Contribution ID: **22**

Type: **Oral presentation**

Using Machine Learning to advance hour-scale heavy rain forecast with high resolution ECMWF Global Model and Local Meso-scale Model Forecasts

Tuesday, 6 October 2020 09:00 (30 minutes)

Numerical weather predictions (NWP) and observational systems have improved greatly both on quantity and quality in recent years. Combining expertise knowledge and machine learning (ML) to extract and synthesize valuable information from these "big data" is expected to one of present major challenges and opportunities to improve the severe weather forecast. A ML correction method based on the physical ingredients was developed in this study to improve the hour-scale flash heavy rain forecast, which contains a key factors extraction technique by blending ML feature engineering and expert knowledge , and several ML models.
The summer flash heavy rain events in the Beijing-Tianjin-Hebei region of China are studied. Firstly, the hourly features were analyzed based on ten years observations. Then a comprehensive verification was conducted with the operational forecasts including the high resolution ECMWF global model and mulitple local forecasts (SMS 9 km and 3km, GRAPES 9km and 3km, BJ-RUC 9km and 3km) , and those with better performances are chosen as ML Components. Data in 2018 and 2019 summer are used for training and test the ML models. Here, we report on initial results.
More than 200 thermo-dynamical features were selected by expertise. Multiple ML feature engineering techniques including the correlation coefficient, the mutual information, and the embedded methods are combined and used to furtherly select features and get rid of the redundancy. The results show that the finer the scale is the more important the physical features are. There is a good correlation between 6hr (or 12hr) accumulated precipitation and the estimated precipitation. However, the direct rainfall forecasts become less important when it comes to the 3hr (or 1hr) accumulations, and the contributions of thermo-dynamical features are significantly enhanced, in particular the moisture at low level and the wind field near surface. This means that the uncertainty of model rain is increasing with the finer of the forecast, which is consistent with the professional cognition. Therefore, by using forecasted physical ingredients with more reliability and the rapidly updating observations may be potential to improve the hour-scale flash heavy rain forecast. Multiple ML models including ET (Extra Tree)/RF (Radom Forest)/Catboost were trained, and then a secondary ensemble was conducted. The results show that the ML correction based on physical features can significantly improve the forecast on 3-hour cumulative precipitation compared to the ECMWF and local meso-scale model. Taking the threshold of 10mm rainfall as an example, the Ts score of ET model is 0.27, which is about 35% higher than those of the numerical forecasts (0.2 or less); the missing alarm rate is about 26% lower than ECMWF forecast, and the false alarm rate is about 28% lower than the local SMS model. A case study of flash heavy rain in August 2019 shows that both the intensity and the heavy rain area are significantly improved in the ML correction system.

Thematic area

1. Machine Learning for Product development - Including NWP Post-processing, Non-linear Ensemble Averaging, Development of new NWP Products

Primary author: Prof. ZHONG , Qi (China Meteorological Administration Training Centre)

Co-authors: Mr FANG , ZU-LIANG (Chinese Academy of Meteorological Sciences); Mr SUN, ZHUO (Hebei Meteorological Observatory)

Presenter: Prof. ZHONG , Qi (China Meteorological Administration Training Centre)

Session Classification: Session 2 (cont.): ML for Earth System Observations

Track Classification: ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction