EGU21-8256, updated on 16 Oct 2021
https://doi.org/10.5194/egusphere-egu21-8256
EGU General Assembly 2021

Building hydrological single-model ensembles using artificial neural networks and a combinatorial optimization approach

Juan F. Farfán-Durán[1] and Luis Cea[2]
[1]Centre for Technological Innovations in Construction and Civil Engineering, University of A Coruña, A Coruña, Spain (j.farfan@udc.es)
[2]Centre for Technological Innovations in Construction and Civil Engineering, University of A Coruña, A Coruña, Spain (luis.cea@udc.es)

In recent years, the application of model ensembles has received increasing attention in the hydrological modelling community due to the interesting results reported in several studies carried out in different parts of the world. The main idea of these approaches is to combine the results of the same hydrological model or a number of different hydrological models in order to obtain more robust, better-fitting models, reducing at the same time the uncertainty in the predictions. The techniques for combining models range from simple approaches such as averaging different simulations, to more complex techniques such as least squares, genetic algorithms and more recently artificial intelligence techniques such as Artificial Neural Networks (ANN).

Despite the good results that model ensembles are able to provide, the models selected to build the ensemble have a direct influence on the results. Contrary to intuition, it has been reported that the best fitting single models do not necessarily produce the best ensemble. Instead, better results can be obtained with ensembles that incorporate models with moderate goodness of fit. This implies that the selection of the single models might have a random component in order to maximize the results that ensemble approaches can provide.

The present study is carried out using hydrological data on an hourly scale between 2008 and 2015 corresponding to the Mandeo basin, located in the Northwest of Spain. In order to obtain 1000 single models, a hydrological model was run using 1000 sets of parameters sampled randomly in their feasible space. Then, we have classified the models in 3 groups with the following characteristics: 1) The 25 single models with highest Nash-Sutcliffe coefficient, 2) The 25 single models with the highest Pearson coefficient, and 3) The complete group of 1000 single models.

The ensemble models are built with 5 models as the input of an ANN and the observed series as the output. Then, we applied the Random-Restart Hill-Climbing (RRHC) algorithm choosing 5 random models in each iteration to re-train the ANN in order to identify a better ensemble. The algorithm is applied to build 50 ensembles in each group of models. Finally, the results are compared to those obtained by optimizing the model using a gradient-based method by means of

the following goodness-of-fit measures: Nash-Sutcliffe (NSE) coefficient, adapted for high flows Nash-Sutcliffe (HF−NSE), adapted for low flows Nash-Sutcliffe (LF−W NSE) and coefficient of determination (R2).

The results show that the RRHC algorithm can identify adequate ensembles. The ensembles built using the group of models selected based on the NSE outperformed the model optimized by the gradient method in 64 % of the cases in at least 3 of 4 coefficients, both in the calibration and validation stages. Followed by the ensembles built with the group of models selected based on the Pearson coefficient with 56 %. In the case of the third group, no ensembles were identified that outperformed the gradient-based method. However, the most part of the ensembles outperformed the 1000 individual models.

Keywords: Multi-model ensemble; Single-model ensemble; Artificial Neural Networks; Hydrological Model; Random-restart Hill-climbing