

Hydrol. Earth Syst. Sci. Discuss., author comment AC2  
<https://doi.org/10.5194/hess-2021-176-AC2>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Reply on RC2

Farhang Forghanparast et al.

---

Author comment on "Deep, Wide, or Shallow? Artificial Neural Network Topologies for Predicting Intermittent Flows" by Farhang Forghanparast et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-176-AC2>, 2021

---

Reveiw2:

1. **Comment:** Artificial Neural Network (ANN) can used as a regression model to simulate streamflow as a continuous variable. This paper added a classification model on top of the regression model to simulate the flow status of intermittent streams. If the classification model outputs a zero-flow status, the flow status of the stream is decided without further running the regression model; if the classification model outputs a flowing status, the regression model will be run to predict a flowrate. Based on this idea, the authors developed two separate ANN models with different structures (wide vs. deep) to simulate streamflow for nine intermittent streams in the Texas, US, and compared the results with that from a solely regression model.

**Response: Thank you for broadly summarizing our study. However, the statement - "If the classification model outputs a zero-flow status, the flow status of the stream is decided without further running the regression model." is only correct for the proposed deep configuration. The continuous model is run irrespective of the flow status decided by the classification model in the wide configuration even when there is an outcome of no flow by the classifier. Therefore, the information flow (i.e., how the input data propagate through the architectures) is different in these two configurations.**

- **Comment:** Although the authors argued that the wide and deep models are different in their structures, I disagree and would say that the only difference is the input data to the regressor of the models: the regressor in the wide model takes all input data including both flowing and non-flowing values, while the regressor in the deep model only takes flowing values as input data. Therefore, the wide and deep models are essentially the same and the difference in results are only due to different input data. The study was actually testing the impact of different input data (a full dataset or a partial dataset) on simulation outputs. This fact compromises the whole structure of the manuscript, and the finding that the wide model that takes the full dataset as input showed better performance in simulating flowrates than the deep model that was built only on part of the input data is not surprising.

**Response: Thank you for this comment. We very much appreciate it as it has helped clarify some differences between the two architectures so that the**

presentation is less confusing.

The structure of machine learning models is based on how the Data and information flow through them (e.g., Schmidhuber, 2012). In a deep network, the information flows through a horizontally stacked (sequential) architecture while in the wide network it flows through a vertically stacked (or parallel) architecture. Our categorization follows this standard of machine learning literature and the differences in information flow were the basis for suggesting that these models have different structures.

We could have also developed another deep (horizontally stacked model) wherein the information would flow sequentially from a classifier and only invoke a regressor that would be trained on the entire dataset (both flow and no-flow) if the classifier cell resulted in a no-flow estimate. This model would be in line with the approach you are proposing here as information would flow sequentially from the classification to regression cells (With the regression cell having an ANN trained on the entire dataset). In this case, the structure of the two models would be exactly the same with only a difference in the training dataset.

Part of the confusion could also be stemming from the fact that the same set of inputs were adopted to evaluate deep and wide architectures. We adopted this approach here to ensure as much similarity between the two models and block other factors (e.g., different model architectures and algorithms) during our comparison. Even then, the differences between the two models can be seen. The two models essentially perform the same when the classifier correctly classifies no-flow (i.e., classifier controlled cases). However, they do provide different results when there is flow or there is a misclassification of flow state (regression controlled cases). The deep model, in its present configuration, does well when there are large flows (which is often seen in intermittent streams) while the wide model does well when the system is dominated by low flows.

In a deep architecture, the classification cell and regression cell need to be trained on the same dataset (as information flow is sequential). On the other hand, this need not be the case with a wide network. The wide network, therefore, has the advantage of coupling with existing data-driven and physically-based modeling tools that may be available for the intermittent stream of interest. We have modified Figure 1 (Please refer to the attached PDF, the additional information for Comment2, Reviewer2, for the updated version of Figure 1) to make this distinction clear and also mentioned the above points to draw a clear distinction between the two configurations and added verbiage to the text to make this clearer.

- **Comment:** In addition, more justification should be added to the Introduction (probably to the paragraph beginning from Line 30) to explain why a data-driven method is chosen simulate streamflow, rather than a hydrological model?

**Response:** Thank you for your comment. Data-driven models are largely preferred in intermittent streams because the assumption of continuum is strictly not valid, especially when the stream dries up. The streamflow in intermittent streams exhibits sharp discontinuities which, in turn, cause significant nonlinearity in the datasets. Being empirical in nature, data-driven

**models are not based on continuum assumption and generally exhibit a greater ability to capture nonlinearities. Both these factors suggest better suitability of data-driven models to modeling intermittent flows and were the primary factors for using them in this study. We have added relevant statements to the revised manuscript to better clarify this issue.**

- **Comment:** By the way, has the authors thought of combining the classification ANN model with a hydrological model to better simulate streamflow in an intermittent stream?

**Response: Thank you for your suggestion. The integration of a classification ANN model with a physically-based hydrological model is certainly possible in the case of a wide ANN architecture but not with a deep ANN architecture. The coupling of a hydrologic model (based on the continuum assumption) with a discrete classifier would help address the issue of modeling 'mixture data' types arising in intermittent streams with a combination of continuum and discrete classifier approaches. This is clearly one of the advantages of the wide formulation. The deep architecture exhibits greater fidelity to the mixture type data in that discrete and continuous portions of the data are modeled separately (but within the same model). The wide architecture does not exhibit complete fidelity in the sense that the discrete and continuous portions of data are used in the regression cell instead of just the continuous portion. However, the wide architecture is more practical in that existing models (both hydrologic and data-driven) based on the continuum assumption can be integrated with a classifier and improved upon. We have added this discussion to the revised version.**

**As our focus here is on comparing deep and wide architectures, the comparison with a hydrological model is clearly out of the scope here. But we certainly envision a future study that is built along these lines. We thank the reviewer for this question as it certainly helps clarify the differences between deep and wide architectures.**

- **Comment:** The structure of the Methodology needs improvement as well. Probably starting with an ANN regression model that is conventionally used to simulate streamflow, followed by the introduction of a classification model on top of the regression model.

**Response: Thank you for your comment. As single-layer ANNs are well-known and widely applied in the field of hydrology, we did not feel an introduction on them was needed in the interest of brevity. However, we agree with the author that a short discussion on MLPs for classification and regression would make the paper complete. Therefore, we have added a section in the supplementary material explaining the workings of ANNs for both regression and classification. Please refer to the attached PDF, the additional information on Comment5, Reviewer2, for the added introductory section on ANNs.**

**6. Comment:** Instead of proposing a deep and wide model, only develop one of them, since they are the same (see previous argument).

**Response:** Thank you for your suggestion. We have clarified why the two architectures are different and as such retained the presentation on both models. However, we thank you for your ideas and suggestions as a comparison of different types of wide models (e.g., data-driven and hydrologic) coupled with a classifier would be of interest to the hydrologic modeling community and expect to continue our research along the lines suggested by you.

- **Comment:** More descriptive information should be provided for the model evaluation testbeds, such as what is the calibration period / testing period, why choose that, etc.

**Response:** Thank you for your comments. We have added a statement to indicate that the first 75% of the records were used for training and the remaining 25% was used for Testing. The choice of this split was based on our goal of evaluating the proposed architectures to make short (a few months ahead) to medium-term forecasts (a few years ahead) necessary for water resources management in these streams.

In addition to the above clarifications, we have also modified Table 1 (Please refer to the attached PDF, the additional information on Comment7, Reviewer2, for the updated version of Table 1) to provide additional details pertinent to the calibration period and validation period.

- **Comments:** The caption of Figures and Tables in this study should be standing alone, with more information added.

**Response:** Thank you for your comments. All the captions of figures and tables were reviewed, and the captions were updated with more information added (Please refer to the attached PDF, the additional information on Comment 8, Reviewer2, for a list of updated captions.)

- **Comment:** As there are many comparisons made in the results, log transformed/no transformation, with/without SMOTE, continuous/wide/deep, it is very easy to confuse readers of what the main point of the study. I would suggest the authors only focus on the comparison of regression vs. regression + classification, taking the pathway of SMOTE and log transformation, since they are shown to provide better results, and other comparisons can be included as supporting information.

**Response:** Thank you for your suggestion. All results associated with the “no-transformation” mode have been removed from the main body and Table 2. That information has been moved to the supplementary material (Table S4). Please refer to the attached PDF, the additional information on Comment9, Reviewer2, for Table S4. Also, after the positive impact of SMOTE-balancing was depicted in Figure 6, only the results of SMOTE and log transformation pathway are presented, per your recommendation. Furthermore, the captions of the figures and tables were updated to clarify the results and various comparisons that are being made in each section.

Please also note the supplement to this comment:

<https://hess.copernicus.org/preprints/hess-2021-176/hess-2021-176-AC2-supplement.pdf>