

EGU22-9734, updated on 02 Dec 2022

<https://doi.org/10.5194/egusphere-egu22-9734>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



High Impact Weather Forecasts in Southern Brazil using Ensemble Precipitation Forecasts and Machine Learning

Cesar Beneti¹, Jaqueline Silveira¹, Leonardo Calvetti², Rafael Inouye¹, Lissette Guzman¹, Gustavo Razera², and Sheila Paz¹

¹SIMEPAR - Parana Environmental Technology and Weather Monitoring Service, Curitiba, Brazil (cesar.beneti@simepar.br)

²UFPEL - Federal University of Pelotas, Pelotas, Brazil (lcalvetti@gmail.com)

In South America, southern parts of Brazil, Paraguay and northeast Argentina are regions particularly prone to high impact weather (intensive lightning activity, high precipitation, hail, flash floods and occasional tornadoes), mostly associated with extra-tropical cyclones, frontal systems and Mesoscale Convective Systems. 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. Intense precipitation events are a common cause of electricity outages in southern Brazil, which ranks as one of the regions in Brazil with the highest annual lightning incidence, as well. Accurate precipitation forecasts can mitigate this kind of problem. Despite improvements in the precipitation estimates and forecasts, some difficulties remain to increase the accuracy, mainly related to the temporal and spatial location of the events. Although several options are available, it is difficult to identify which deterministic forecast is the best or the most reliable forecast. Probabilistic products from large ensemble prediction systems provide a guide to forecasters on how confident they should be about the deterministic forecast, and one approach is using post processing methods such as machine learning (ML), which has been used to identify patterns in historical data to correct for systematic ensemble biases.

In this paper, we present a study, in which we used 20 members from the Global Ensemble Forecast System (GEFS) and 50 members from European Centre for Medium-Range Weather Forecasts (ECMWF) during 2019-2021, for seven daily precipitation thresholds: 0-1.0mm, 1.0mm-15mm, 15mm-40mm, 40mm-55mm, 55mm-105mm, 105mm-155mm and over 155mm. A ML algorithm was developed for each day, up to 15 days of forecasts, and several skill scores were calculated, for these daily precipitation thresholds. Initially, to select the best members of the ensembles, a gradient boosting algorithm was applied, in order to improve the skill of the model and reduce processing time. After preprocessing the data, a random forest classifier was used to train the model. Based on hyperparameter sensitivity tests, the random forest required 500 trees, a maximum tree depth of 12 levels, at least 20 samples per leaf node, and the minimization of entropy for splits. In order to evaluate the models, we used a cross-validation on a limited data sample. The procedure has a single parameter that refers to the number of groups that a given

data sample is to be split into. In our work we created a twenty-six fold cross validation with 30 days per fold to verify the forecasts. The results obtained by the RF were evaluated through estimated value versus observed value. For the forecast range, we found values above 75% for the precision metrics in the first 3 days, and around 68% in the next days. The recall was also around 80% throughout the entire forecast range, with promising results to apply this technique operationally, which is our intent in the near future.