

Ensemble post-processing of sub-seasonal to seasonal precipitation forecasts based on Probabilistic Double Machine Learning method

Shengsheng Zhan¹, Aizhong Ye^{1*}, Lingyun Wu¹, Chenguang Zhao¹

¹ State Key Laboratory of Earth Surface Processes and Hazards Risk Governance, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
Email: azye@bnu.edu.cn

Abstract: Subseasonal-to-seasonal (S2S) precipitation forecasting is crucial for hydrological modeling; however, its accuracy often falls short of the requirements for hydrological forecasts, necessitating post-processing. A novel improved version of the Double Machine Learning (DML) method, termed Probabilistic Double Machine Learning (PDML), is proposed for ensemble post-processing of S2S forecasts. The new PDML method (Figure 1) extends the classifier from binary classification to multi-class classification, improves the regressor from single-value output to probability distribution output, and combines the classifier and regressor based on total probability theorem. PDML not only quantifies uncertainty through ensemble output but also provides additional consideration for extreme precipitation events in the classification and regression progress. Various machine learning methods are compared within the PDML framework, including the state-of-the-art Kolmogorov-Arnold Networks. The results indicate that deep learning models based on Recurrent Neural Networks (RNN) and the U-NET architecture perform the best within the PDML framework. It achieves post-processing of S2S forecasts across different timescales and outperforms the statistical Ensemble Pre-Processor (EPP) method. On average, it improves the original forecast's correlation coefficient, critical success index, and root mean square error by 85.8%, 294.6%, and 45.3%, respectively, and achieves an 8.6% improvement on the continuous ranked probability score compared to EPP. The results demonstrate that PDML can effectively perform ensemble post-processing of precipitation forecasts across different timescales, quantify uncertainty, and facilitate further hydrological modeling.

Keywords: Sub-seasonal to seasonal; Precipitation Forecasts; Post-processing; Probabilistic Double Machine Learning

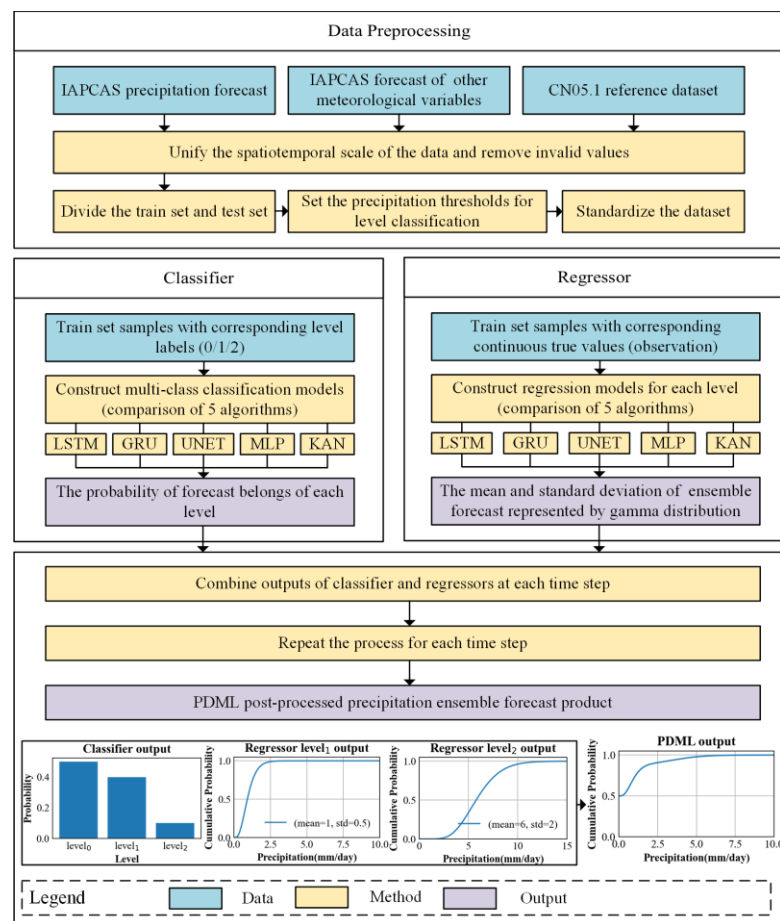


Figure 1. Flowchart showing the structure of the PDML method