

Summary:

The paper “Streamflow prediction in data-scarce regions with semi-supervised deep learning” presents a two-stage semi-supervised learning framework for streamflow prediction in regions where observed discharge data are limited. The authors use Contrastive Predictive Coding (CPC) to pre-train an encoder and LSTM network using unlabeled meteorological forcing data, and then fine-tune the model with a limited amount of labeled streamflow data.

Overall, this is a timely and valuable contribution. The paper is methodologically sound, clearly motivated, and relevant to hydrological prediction in poorly gauged or ungauged basins. The results support the usefulness of semi-supervised learning for data-scarce streamflow prediction. Therefore, I believe the manuscript is worth publishing after minor revisions.

Response: Thank you for taking the time to review our work and provide useful comments which have helped us improve the quality of our manuscript.

Minor comments:

1. The authors should more clearly explain why CPC was selected over other self-supervised learning methods, and briefly discuss whether alternative SSL methods may perform differently.

Response: Thanks for the comment. We will add the motivation for selecting CPC over other self-supervised learning methods for time series forecasting tasks. The main reason is that CPC’s predictive mechanism is closely aligned with the streamflow prediction task, predicting future states from past observations. Thus, we hypothesized that CPC could be effective for streamflow prediction tasks as an initial exploration, while we did not investigate the effectiveness of other self-supervised learning methods in this study. We will also add a brief discussion of another self-supervised learning method in the Future Work section. This method could provide an alternative approach for leveraging unlabeled data (i.e., meteorological forcing time series), with a different learn mechanism from the CPC we used. Its performance may therefore differ from that of CPC, and its effectiveness needs future investigation.

2. The study uses CAMELS-DE, which is a high-quality dataset from Germany. The authors should discuss more explicitly how well the proposed method may transfer to regions with poorer data quality or stronger hydroclimatic variability.

Response: We agree that the CAMELS-DE dataset is a high-quality dataset with large amount of meteorological forcing and observed streamflow time series. However, this study initially focuses on the methodological development of SSL for improving streamflow prediction in data-scarce regions, rather than investigating how data quality affects the effectiveness of SSL methods. A systematic investigation of this effect would be valuable and is left for future work, as it is beyond the scope of the present study.

While we did not investigate the scenario involving low-quality data, we hypothesize that our proposed SSL methods still could still outperform baseline LSTM models. If only a limited amount of low-quality observed streamflow data is available in a target basin, baseline model can only leverage these data to learn data representations, that may result in limited prediction performance. In contrast, our SSL methods can additionally leverage a large amount of unlabeled meteorological forcing time series to learn representations that are effective for downstream streamflow prediction training. This advantage of SSL result in an improvement in performance compared to the baseline models.

Furthermore, we have highlighted the potential of SSL for streamflow prediction in data-scarce regions in Section 4.5 (Implications for streamflow prediction). Such regions may also be characterized by low-quality observational data, with issues such as temporal discontinuities and sparse spatial coverage.

3. The results show that SSL performs best when labeled data are very limited, while supervised learning performs better when more labeled data are available. The authors should emphasize this practical threshold more clearly in the conclusions.

Response: In the Conclusion section, we highlighted the observed threshold in our experiments (see the first conclusion). The SSL method improves performance of regional models compared to the baseline SL method for data-scarce conditions, when less than 10% of one-year labeled sequences are available for training. However, the SL baseline slightly outperforms the SSL method under data-rich conditions, when $\geq 10\%$ of one-year labeled sequences are available.

4. The manuscript reports improvements for flood and low-flow prediction, but the discussion of extreme-event performance could be slightly expanded, especially regarding cases where both SSL and SL underestimate flood peaks.

Response: Thanks for the comments. We will discuss the events shown in Figure. 9 in greater detail, as both methods underestimate the peak flow. We will provide our analysis of the possible reasons for this underestimation.

5. The authors mention limited computational resources and no detailed hyperparameter tuning for CPC pre-training. This limitation should be stated more clearly in the main discussion, as model performance may depend on pre-training duration and parameter choices.

Response: In the Results and Discussion Section, we have summarized and highlighted the limitations of this work in a sub-section named “4.4 Limitations and future work”. We indicated that we did not perform hyperparameter fine-tuning for CPC pre-training stage (e.g., the number of pre-training epochs), due to limited computational resources (1 hour per pre-training epoch). Despite these constraints, the proposed approach achieves competitive performance, demonstrating the potential effectiveness of the methodology for streamflow prediction in data-scarce regions. Further performance improvement is expected through

scaling both the pre-training dataset, and the computational effort devoted to hyperparameter optimization. If anything remains unclear or needs to further highlight, please feel free to tell us, and we will further clarify.

6. I strongly recommend the authors citing “Analysis of historical global warming impacts on climatological trends for the partially gauged Hirmand River Basin based on multiple data products and bias correction methods” because it is relevant to streamflow and hydrological prediction in partially gauged/data-scarce basins.

Response: Thanks for the recommendation. However, after carefully reviewing the suggested paper, we found that it is not directly aligned with the main focus of our manuscript, i.e., proposing a Semi-Supervised Learning (SSL) method to improve streamflow prediction under data-scarce conditions. Although the suggested study also addresses the challenge of limited observational data, it tackles this issue using a data pre-processing approach that fills spatial gaps through interpolation. In contrast, our study does not employ any data pre-processing or spatial interpolation to address data scarcity. Therefore, we believe that the suggested paper is not sufficiently relevant to the methodological focus of our study.