

Earth Syst. Sci. Data Discuss., author comment AC5  
<https://doi.org/10.5194/essd-2022-52-AC5>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC3

Ibrahim Demir et al.

---

Author comment on "WaterBench-Iowa: a large-scale benchmark dataset for data-driven streamflow forecasting" by Ibrahim Demir et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-52-AC5>, 2022

---

### Specific Comments:

Line 17: "To some extent, WaterBench makes up for the lack of unified benchmarks in earth science research. We highly encourage researchers to use the WaterBench for deep learning research in hydrology."

**Suggestion:** I would replace this sentence with a summary of the results; it will be the good performance that will encourage.

**Answer:** We have revised this sentence as you advised.

Line 84: "*This dataset solves the difficulty of data acquisition and does not require domain knowledge from meteorology.*"

**Question:** What the authors mean by this sentence: "... does not require knowledge of the domain from meteorology?"

**Answer:** Yes, and we have updated the descriptions as you suggested.

Line 89: "*Scientific advancement, intrinsically, is supposed to be cumulative, and in order to have better generalized deep learning-based flood forecasting models, scientists need to build on top of what their fellow researchers have done.*"

**Comment:** Certainly, scientists need to build on what their fellow researchers have done, but they also have to improve on what has been done.

**Answer:** We have updated this sentence to "*scientists need to build **and improve** on what their fellow researchers have done*"

Line 90: "*We believe that this could only be done by using the same set of testing mechanisms ...*"

**Comment:** This sentence is true after being certain that everything has been done correctly.

**Answer:** Yes you are correct. We will rephrase the sentence.

Line 102: "*There remains a need for a dataset that is more convenient to use in deep learning research given that most of the deep learning researchers are not domain*

experts.”

**Question:** What the authors mean by this sentence?

**Answer:** Our paper preprocesses some complex data, builds an optimization task that computer scientists can understand, and provides optimization goals (i.e. the metrics like NSE and KGE we want to maximize). As a result, data scientists who lack meteorological and hydrological knowledge can use our data for machine learning or AI modeling. Academic collaborations are common between domain experts and data scientists. Benchmark datasets support a new type of collaboration that involves meteorologists and hydrologists processing data before passing it on to computer scientists for artificial intelligence modeling and analysis.

Line 110: *"This study proposes a flood forecasting dataset that follows FAIR data principles ..."*

**Suggestion:** I would write some references and details about FAIR.

**Answer:** We have rephrased this sentence. "Our dataset follows FAIR data principles, which is findable and accessible through DOI, and the data is richly described with references."

Line 115: *"... it could be used in other studies with similar goals, such as physically based modeling with physical equations."*

**Suggestion:** I think "with physical equations" is not needed.

**Answer:** We have rephrased this sentence.

Line 165: *"From the tables above, it is shown that our dataset is limited to a certain range of precipitation since it contains the catchments only in Iowa."*

**Question:** What the authors mean by this sentence?

**Answer:** We have removed this sentence since we updated our title and descriptions that this is WaterBench-Iowa.

Line 167: *"Geologically, all the catchments are located in two HUC watersheds, the Upper Mississippi and Missouri, and the study results may not be applicable to other regions in the U.S."*

**Question:** Do the authors think that this study is only about the analyzed area?

**Answer:** No. The trained models will be only applicable to the studied area, but the data-driven model and structures can be used in other regions around the world. We have rephrased this sentence.

Line 168: *"WaterBench is also subject to a relatively high missing data rate for streamflow since the reliable hourly dataset is limited in USGS for some of the watersheds in Iowa."*

**Question:** How have you thought about solving this limitation? How much data is missing?

**Answer:** This is mainly due to the freezing of rivers in winter, and we do not consider these data at present.

Line 174: *"Since the total precipitation amount is the product of precipitation intensity and area, larger watersheds typically have higher streamflow rates."*

**Comment:** I do not think this comment is needed, because river flow discharge depend on a number of conditions.

**Answer:** Yes. You are correct. We will modify this sentence.

Line 179: "2.2.2 *Time of Concentration.*"

**Question:** Isn't the time of concentration calculated for each grid point of the considered domain? Is the velocity not estimated as a function of slope?

**Answer:** The time of concentration is calculated based on the most distant point in the watershed to the outlet divide by an average velocity of 0.75m/s. This has been confirmed the 0.75m/s is very effective in Iowa basins and used in the following studies:

Mantilla, R., Gupta, V. K. & Troutman, B. M., "Scaling of peak flows with constant flow velocity in random self-similar networks," *Nonlinear Processes in Geophysics*, vol. 18, no. 4, pp. 489–502, 2011.

Mandapaka, P. V., Krajewski, W. F., Mantilla, R. & Gupta, V. K., "Dissecting the effect of rainfall variability on the statistical structure of peak flows," *Advances in Water Resources*, vol. 32, no. 10, pp. 1508–1525, 2009.

Sloan, B. P., Mantilla, R., Fonley, M., & Basu, N. B., "Hydrologic impacts of subsurface drainage from the field to watershed scale," *Hydrological Processes*, vol. 31, no. 17, pp. 3017-3028, 2017.

Fonley, M., Mantilla, R., Small, S. J., & Curtu, R., "On the propagation of diel signals in river networks using analytic solutions of flow equations," *Hydrology and Earth System Sciences*, vol. 20, no. 7, pp. 2899-2912, 2016.

Line 188: "A steep slope may cause a higher velocity and lower infiltration rate, which normally causes a larger streamflow rate at a precipitation event"

**Suggestion:** Therefore, it would be very important to have the velocity data for the different considered grid points.

**Answer:** It is true and your suggestion is very helpful and we will include them in the future version of the benchmark.

Line 206: "Since there were a few missing values in the original data caused by station system breakdown or internet outages."

**Question:** How is missing data handled?

**Answer :**In the dataset, we provided the original data with missing values. In our 120 hour prediction tasks, we removed the data with missing values. Users can fill missing values using interpolation.

Line 221: "2.2.7 *Evapotranspiration (ET).*"

**Question:** Do the authors believe that this comprehensive reference dataset for streamflow prediction is intended for climate or meteorological studies? For climate studies the monthly ET parameter may be useful; for meteorological-hydro studies perhaps, it would be more functional to use soil moisture data (also useful for climate studies) and at a higher temporal resolution.

**Answer:** Our goal in creating this task is to implement real-time streamflow forecasts that take advantage of the efficient performance of machine learning. Thus, we did not include many meteorological features which cannot be measured real-time or near-real-time. Your suggestion is very helpful that the soil moisture data would be useful. We will include the real-time hourly soil moisture data if there is one in our further studies.

Line 145: "Since general statistics such as mean squared error (MSE) and root mean

*squared error (RMSE) are not dimensionless...*

**Suggestion:** However, it would be interesting to see some results.

**Answer:** Since we have 125 watersheds with the streamflow ranges from 10 cfs to 1000 cfs, we only can provide dimensionless statistics. However, we will provide more analyses such as CDF in our revised manuscript.

Line 255: *"This means that there will be 120 different NSE and KGE values for different hours at each watershed."*

**Question:** What the authors mean by this sentence?

**Answer:** Since we worked on the prediction up to 120 hours in the test year, we have the KGE and NSE model efficiency results from the 1-hr ahead predictions to 120-hr ahead predictions at the test year.

Line 256: *"It should be noted that since the watersheds here are not filtered, it is possible for some watersheds to be greatly affected by human activities, including mitigation, construction, irrigation, urban drainage, etc. activities in watersheds."*

**Question:** Doesn't this sentence conflict with what is stated in line 120?

**Answer:** No. In line 120, we stated "The WaterBench is not selected based on human activities, which is a reaction to the real situation in Iowa." This means human activities in watersheds would be a challenge for prediction tasks (i.e., increasing soil erosion and more runoff with time). However, some benchmark dataset such as CAMELS only selected the rural areas without human activities.

Line 275: "Table 4. The median NSE and KGE among 125 watersheds at the prediction hour 1, 6, 12, 24, 48, 72, 96, and 120."

**Question:** what do the second and third columns indicate?

**Answer:** Sorry for the improper formatting of the table. It is now fixed. The second and third columns are NSE values.

Line 287: "Table 5. The median NSE and KGE among 125 watersheds at the prediction hour 1, 6, 12, 24, 48, 72, 96, and 120."

**Question:** what do the second and third columns indicate?

**Answer:** Sorry for the improper formatting of the table. It was an error of the table format and now fixed. The second and third columns are NSE values.

Line 297: *"The second limitation is the scale effect."*

**Comment:** Unfortunately, it is not clear from the work what scales are being considered. Would the scale effects be the size of the drainage basin?

**Answer:** Yes, the scale effects we mentioned in the manuscript are the size of the drainage basin. We will rephrase this sentence in the manuscript.

Line 300: "The scale effect observed in our benchmark indicates the prediction on small watersheds is still a challenge."

**Question:** Given the Classification of Watersheds by Size from the values in Table 2 it is evident the small to medium sized watersheds in the considered domain are a large number. I do not understand the statement made in line 300.

**Answer:** We have around half of the basins over 1000 km<sup>2</sup> in Table 2, and it is observed that the larger watersheds are better overall performance from Figure 6.

Line 315: *"Although the data is limited to the Midwest, we believe that any studies on this dataset could provide insights for other streamflow forecasting and rainfall-runoff modeling studies at other watersheds."*

**Comment:** This sentence written in this way is contrary to what is stated in line 168.

**Answer:** We have rephrased the sentence in Line 168 in our revised manuscript.