IAHS2017-62
IAHS Scientific Assembly 2017

Evaluation of different multi-model ensemble combinations for flood forecasting

Martijn Booij (1), Loek Zomerdijk (1), Yue-Ping Xu (2), and Maarten Krol (1)

(1) University of Twente, Water Engineering and Management, Enschede, Netherlands (m.j.booij@utwente.nl), (2) Institute of Hydrology and Water Resources, Zhejiang University, Hangzhou, China

Flood forecasting in small mountainous catchments often is difficult due to short lead times of precipitation forecasts. Multi-model ensemble (grand-ensemble) forecasts can potentially improve these forecasts. However, techniques to combine different ensemble prediction systems (EPSs) need to be developed. This study aims to develop an ensemble flood forecasting system for Quzhou (east China) for lead times of 1-10 days and to evaluate different combinations of multi-model ensemble flood forecasts. The lumped hydrological GR4J model is used to forecast flow with ensemble precipitation forecasts of four different weather centres (European Centre for Medium-Range Weather Forecasts (ECMWF); Chinese Meteorological Administration (CMA); UK Met Office (UKMO) and US National Centers for Environmental Prediction (NCEP)) as input. The EPSs of these centres have different ensemble sizes and include one control forecast. The ensemble forecasts are bias corrected with the Quantile Mapping method. The bias-corrected precipitation forecasts are used as input into the hydrological model. The flood forecasts are evaluated on three important components of skill: reliability, resolution and sharpness. Six different grand ensemble flood forecasts are constructed after the evaluation of the single model forecasts. The first is a combination of the members where the EPSs are not weighted and, as a consequence, EPSs with more ensembles have more influence on the grand ensemble. The second is a combination of the models where the models are weighted so that their influence on the grand ensemble is equal. Other combinations are constructed using weighted contributions based on the skill of the evaluated EPSs. As expected, evaluation of the flood forecasts show that skill decreases with lead time and with increasing exceedance threshold. Two recognizable components of the forecast error, the meteorological error and the hydrological model error, both increase with lead time, with an increasing contribution of the meteorological error compared to the hydrological error with lead time. All forecasts have relatively good performance in terms of reliability, resolution and sharpness. In general, ECMWF proves to be the most skilful model and CMA the least skilful model for the Quzhou catchment area. All evaluations of the grand ensemble hydrological forecasts have added value. They show improved root mean squared errors (RMSE), continuous ranked probability skill scores (CRPSS), reliability and resolution as compared to the single EPSs. Simple combination methods of the grand ensembles show similar skill as combinations of ensembles forecasts using weighted contributions based on skills. A model with less skill might be able to add model structure errors that are missing in other EPSs with good skill and might have good performance on days when the other models show low performance. Generally it can be concluded that there is no significant difference between the different combination methods. It is recommended to use an approach where the models are equally weighted so that the influence of models is not dependent on ensemble size.