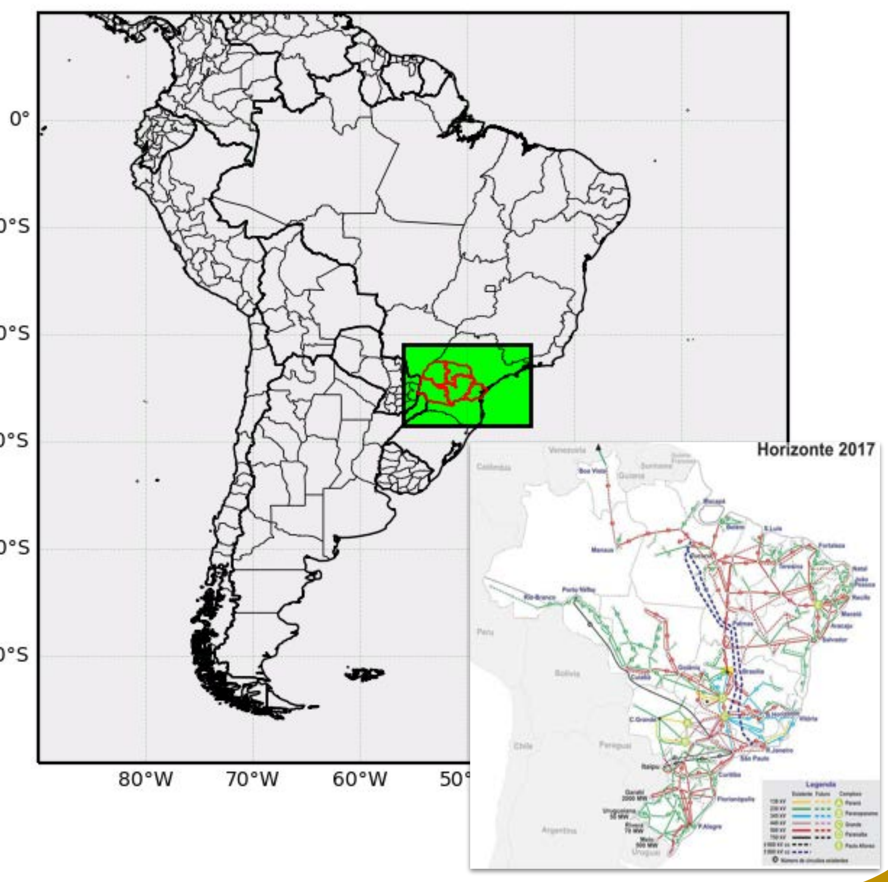
Post Processing Quantitative Precipitation Forecasts using Machine Learning in Southern Brazil

L Guzman[1], J. Silveira[1], C. Beneti[1], S. Paz[1], R. Inouye[1], L. Calvetti[2]
[1]SIMEPAR - Parana Environmental Technology and Weather Monitoring Service, Curitiba, Brazil
[2]UFPEL - Federal University of Pelotas, Pelotas, Brazil

Introduction

In South America, **Southern Brazil,** Paraguay and northeast Argentina are regions particularly prone to high-impact weather (intense lightning activity, high precipitation, hail, flash floods, and occasional tornadoes).
Most of the precipitation is associated with extra-tropical cyclones, frontal systems, and **Mesoscale Convective Systems,** responsible for more than 80% of the **precipitation** and severe weather events which occur in this area.
Precipitation is well distributed during the year, but the **South American Low-Level Jet**, east of Andes, brings more moisture and convective instability to the region during spring and summer, favoring the occurrence of high lightning incidence and strong precipitation, in this period of the year.
In the south of Brazil, agricultural industry and electrical power generation are the main economic activities.
This region is responsible for 35% of all hydro-power energy production in the country, with long transmission lines to the main consumer regions, which are severely affected by these extreme weather conditions.
The MCS play an important role on the hydrological cycle and on the incidence of severe weather events, highlighting the importance of improving the knowledge of those weather systems, with the goal of better forecasts.



Data and Methodology

The post-processing of Numerical Weather Prediction (NWP) is one alternative on the path to reducing weather forecast uncertainties, especially in quantitative precipitation forecasting. By post-processing large ensemble system simulations, in addition to providing ideas about the uncertainty of the model outputs, possible errors of space and time mislocation of precipitation events may be prevented.
In this work, the quantitative precipitation forecasting problem was addressed with Machine Learning (ML) techniques. A CatBoost-based approach that uses the residue of the previous model to adjust the next model to be trained, was implemented for NWP post-processing. A multimodel ensemble prediction system, combining 21 model outputs from the Global Forecast System (GFS) and 51 from the European Center for Medium-Range Weather Forecasts (ECMWF), was designed for forecast lead times up to 15 days between January 2020 and May 2022.
Daily precipitation thresholds of 1 mm, 15 mm, and 40 mm were considered for the analysis. ML predictors were derived from ensemble forecasts in South Brazil. The training set was submitted to preprocessing (normalization and imputing missing data). A Binary CatBoost Classifier algorithm was used to train models for each combination of precipitation threshold and forecast horizon, with 45 ML models. A second approach was a ML model that combines daily precipitation forecasts of the previous five days to generate a forecast for each date.

Results

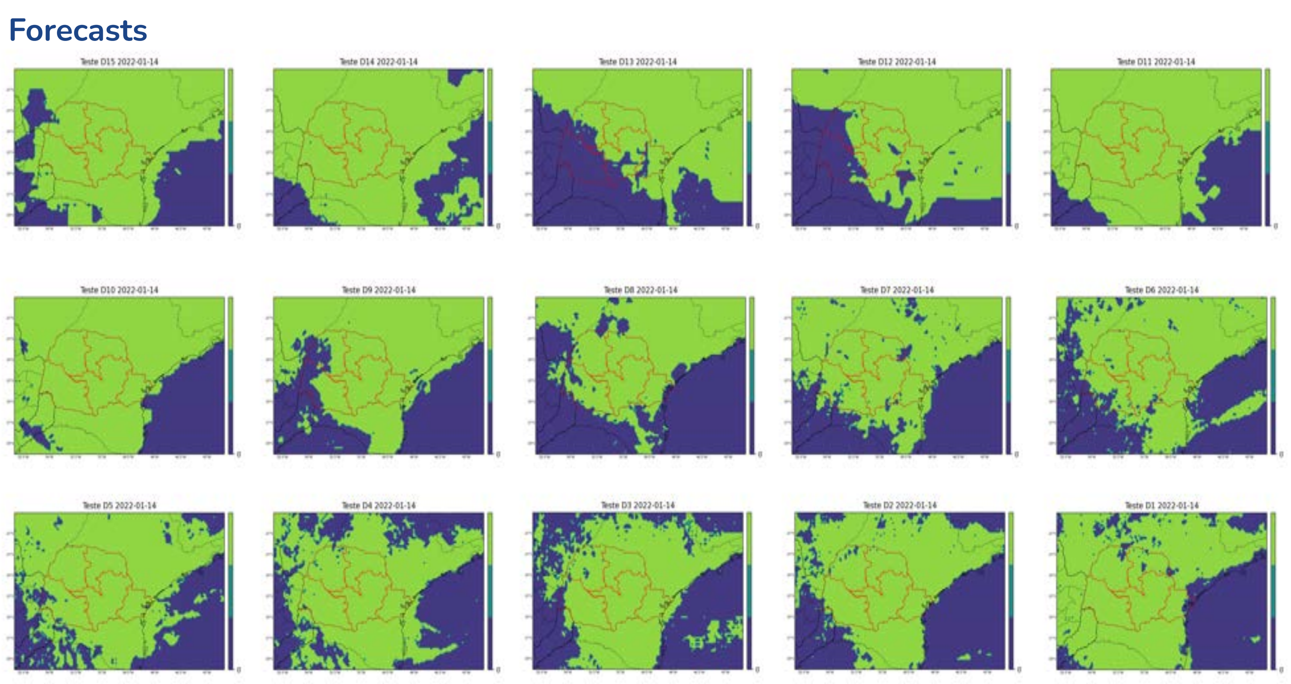


Figure 1: 15 days ML-ENS-QPF Forecasts and Observed Precipitation (>1mm/day).

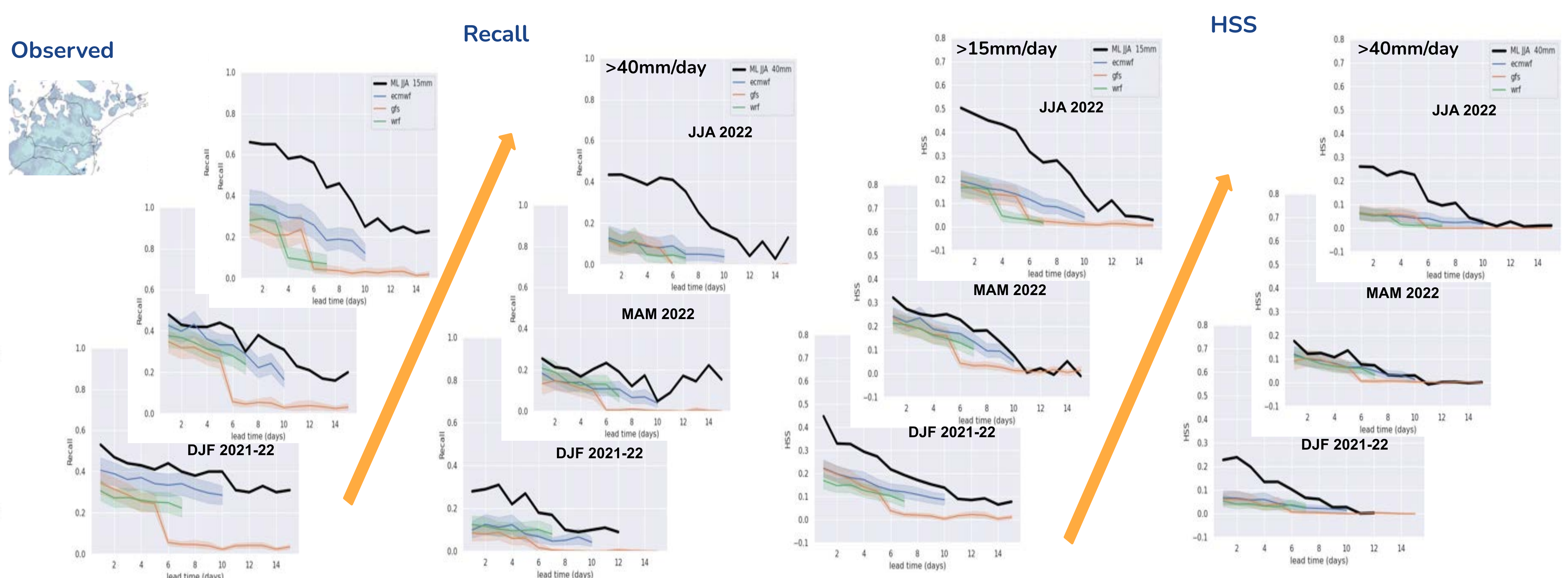


Figure 2: Evolution of the ML-ENS-QPF and deterministic forecasts (of >15mm/day and >40mm/day) recall and Heidke skill score, from DJF/2021-22 until JJA/2022.

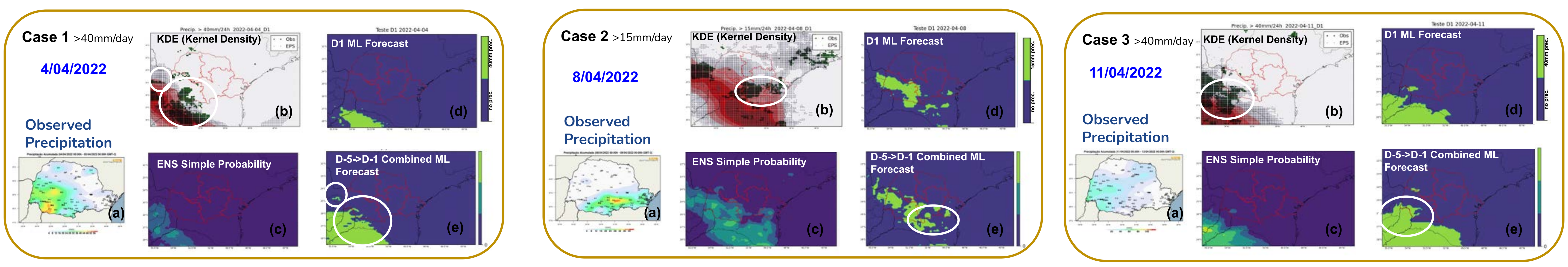


Figure 3: Examples of precipitation events of >15mm/day and >40mm/day over the study region: (a) observed precipitation by Siprec, (b) KDE (Kernel Density) in red and Siprec in green, (c) Ensemble simple probability contours, (d) D-1 ML Model forecast and (e) D-5 to D-1 Combined ML Forecast, for 4/04/2022, 8/04/2022 and 11/4/2022.

Conclusions

Based on hyperparameter sensitivity tests, the CatBoost required a maximum tree depth of 8 levels and a learning rate of 0.1.
The results showed that skill score values were strongly dependent on the precipitation thresholds considered.
Recall values showed that ML models were more successful in forecasting precipitation events than any deterministic model output analyzed.
More than 50% of precipitation events of more than 15mm and around 40% of precipitation events of more than 40mm were correctly forecasted with ML models from D1 to D6. The best results of the ML daily models were for shorter ensemble forecast lead times, D1 being the best ML model.
ML models considering precipitation forecast of the previous five days performed better than the ML daily models.

14–17 November 2022 | #AIforEOWS
ECMWF-ESA Workshop on Machine Learning for Earth Observation and Prediction
ECMWF esa