

The authors address a very important issue in relation to the application of data-driven models in river flow forecast. The paper is well written and organized and the proposed approach and illustrative case study are clear and useful. It is recommended to be accepted after major revision. The quality and contribution of the paper could be improved if the authors considered the reviewers suggestions and also following points:

Add more recent papers, especially Neuro-fuzzy and WNN applications in rainfall-runoff and hydrosystems , such as

Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramasastri, K.S. (2004) A neuro-fuzzy computing technique for modeling hydrological time series, *Journal of Hydrology*, 291(1-2): 52-66, DOI: 10.1016/j.jhydrol.2003.12.010.

Nayak, P.C., Sudheer, K.P., Rangan, D.M and Ramasastri, K.S. (2005) Short-term flood forecasting with a neurofuzzy model, *Water Resources Research*, 41, W04004, DOI: 10.1029/2004WR003562

Nayak, P.C., Sudheer, K.P., and Jain, S.K. (2007) Rainfall-runoff modeling through hybrid intelligent system, *Water Resources Research*, 43, W07415, doi:10.1029/2006WR004930

Krishna B, Y R S Rao and **Nayak, P. C.** (2012) “Wavelet neural network model for river flow time series” *Journal of Water Management*, 165(8): 425 –439

Nayak, P. C., Venkatesh, B., Krishna, B. and Jain, S. K (2013) “Rainfall runoff modelling using conceptual, data driven and wavelet based computing approach” *Journal of Hydrology*, 493: 57-67 DOI: 10.1016/j.jhydrol.2013.04.016

Methodology section

This section would be improved by some further explanation of the wavelet technique as used in the study (e.g. the discrete form and which form of mother wavelet has been used - Daubechies or Morlet - and define it) and further explanation of the activation functions used in the ANN and membership functions used in neurofuzzy model. How these parameters are optimized. What are their threshold values. It is not very clear to readers who are unfamiliar with the techniques.

Insufficient details on model selection. In an ideal world, each model would be able to fit any nonlinear function to an arbitrary accuracy. However, getting to the best model you can in the quickest amount of time with the least number of parameters seems to be very useful goal. Can the authors say anything about whether a certain type of model was faster to develop than others.

In the manuscript, there is no information about calibration procedure adopted in modeling. How parameters are optimized, how different models are behaving in estimating runoff process. What are optimal architectures for different models they developed. More description of model selection and generalisation is required. Which models are good in estimating peak flows.

All these models are capable of fitting a complicated nonlinear function from a suitable model architecture. Instead of comparing model behavior based on statistics listed in Table-2, I would prefer to see more discussion on what additional insight or advantage be gained from wavelet may reveal the model dynamic or physical process of the system.