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

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

Dear Dr. Roger Moussa, 08.07.2025

Attached, please find the revised version of the manuscript "*Can discharge be used to inversely correct precipitation?"* co-authored with R. Loritz, H. Gupta and E. Zehe, to be considered for publication in Hydrology and Earth System Science.

After carefully reviewing the two comments from Anonymous Reviewer 3 in the previous round, we have decided to implement both the changes suggested by the reviewer. We downloaded and pre-processed the observational EOBS gridded data for all the catchments included in our training dataset. Next, we repeated the entire analysis using this observational product as the new training target. Additionally, we removed mean precipitation (*pmean*) from the list of static attributes in this updated setup. Lastly, we have made other minor changes facilitated by a fresh reading.

Our main finding remains largely unchanged in the updated version. We observed that incorporating discharge information improved the performance of the LSTM network during the unseen testing period and resulted in more hydrologically consistent storm estimates in the out-of-sample catchments. Additionally, the forward modelling using traditional hydrological models once again produced higher mean NSE values for the runs based on inversely generated precipitation estimates.

We would like to thank the Editor and Anonymous Reviewer 3 again for giving us another opportunity to revise our manuscript.

Please get in touch with me if you need any additional information.

Thank you very much for your consideration.

Best regards,

Ashish

On behalf of Ralf, Hoshin and Erwin

Email: ashish.jaseetha@kit.edu

Reviewer 3:

The authors would like to thank Anonymous Reviewer 3 for carefully reviewing our manuscript and providing their critical yet helpful and detailed comments. We have followed the reviewer's suggestions concerning both the comments. 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:

Following a first round of review, this review comments on a second version of the paper by Manoj et al., which describes a method to estimate catchment-scale precipitation from streamflow records using a machine learning approach. We reiterate positive comments we made in our first review, including how significant the paper is to practical hydrological problems and the use of large datasets to validate the results. We feel that the detailed comments indicated in our first review have been addressed satisfactorily by the authors. As a result, we won't mention them in this review.

Unfortunately, the authors have brushed aside our two main comments, despite the fact that we offered simple alternatives, and answered them with cosmetic changes in their manuscript as explained below. We respect their opinion, but cannot approve the publication of their paper under these conditions. As a result, we recommend a major revision of the paper.

The following sections clarify our position regarding the two fundamental points we raised in our first review.

Comment #1: Our first fundamental criticism of the paper is that it aims to generate an improved ERA5-Land precipitation product (referred to as ERA5-P hereafter) using a statistical model, but trains this model to reproduce ERA5-P. When the training is completed, the model being imperfect will introduce errors in its prediction, hence producing a contaminated version of ERA5-P (ERA5-P plus residual errors from the statistical model). The whole premise of the paper is to suggest that this contaminated version is significantly superior to the original ERA5-P. In other words, the residual error of the statistical model contains valuable information, even though it is being minimised by the training algorithm.

What we said in our first review is that this is possible, but it would be due to chance because the authors expect to find something valuable out of what the training algorithm discards. What is likely to happen is that this lucky outcome may be due to the author's focus on selected rainfall metrics. Other metrics may show that the contaminated ERA5-P is worse than the original ERA5-P, for example, rain event timing, long-term seasonality, sensitivity to warming climate conditions, zero rainfall simulations, etc… To summarise, what is lacking in the paper is a clear definition of a reference precipitation dataset that all alternative rainfall products (ERA5-P and LSTM outputs) are measured against, and try to replicate. In our previous review, we suggested the use of E-OBS as a training target, because this dataset is often used as a reference in the paper. In their response to our comment, the authors objected that "while E-OBS is superior to ERA5 in regions with dense data, using ERA5 has advantages in data-scarce regions". We agree with this statement, but believe that it relates to the applicability of the algorithm, which should not precede a thorough and logically robust testing in a controlled environment.

Overall, we repeat our request for the selection of a reference observed rainfall precipitation dataset to be used as a training target for the LSTM and computation of all performance metrics in the paper.

We have implemented the reviewer's suggestion by using the observational EOBS product as the target for our model runs. Since the original Caravan dataset (Kratzert et al., 2023) did not include the EOBS estimates for the training catchments; we preprocessed the data to derive the average precipitation estimate for each catchment. We then repeated our experiments using an ensemble network of three LSTM models (with different initialisation seeds) and report the mean results for both *with_discharge* and *without_discharge* runs. Our main finding remains largely the same in the new updated version.

In addition to looking at the gain over all the days, we also explored the performance gain over days with higher magnitude precipitation (shown in Figure 1). We could see that gains are considerably greater on days with higher recorded precipitation (increase in median NSE value of about 29% from 13% for days with more than 5 mm precipitation). This is logical because the discharge information is more effective in capturing extreme conditions. In contrast, the information gain is limited under average flow conditions.

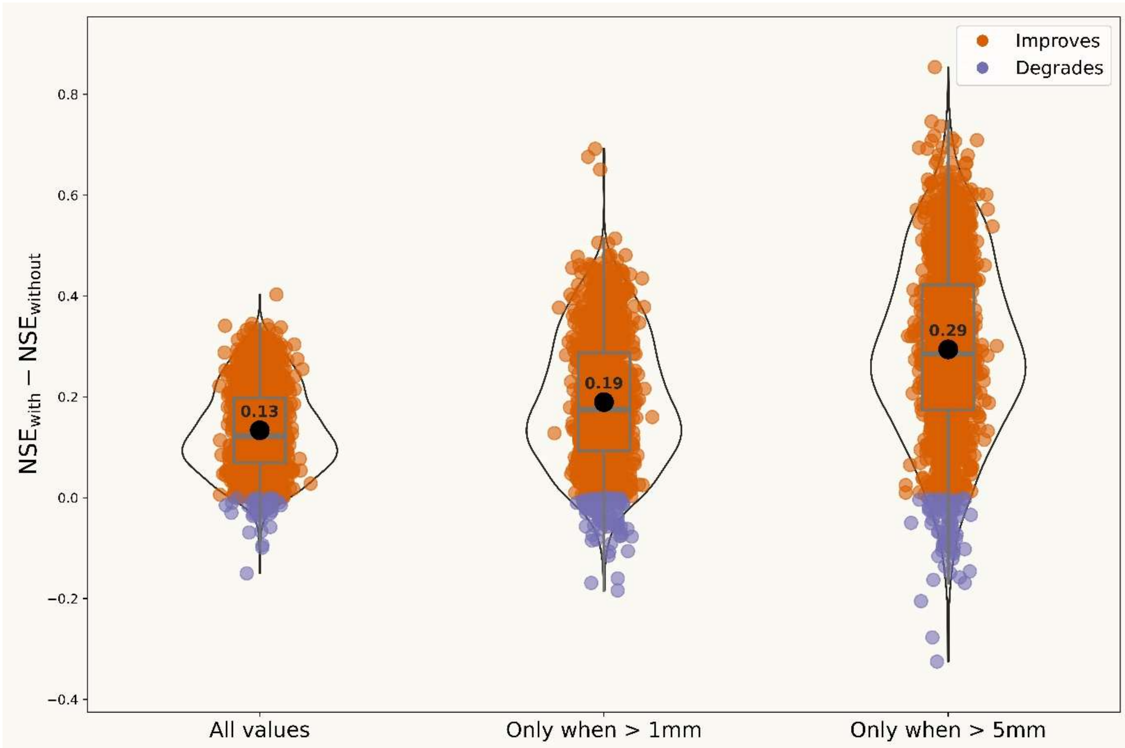

Figure 1 Comparison of performance gain for the with_discharge vs without_discharge models in NSE for different precipitation amounts. The first violin plot illustrates the average improvement across all days in the testing period. The second and third plots display the mean performance gains over the catchments, specifically focusing on days where precipitation exceeded 1 mm and 5 mm, respectively.

For the continental analysis, we again calculated all the performance metrics and now compare both the *with_discharge* and *without_discharge* models to EOBS and ERA5 Land (Figure 2). The predictions from *without_discharge* model are also added for the out-of-sample analysis (Figure 3). For the out-of-sample predictions, we again observe that the LSTM estimate overestimates the EOBS value (new training target) in three out of four catchments; the runoff coefficients (Table 1) and timing of the peaks again point to the overall reliability of the estimate.

The forward hydrological model runs using HBV and CATFLOW were also repeated for the new estimate from the *with_discharge* model, and we again observed higher NSE values over the evaluation period compared to runs with ERA5 Land.

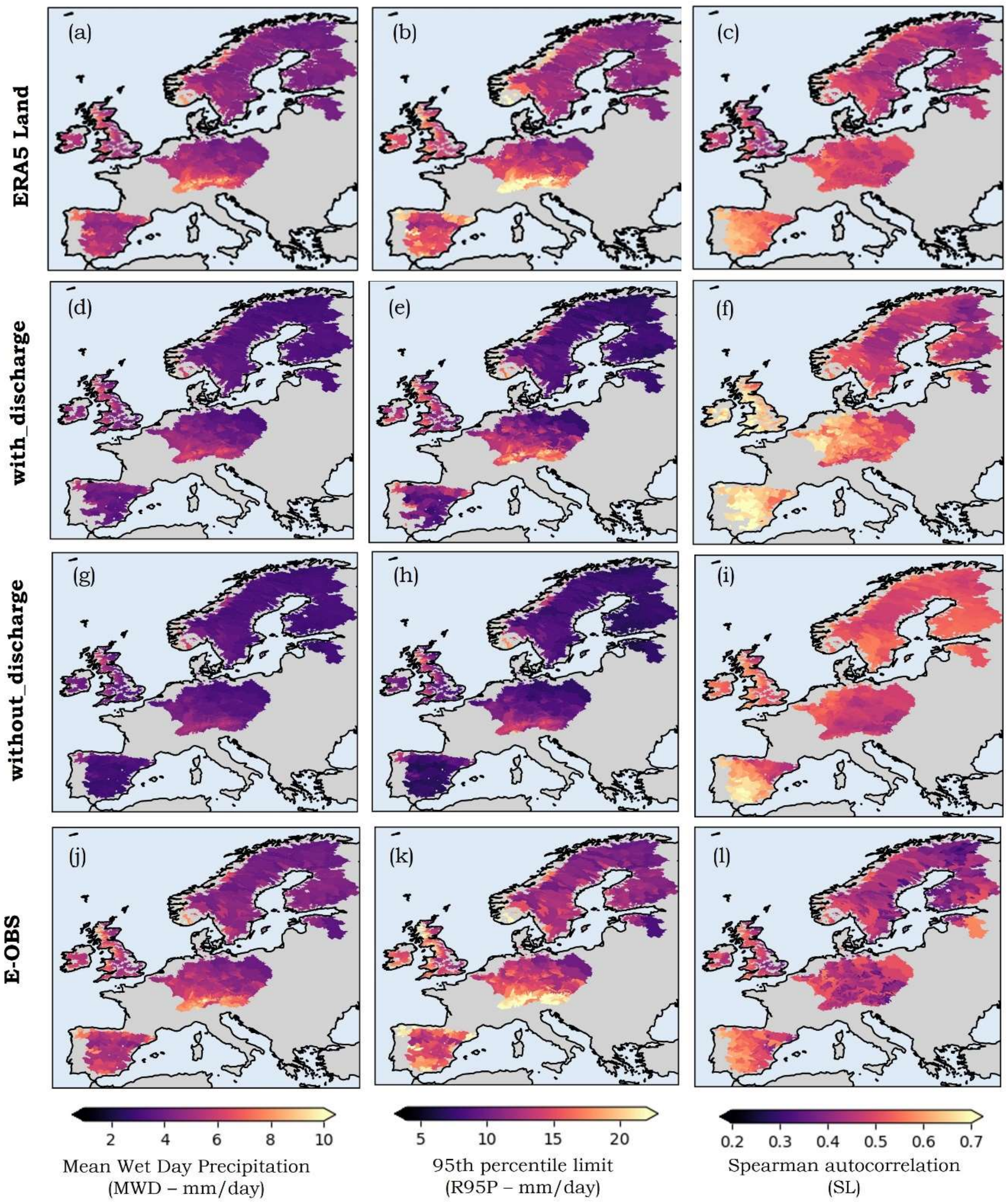

Figure 2 The spatial patterns of the different time series metrics (Appendix C) mean wet day precipitation (MWD) – mm/day, 95th percentile limit (R95P) – mm/day, and Spearman autocorrelation values (SL) over the study catchments for the different precipitation estimates - ERA5 Land (top row): a) to c), *with_discharge* LSTM model (second row): (d) to (f), *without_discharge* LSTM model (third row): (g) to (i) and E-OBS (bottom row): (j) to (l) from 2006 to 2020 (2015 for CAMELS-GB catchments).

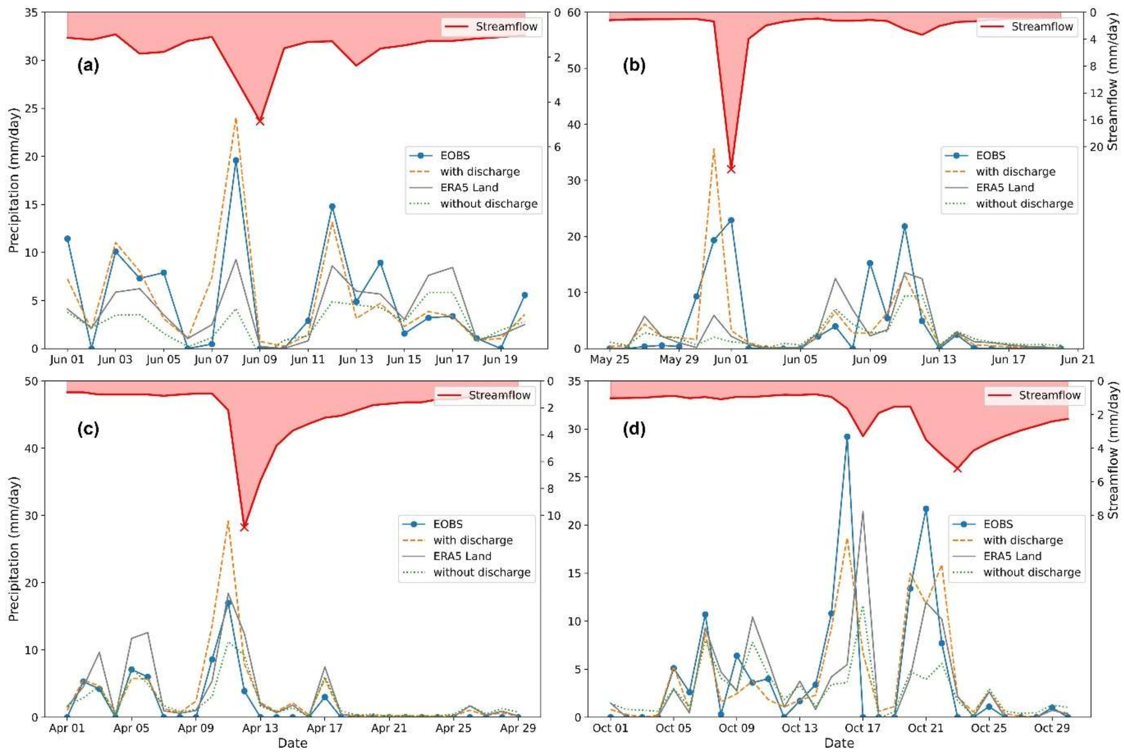

Figure 3 Precipitation estimates for flood events at four out-of-sample catchments: (a) Elsenz Schwarzbach, (b) Ernz, (c) Sueiro, and (d) Hoelzlebruck. The red line represents the observed daily streamflow, with a cross marking the day of the flood event. The orange curve indicates the precipitation predicted by the *with_discharge* LSTM model, while the green curve shows the precipitation predicted by the *without_discharge* model. The blue line reflects the original gauge-based EOBS time series, and the grey line represents the estimate from the ERA5 Land.

Table 1 Event characteristics (storm volume and runoff coefficients) for the four out of sample catchments

Event Characteristics		**Elsenz-Schwarzbach**	**Ernz**	**Sueiro**	**Hoelzlebruck**
Precipitation (mm)	ERA5 Land	12.51	9.60	41.81	32.12
	with_discharge	32.79	42.75	58.53	50.85
	without_discharge	4.92	6.20	29.46	22.92
	E-OBS	20.07	51.72	29.50	44.90
Discharge (mm)		5.98	26.88	23.39	19.14
Runoff Coefficient (-)	ERA5 Land	0.48	2.80	0.56	0.60
	with_discharge	0.18	0.63	0.40	0.38
	without_discharge	1.21	4.34	0.79	0.84
	E-OBS	0.30	0.52	0.79	0.43

Comment #2: The second fundamental comment made in our previous review related to the use of the mean catchment rainfall derived from ERA5-P as a predictor in the authors' statistical model. At the same time, ERA5-P is the training target of the author's statistical model. As a result, ERA5-P derived data are parts of both predictors and predictands. This is a fundamental flaw in statistical modelling, which cannot be accepted if aiming at publishing in a scientific journal such as HESS. At the same time, the authors have maintained the use of this predictor in their revised manuscript.

In their response, the authors first argued that this issue is relevant to real-time forecasting ("it is important to note that our end goal is different from an operational forecast"). We disagree with this view. The issue we are raising here relates to the problem of predicting output data while using part of this data as a predictor. This creates a risk of inflating model performance compared to a realistic use of the model, where the output data is not available, by definition (otherwise, a prediction model would not be required). In addition, this configuration prevents the model from being realistically used: mean catchment rainfall is one of the core data one would expect to extract from a rainfall product. If this data is required as part of the model inputs, we do not see much use of the LSTM rainfall product presented by the authors.

Overall, we repeat our request to remove mean catchment rainfall as a predictor in the paper. We believe this is extremely simple to do, as most computations have already been done by the authors.

We appreciate the reviewer's detailed comments regarding the possible short comings of using mean catchment rainfall derived from ERA5-P as a predictor. Our initial choice to include static attributes was based on incorporating climatic indicators relevant to our catchments, and it made sense to include mean precipitation. Additionally, we were influenced by other hydrological modelling studies (Gharari et al., 2021) that utilised average runoff information to enhance calibration for the traditional streamflow prediction problem.

After considering the detailed concerns raised by the reviewer and the editor regarding the statistical modelling setup that used LSTM, we have decided to exclude *pmean* from the paper. We have re-run all our model simulations with the new setup, which does not include *pmean* as a predictor. The results align with previous studies (Heudorfer et al., 2024; Li et al., 2022) that suggest the physiographic characteristics of static features may be irrelevant; what truly matters is the presence of unique identifying information.

References

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 Resources Research*, *57*(6). 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

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., & Matias, Y. (2023). Caravan - A global community dataset for large-sample hydrology. *Scientific Data*, *10*(1), 61. https://doi.org/10.1038/s41597-023-01975-w

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 Resources Research*, *58*(8). https://doi.org/10.1029/2021WR031794