WaterBench: A Large-scale Benchmark Dataset for Data-Driven Streamflow Forecasting

Response to Reviewers' Comments

ESSD-2022-52 | Data description paper
Submitted on 09 Feb 2022
Ibrahim Demir, Zhongrun Xiang, Bekir Demiray, and Muhammed Sit
Special Issue: Benchmark datasets and machine learning algorithms for Earth system science data (ESSD/GMD inter-journal SI)

Topical Editor decision: Publish subject to minor revisions (review by editor)

Comments to the author:

Dear authors,

thank you for submitting a detailed response to the 3 reviewers and a revised manuscript. Before accepting this paper for publication, I have a couple of minor points, which I ask you to improve:

1) the zenodo record which you refer to should contain a README file describing what exactly users will find in the csv and zip files. This description should include a link to the code repository and a link to itself (i.e. the zenodo doi) as well as to the paper in ESSD. And it should include a (brief) description of the file structure and format.

Answer: Thank you for the suggestion, and we have updated the Zenodo record with a readme file.

2) according to the ESSD publication rules, code must be provided in a non-alterable form, for example as another zenodo dataset with a doi. You can either upload a zip or tar archive of your code from git to the same zenodo record that contains the data, or you can create a new record. The link in the abstract should not point to a github repo. The github repo can be mentioned in the "Data and Code Availability" section like "The most recent code version can be found at https://github.com/uihilab/WaterBench.".

Answer: We have updated the abstract, data and code availability section as you suggested.

3) if you don't mind, I would suggest rephrasing the sentence on the domain knowledge again: "This dataset solves the difficulty of data acquisition and does not require domain knowledge of the domain of meteorology." to "While knowledge of the application domain is essential to find scientifically robust ways to prepare the input data and to interpret the results of machine learning models, such knowledge is not always accessible to deep learning experts. If there are well-defined benchmark datasets with a clear description of the machine learning task to solve and with well-defined and domain-science informed evaluation metrics, then it becomes possible for non-domain experts to solve such challenges and to introduce novel machine learning methods to the field."

- if you accept this suggestion, you may be able to shorten the text between lines 97 and 104 in your track changes manuscript.

Answer: We have rephrased and shortened the sentences as you suggested.

4) line 124: you can add https://www.nature.com/articles/sdata201618 as reference to FAIR data.

Answer: We have updated this citation.

5) the benchmark task itself should appear in the abstract - You can move the sentence "we define a sample benchmark task of predicting the hourly streamflow for the next five days for future comparative studies." from line 264 to the abstract and slightly rewrite the next sentence "The benchmark task defined here simulates rainfall..."

-- please rephrase that sentence, though, as "simulate ... forecasts ... real life" don't go well together.

Answer: We have updated the abstract, and rephrased the sentence as you suggested.

6) Acknowledgements should include the source of funding of this study and possibly a comment to thank the reviewers.

Answer: We have updated the acknowledgement as you suggested.