Can discharge be used to inversely correct precipitation? (hess-2024-375)

Ashish Manoj J, Ralf Loritz, Hoshin Gupta, Erwin Zehe

The authors would like to thank Anonymous Reviewer 3 for going through our manuscript and giving their critical comments and suggestions. The following responses have been prepared to address all the reviewers' comments point-by-point. We have responded (in black) to the reviewer's comment (in blue).

General comments:

The paper by Manoj et al. describes a method to estimate catchment scale precipitation from streamflow records using a machine learning approach. The paper is well-written, and its objective is highly significant in the context of hydrological sciences, where precipitation data remain scarce and critical to improve water resources modelling. Using streamflow as a predictor to estimate rainfall is not new, but it makes perfect sense as streamflow reflects recent rainfall history. The LSTM model is perfectly justified for this task, considering the high level of performance reached by this type of machine learning approach. We particularly appreciated exploring a large sample of catchments, which reinforces the author's conclusions. We were also impressed with the final validation exercise using different hydrological models.

Overall, this paper contains many valuable elements and a tremendous amount of work. However, it suffers from two fundamental flaws requiring a major revision before its acceptance for publication:

We thank the reviewer for highlighting the work's significance and summarizing the main strengths. We propose to make the following changes to address the well justified concerns raised by the reviewer.

Comment #1: The fundamental aim of the paper stated in the introduction is to generate better precipitation estimates compared to currently available reanalysis products such as ERA5-Land. The authors are clear about the issues of reanalysis products in various parts of the manuscript. For example, related to a particular flood event, they indicate: "Our previous work (Manoj J et al., 2024) indicated that ERA5 Land could not accurately replicate the characteristics of the convective storm that caused this annual flood event" (line 240). Consequently, we do not see the point in training an LSTM using ERA5-Land precipitation as a target. The best we can expect from this approach is to generate rainfall series identical to ERA5-Land precipitation, which is known to be problematic.

We agree that the ERA5 Land has issues representing the driving precipitation estimates for specific event scales (Essou et al., 2016; Manoj J et al., 2024). As stated in the beginning of our abstract, our aim was to see whether the inverse data assimilation using streamflow information could be used to overcome at least some of these well documented deficiencies.

a) Our approach, utilizing a regional LSTM model trained on much larger catchments, demonstrated effectiveness in adjusting the underestimated precipitation values for these events at smaller (out of sample) scales. Notably, only about 9% of the catchments in our training dataset had areas smaller than 100 km^2. We could show that discharge response encodes sufficient information about the driving precipitation to correct ERA5 Land in the right direction.
b) Reanalysis data, by definition, are a mix of observations and past short-range weather forecasts rerun with modern weather forecasting models (ECMWF, 2023). Different data assimilation methods are used for this. Our idea was that the inversion technique could be used as another final layer of post-processing (using the LSTM in this case) for the model outputs to ensure that the final product is more consistent with the variabilities observed in the discharge record.

Furthermore, any performance comparison between LSTM outputs (trained on ERA5-Land precip) and original ERA5-Land precip using an independent dataset as a reference (E-OBS in this case) are logically flawed: the LSTM was not trained to reproduce anything else than ERA5, so any perceived "improvement" between its outputs and ERA5-Land precip when simulating an independent dataset (E-OBS in this case) is due to chance. Fortunately, the solution to this problem is simple: instead of ERA5-Land, the authors could set the LSTM training target to rainfall observation (i.e. E-OBS). The comparison between LSTM and ERA5-Land would become meaningful and clarify if precipitation estimation can be improved compared to using ERA5-land.

We would like to clarify that our goal was to enhance the ERA5 Land estimates by incorporating streamflow information along with only other meteorological forcings from ERA5 Land, rather than generating a new precipitation product that uses again another precipitation as input.

c) It is important to emphasize that "true" precipitation estimates are only available at observational stations and not at the catchment scale. The performance comparison using EOBS and the runoff coefficient was intended to provide insight into the feasibility of different precipitation estimates from a hydrological perspective. While we acknowledge the existence of even better regional products (e.g., HYRAS – German Weather Service) for some of the study catchments, we believe that these various products should not be viewed as independent of one another. Instead, they contain complementary information as they represent the same physical truth i.e. precipitation occurring over a catchment, albeit with different uncertainties and errors.
d) Studies evaluating daily precipitation from EOBS and ERA5 over Europe (Bandhauer et al., 2022) have shown that while E-OBS is superior to ERA5 in regions with dense data, using ERA5 has advantages in data scarce regions. This was also seen in the out of sample analysis. For the Sueiro catchment (camelses_1414), the closest observational station is located more than 60 km away (Figure 1 in this document), this explains why the EOBS performs rather poorly in representing the driving forcings for the summer flood event (Figure 5C in original draft). Additionally, the runoff coefficient estimate for E-OBS was around 1.05, which indicates a hydrologically infeasible value (Table 2 in the original draft) when compared to the estimates from ERA5 Land and LSTM. Compared to purely interpolated products like EOBS, reanalysis products are usually released and updated more frequently. This again points to the value of reanalysis products like ERA5 for tackling the *prediction in ungauged basin* (PUB - Hrachowitz et al., 2013) problem.
e) While our workflow could indeed be extended in multiple directions to generate more coherent precipitation products, we feel this is currently beyond the scope of our initial study, which aimed to explore whether discharge had sufficient information to help in tackling the inverse problem over a large sample space.

We will clarify these points and restructure the Introduction and Discussion sections to highlight them more effectively.

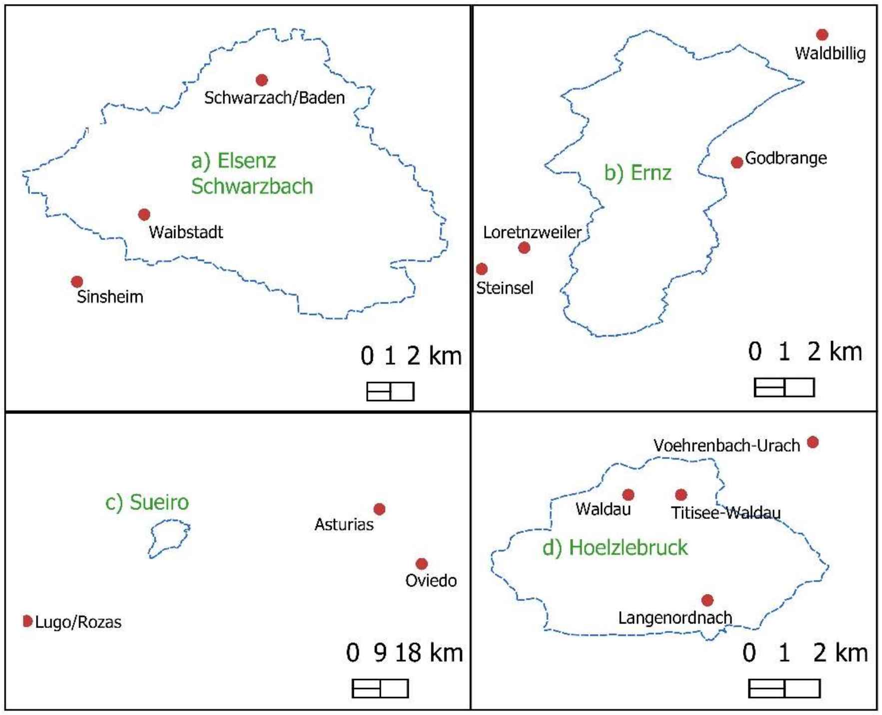

Figure 1 Spatial maps showing the proximity of observational stations (used for deriving the EOBS gridded product) to the four out of sample catchments considered in the present study.

Comment #2: When training their LSTM, the author used the mean catchment rainfall as a predictor (Pmean, see line 138). In other words, they use some of the predictand data as a static predictor. This is a major flaw in a regression setup: it gives the LSTM model a distinct advantage over an operational situation where, obviously, the mean catchment rainfall is not known. Here again, the solution to this issue is straightforward: remove this predictor from the list of static predictors.

While we used the LSTM model in the commonly used streamflow prediction (regression) mode, it is important to note that our end goal is different from an operational forecast. The model takes in future streamflow as a predictor, which implies that the real-time forecast implications of our methodology are limited. The approach could be seen as a data assimilation post-processing step to ensure that the final precipitation estimates are more consistent with the variabilities observed in the discharge record. We would like to highlight that hydrological modellers have previously constrained their models (Gharari et al., 2021) using average annual runoff to improve calibration for the classical streamflow prediction problem.

To address the reviewer's comment regarding the removal of *p_mean* from the list of static attributes due to concerns about data leakage, we retrained and tested the regional scale LSTM model, removing *p_mean* while keeping all other conditions the same.

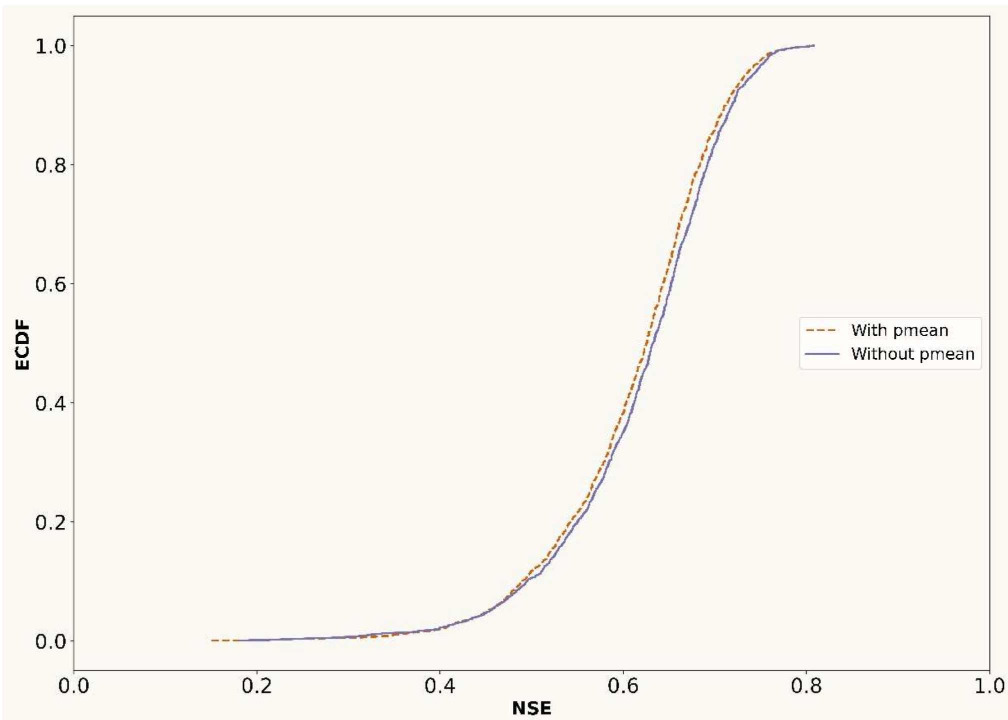

Figure 2 Empirical Cumulative Distribution Function (ECDF) of NSE values, comparing model performance in runs with and without *p_mean*.

The ECDF plot (Figure 2) indicate that the model performance remains fairly consistent for the two runs. We believe that this can be attributed to two main reasons:

A. Our training dataset had a large number of catchments (1804) cutting across various hydroclimatically diverse regions over Europe, ensuring that our model could learn robust dependencies from the meteorological (dynamic) forcings itself.
B. Recent research (Heudorfer et al., 2024; Li et al., 2022) on Entity Aware (EA) deep learning models (models that are provided with static features- predominantly in the form of physiographic proxies, next to dynamic forcing features) have suggested that the information in static features is not being effectively leveraged.

Li et al. (2022) demonstrated that an LSTM model using randomly initialized numbers as static features outperformed a model that used actual physiographic static features, as long as the number of random static features was greater than the number of physiographic static features. This indicates that the specific physiographic characteristics of the static features may be irrelevant; what truly matters is the presence of unique identifying information. Heudorfer et al. (2024) report that while static features serve as unique catchment identifiers, resulting in excellent in-sample performance when confronted with out-of-sample data, the model is unable to generalize from static features and instead relies almost exclusively on meteorological data for prediction.

Aside from these two fundamental problems, we also have a few general comments:

Comment #3: some aspects of the method lack clarity. We got a bit lost in all the cases considered by the authors at the end of the manuscript. We suggest clarifying several elements using summary tables in the method section (and not later in the paper):

- the list of all LSTM configurations tested with their inputs (including lagged inputs) and their outputs,
- the list of all performance metrics,
- the list of all test cases including the number of catchments, the forcings (if using hydrological models) and the outputs tested.

We will include tables in the Data and Methods section that detail all the datasets used, as well as the various model runs and test cases. To enhance readability, we also plan to provide information about the LSTM configurations and the different hydrological models in a new appendix section.

Comment #3: The LSTM model was trained on mean squared error, emphasising large rainfall values. We recommend testing other configurations where training is done on transformed values, e.g. square roots and log transform, to check if certain rainfall metrics can be improved further.

We appreciate the reviewer's suggestion that testing additional functions could enhance some of the rainfall metrics. However, since the primary focus of our study was on accurately representing heavy precipitation events that lead to floods, we chose to use the mean square error training function because it was the simplest and most commonly used. We will include this information and suggest exploring other configurations as part of future research in this direction.

Detailed comments

Comment #4: Line 68, "we conjecture that the catchment-average precipitation can be inversely identified": this problem is still numerically ill-posed due to catchment memory. We suggest rephrasing to "we conjecture that streamflow data can reduce the uncertainty associated with this process by providing valuable information on recent rainfall history".

We will rephrase this line as suggested by the reviewer.

Comment #5: Line 115, "The Caravan dataset uses the ERA5 Land as meteorological forcing": it would be useful to remind that this is far from satisfactory as ERA5 is known to have important limitations when simulating rainfall.

This information will be added to the revised draft.

Comment #6: Line 185, "model "with_discharge" outperforms the model "without_discharge" not only on average but also concerning the best-performing catchments.": It would also be useful to show the distribution of pairwise NSE differences. This would answer the question: "How many catchments reach better NSE when using streamflow predictors?".

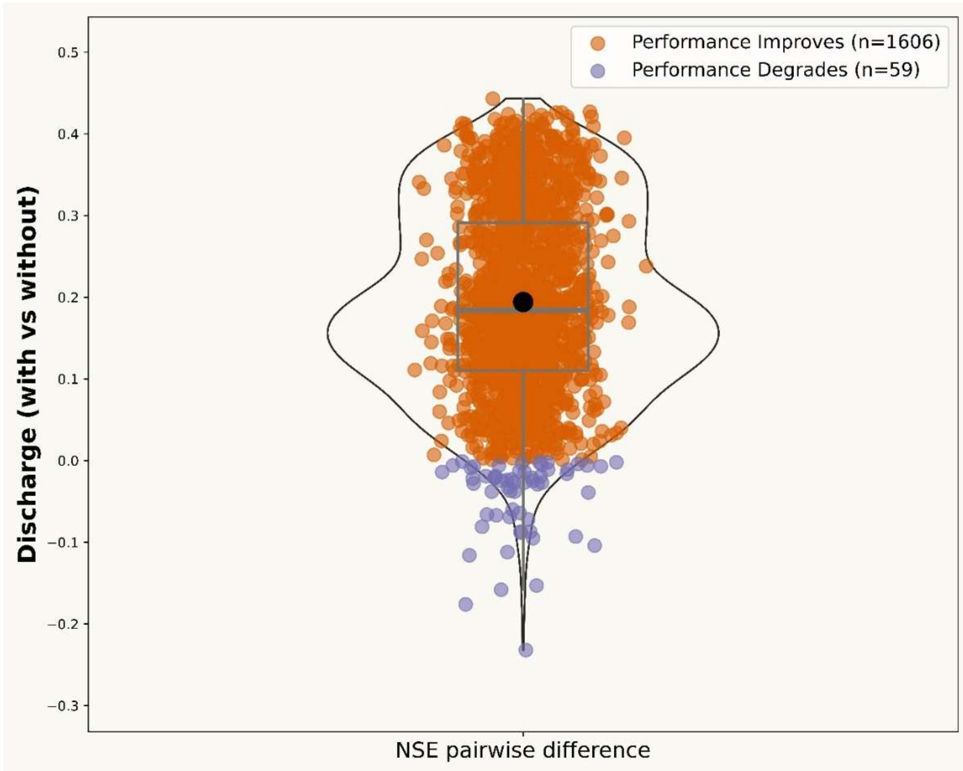

Figure 3 Violin plot displaying the pairwise differences (*with_discharge* vs *without_discharge* models) in NSE for the study catchments.

We agree with the reviewer's suggestion that displaying the distribution of pairwise NSE differences would better clarify our main results. Therefore, we will include pairwise NSE difference plots (Figure 3 in this document) in the main manuscript and move the distribution and ecdf plots (originally Figure 2 in the manuscript) to a new Appendix section dedicated to the LSTM models.

Comment #7: Figure 3: This map is a bit confusing because for LSTM and ERA5-Land, the data generated by the authors is in the form of points (i.e. catchment average), not surfaces. Please update the map accordingly.

Although we only have average information for the catchment area, these averages are derived from grid points that encompass the entire area. We believe that representing them as point data at the catchment outlets would not accurately reflect the fact that the information represents the whole catchment area.

Comment #8, Line 204 “preserves spatial gradients”: what do the authors mean by “preserve”? Please clarify. It is hard to assess spatial patterns from maps as small as Figure 3. We suggest an additional metric and a figure to clarify this point.

We agree that it is indeed hard to assess the spatial gradients without additional metrics. Since evaluating such spatial gradients is not the main focus of the present study, we will remove this sentence from the revised draft.

Comment #9, source code: please list software requirements in the source code. This includes the list of software packages required and their versions. If the authors are using Anaconda, it can be done by adding to their repository a conda environment configuration file, also referred as “yml” file (conda contributors, 2025), which lists all Python package and their version.

A dependencies file detailing all the software packages and their versions will be added to the GitHub repository.

References

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O.E., Frei, C., 2022. Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. Int. J. Climatol. 42, 727–747. https://doi.org/10.1002/joc.7269

ECMWF, 2023. Factsheet on Reanalysis [WWW Document]. 2023. URL https://www.ecmwf.int/en/about/media-centre/focus/2023/fact-sheet-reanalysis

Essou, G.R.C., Sabarly, F., Lucas-Picher, P., Brissette, F., Poulin, A., 2016. Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling? J. Hydrometeorol. 17, 1929–1950. https://doi.org/10.1175/JHM-D-15-0138.1

Gharari, S., Gupta, H. V., Clark, M.P., Hrachowitz, M., Fenicia, F., Matgen, P., Savenije, H.H.G., 2021. Understanding the Information Content in the Hierarchy of Model Development Decisions: Learning From Data. Water Resour. Res. 57. https://doi.org/10.1029/2020WR027948

Heudorfer, B., Gupta, H., Loritz, R., 2024. Deep Learning Models in Hydrology Have Not Yet Achieved Entity Awareness. https://doi.org/10.22541/essoar.172927199.90156076/v1

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H. V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)-a review. Hydrol. Sci. J. 58, 1198–1255. https://doi.org/10.1080/02626667.2013.803183

Li, X., Khandelwal, A., Jia, X., Cutler, K., Ghosh, R., Renganathan, A., Xu, S., Tayal, K., Nieber, J., Duffy, C., Steinbach, M., Kumar, V., 2022. Regionalization in a Global Hydrologic Deep Learning Model: From Physical Descriptors to Random Vectors. Water Resour. Res. 58. https://doi.org/10.1029/2021WR031794

Manoj J, A., Loritz, R., Villinger, F., Mälicke, M., Koopaeidar, M., Göppert, H., Zehe, E., 2024. Toward Flash Flood Modeling Using Gradient Resolving Representative Hillslopes. Water Resour. Res. 60. https://doi.org/10.1029/2023WR036420