Geophysical Research Abstracts
Vol. 15, EGU2013-**PREVIEW**, 2013
EGU General Assembly 2013

Finding diversity for building one-day ahead Hydrological Ensemble Prediction System based on artificial neural network stacks

Darwin Brochero (1), Francois Anctil (2), Christian Gagné (3), and Karol López (4)

(1) Chaire de recherche EDS en prévisions et actions hydrologiques, Department of Civil and Water Engineering, Université Laval, Quebec, Canada (darwin.brochero.1@ulaval.ca), (2) Chaire de recherche EDS en prévisions et actions hydrologiques, Department of Civil and Water Engineering, Université Laval, Quebec, Canada (francois.anctil@gci.ulaval.ca), (3) Computer Vision and Systems Laboratory (CVSL), Department of Electrical Engineering and Computer Engineering, Université Laval, Université Laval, Quebec, Canada (christian.gagne@gel.ulaval.ca), (4) Computer Vision and Systems Laboratory (CVSL), Department of Electrical Engineering andComputer Engineering, Université Laval, Quebec, Canada (karol-lina.lopez.1@ulaval.ca)

In this study, we addressed the application of Artificial Neural Networks (ANN) in the context of Hydrological Ensemble Prediction Systems (HEPS). Such systems have become popular in the past years as a tool to include the forecast uncertainty in the decision making process. HEPS considers fundamentally the *uncertainty cascade model* [4] for uncertainty representation. Analogously, the machine learning community has proposed models of multiple classifier systems that take into account the variability in datasets, input space, model structures, and parametric configuration [3]. This approach is based primarily on the well-known "no free lunch theorem" [1].

Consequently, we propose a framework based on two separate but complementary topics: data stratification and input variable selection (IVS). Thus, we promote an ANN prediction stack in which each predictor is trained based on input spaces defined by the IVS application on different stratified sub-samples. All this, added to the inherent variability of classical ANN optimization, leads us to our ultimate goal: diversity in the prediction, defined as the complementarity of the individual predictors.

The stratification application on the 12 basins used in this study, which originate from the second and third workshop of the MOPEX project [2], shows that the informativeness of the data is far more important than the quantity used for ANN training. Additionally, the input space variability leads to ANN stacks that outperform an ANN stack model trained with 100% of the available information but with a random selection of dataset used in the early stopping method (scenario R100P).

The results show that from a deterministic view, the main advantage focuses on the efficient selection of the training information, which is an equally important concept for the calibration of conceptual hydrological models. On the other hand, the diversity achieved is reflected in a substantial improvement in the scores that define the probabilistic quality of the HEPS. Except one basin that shows an atypical behaviour, and two other basins that represent the difficulty of prediction in semiarid areas, the average gain obtained with the new scheme relative to the R100P scenario is around 8%, 134%, 72%, and 69% for the mean CRPS, the mean ignorance score, the MSE evaluated on the reliability diagram, and the delta ratio respectively. Note that in all cases, the CRPS is less than the MAE, which indicates that the ensemble of neural networks performs better when taken as a whole than when aggregated in a single averaged predictor.

Finally, we consider appropriate to complement the proposed methodology in two fronts: one deterministic, in which prediction could come from a Bayesian combination, and the second probabilistic, in which scores optimization could be based on an "overproduce and select" process. Also, in the case of the basins in semiarid areas, the results found by Vos [5] with echo state networks using the same database analysed in this study, leads us to consider the need to include various structures in the ANN stack.

References

[1] Corne, D. W. and Knowles, J. D.: No free lunch and free leftovers theorems for multiobjective optimisation problems. in Proceedings of the 2nd international conference on Evolutionary multi-criterion optimization, Springer-Verlag, 327–341, 2003.

[2] Duan, Q.; Schaake, J.; Andréassian, V.; Franks, S.; Goteti, G.; Gupta, H.; Gusev, Y.; Habets, F.; Hall, A.; Hay, L.; Hogue, T.; Huang, M.; Leavesley, G.; Liang, X.; Nasonova, O.; Noilhan, J.; Oudin, L.; Sorooshian, S.; Wagener, T. and Wood, E.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. J. Hydrol., 320, 3–17, 2006.

[3] Kuncheva, L. I.: Combining Pattern Classifiers: Methods and Algorithms, Wiley-Interscience, 2004.

[4] Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J., and de Roo, A. P. J.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), Hydrol. Earth Syst. Sci., 9, 381–393, 2005.

[5] de Vos, N. J.: Reservoir computing as an alternative to traditional artificial neural networks in rainfall-runoff modelling Hydrol. Earth Syst. Sci. Discuss., 9, 6101–6134, 2012.