Building Hydrological Single-Model Ensembles Using Artificial Neural Networks and a Combinatorial Optimization Approach

Juan F. Farfán-Durán[1]* and Luis Cea [1]

[1]Environmental and Water Engineering Group, Department of Civil Engineering, University of A Coruña, A Coruña, Spain, *j.farfan@udc.es

1. Introduction

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 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, least squares, genetic algorithms and artificial intelligence techniques such as Artificial Neural Networks (ANN) [1, 2].

2. Basic Idea

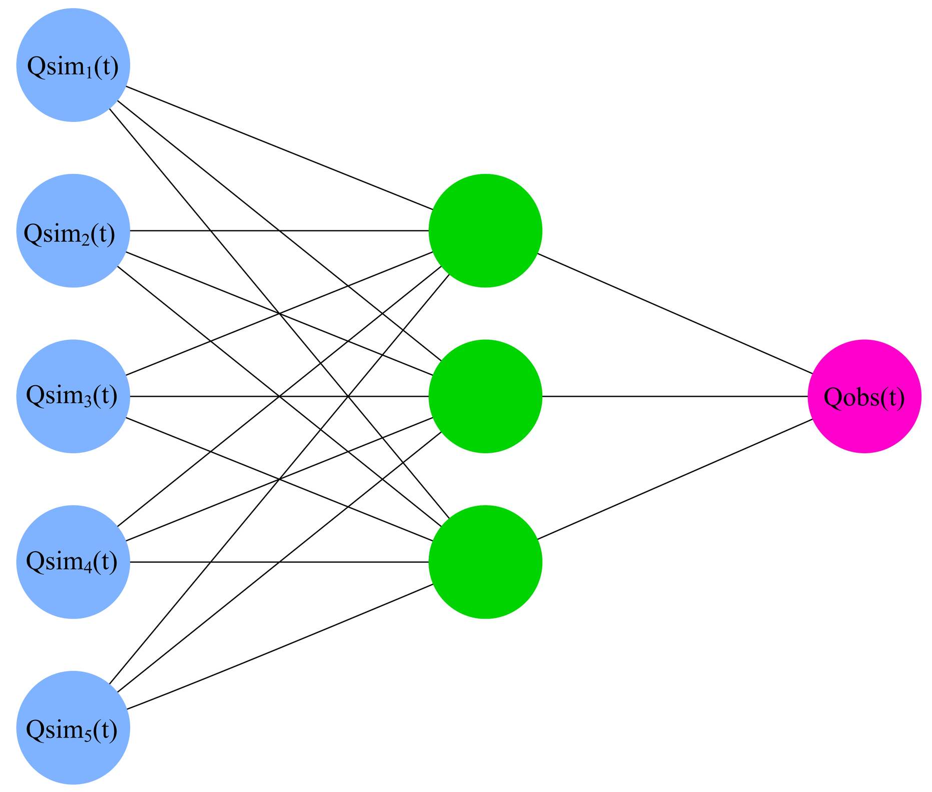
The present study uses hydrological data on an hourly scale between 2008 and 2013 for a calibration perdiod and 2013 to 2016 for the validation period corresponding to the Mandeo basin, located in the Northwest of Spain. We address the construction of ANN-based single model ensembles by means of a combinatorial optimization approach. To this end, we use a lumped hydrological model run with a number of parameter sets sampled from its feasible space, in order to obtain a collection of individual hydrological models.

3. MHIA model

The hydrological model used in this work is the MHIA (acronym for *Lumped Hydrological Model* in Spanish), a lumped model developed by the authors. The model performs a balance of the volume of water in the soil taking into account the following processes: precipitation, infiltration, percolation, evapotranspiration and exfiltration. The requiered input data are the time series of precipitation and temperature. The model has 13 parameters that must be calibrated from time series of observed discharge. The MHIA model is run using 1000 sets of parameters randomly sampled from their feasible space. Then, the 1000 resultant models are classified in 3 samples: 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.

4. Proposed approach

The proposed approach uses the Random-Restart Hill-Climbing algorithm [3]. It starts by training an ANN using a random ensemble of models and evaluating its goodness of fit. Then, an iterative process is started in which a new random state is generated and evaluated. If this ensemble outperforms the previous ensemble in terms of an array of objective functions, it is saved instead of the previous one. The results are compared to those obtained by optimizing the model using a gradient-based method. The applied goodness-of-fit measures are: Nash-Sutcliffe (NSE), Nash-Sutcliffe for High Flows (HF−NSE), Nash-Sutcliffe for Low Flows (LF−NSE) and R^2 coefficient. This process is applied for each of the 3 samples of single models.



6.Preliminary Results

The results show that the RRHC algorithm can iden-tify adequate ensembles. The ensembles built using the group of models selected based on the NSE coeff-cient outperformed the model optimized by the gradi-ent method in 64 % of the cases in at least 3 of 4 coeffi-cients, for both calibration and validation periods (Fig 1). 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. This indicates that the individual models contribute more information to the ANN by forming the groups (1) and (2). The following table contrasts the results between the MHIA model optimized with a gradient-based method and the best ensemble iden-tified with criterion 1).

Model	Period	NSE	HF-NSE	LF-NSE	R_2
Gradient-based model	Calibration	0.88	0.91	0.86	0.94
	Validation	0.76	0.82	0.7	0.89
Best identifed ensemble	Calibration	0.90	0.91	0.87	0.95
	Validation	0.81	0.85	0.77	0.91

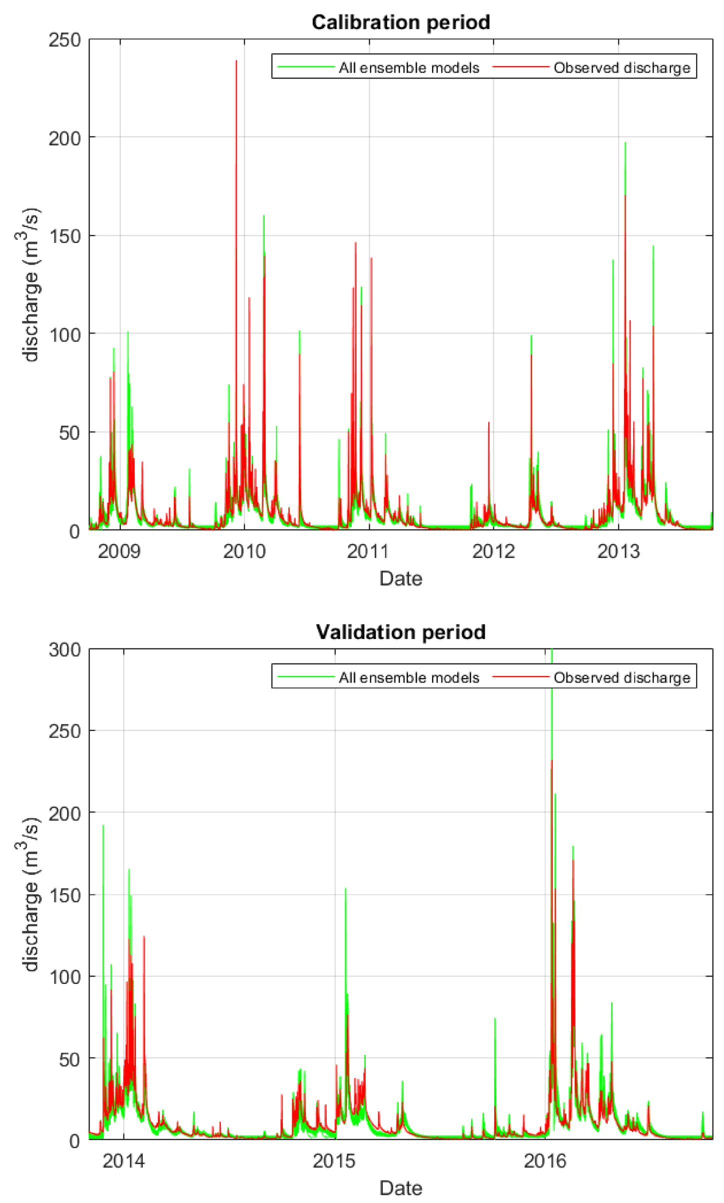


Fig 1: ANN-based ensembles build with the group 1) of individual models that outperformed the MHIA model optimized with a gradient-based method.

Conclusion The proposed combinatorial optimization approach has been shown consistently to be successful in finding adequate combinations of individual models for the construction of single model ensembles. However, its application to multi-model ensembles has not been tested and may serve as a basis for future studies.

5. Algorithm

Pseudocode for Random-Restart Hill-Climbing Algorithm

$i = 0$
while $i < 50$ **do**
  Set an initial random ensemble:
  current = initial random ensemble,
  current ∈ *Sample*
  best = *current*
  while Stop criterion is not reached **do**
    Train ANN:
    fit(*current*)
    if *fit*(*current*) >= *fit*(*best*) **then**
      best = *current*
    else
      current = random ensemble,
      current ∈ *Sample*
    end if
  end while
  $i = i + 1$
end while

i denotes the number of ensembles we want to identify, hence the number of restarts of the Hill-Climbing algorithm, **fit()** denotes goodness of fit, **current** is the combination of individual models used as input to the ANN, **best** is the best combination of models identified, **Sample** denotes the group of individual models from which we build the ensembles. The ANN is trained with the calibration period.

7. References

[1] Cynthia Andraos and Wajdi Najem. Multi-model approach for reducing uncertainties in rainfall-runoff models. In *Advances in Hydroinformatics*, pages 545–557. Springer, 2020.

[2] Juan F. Farfán, Karina Palacios, Jacinto Ulloa, and Alex Avilés. A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in andean watersheds. *Journal of Hydrology: Regional Studies*, 27:100652, 2020.

[3] Stuart J Russell and Peter Norvig. Artificial intelligence-a modern approach, third international edition., 2010.

Acknowledgements

This study is financed by the Galician government (*Xunta de Galicia*) as part of its pre-doctoral aid programme (*Axudas de apoio á etapa predoutoral 2019*) Register Nº ED481A-2019/014. The meteorological data was obtained from the agency MeteoGalicia. The streamflow data has been provided by the regional water administration *Augas de Galicia*.