776	0.833
125	0.689	0.662	0.772	0.829
150	0.684	0.653	0.769	0.825
175	0.688	0.667	0.772	0.830
200	0.688	0.664	0.772	0.830
300	0.687	0.659	0.771	0.828
400	0.688	0.665	0.772	0.830
500	0.690	0.662	0.773	0.830
Minimum	0.684	0.653	0.769	0.825
Maximum	0.692	0.673	0.776	0.833
Range	0.008	0.021	0.007	0.008
Mean	0.688	0.663	0.772	0.829
Median	0.688	0.663	0.772	0.830

Results and discussion

Domain- specific evaluation metric

Confusion matrix

		True labels	
		0	1
Predicted labels	0	TP	FP
	1	FN	TN

Point-based system:

- True positive (TP) = +1 point (majority class),
- True negative (TN)= +2 points (minority class),
- False positive (FP)= -1 point (misclassified majority class)- could be informative,
- False negative (FN)= -2 points (misclassified minority class).

$$Total\ score = TP + 2TN + FP - 2FN$$

$$max\ score = TP + 2 * TN$$

$$Score\ ratio = \frac{Total\ score}{max\ score}$$

Properties:

- Score ratio values range between -1 and 1.
- A value of -1 represents an ideally bad model (FP and FN values).
- A value of 1 represents an ideally good model (TN and TP values).
- A value of 0 means that the model made the same amount of FN as TN and TP as FP.

Number of trees	Score ratio
100	0.609
125	0.601
150	0.594
175	0.602
200	0.602
300	0.599
400	0.601
500	0.603

Conclusion

- Ionospheric VLF data exclusion is a process that can be automated with the utilization of ML classification methods.

- Further research is needed in order to fine-tune the workflow, such as testing different undersampling and oversampling techniques, the application of other ML methods, etc.

- For first-time research into the subject, the obtained results are satisfactory. The application of alternative evaluation metrics confirmed that the best overall model is comprised of 100 trees.

- The development of a domain-specific evaluation metric based on a point system from the confusion matrix demonstrated the benefits of having a single-number metric to quickly determine a few of the best models.

Thank you for your attention.