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 2 for carefully reviewing our manuscript and giving their insightful comments and overall positive feedback. 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).

This paper illustrates "doing hydrology backwards" by developing an LSTM model to predict precipitation based on reanalysis products, given meteorological inputs and the added input of catchment discharge. The authors show that a model given discharge does a better job at predicting precipitation, indicating that discharge encodes significant information about recent precipitation beyond other meteorological forcings. They also find that while the LSTM underestimates precipitation totals relative to the ERA5 training dataset, it better reproduces events that are poorly captured by ERA5.

This is a very interesting paper and appropriate for this journal. It effectively shows that a machine learning approach can be used to improve uncertain precipitation forcings, especially for short time-scale events that are not well represented in reanalysis. With this, I have several comments listed below that could improve and clarify some aspects. As a note, I see some of these may overlap with the first reviewer who also made good points and authors have already responded.

We thank the reviewer for constructive, supportive suggestions and for highlighting the work's potential. We have prepared the following points to address the main comments raised by the reviewer.

General comments:

This study poses that an LSTM model that is trained to reproduce ERA5 precip can actually estimate precip better than the ERA5 product itself. This is based on the input of "future" discharge, which encodes observed precipitation events that are not typically well captured by ERA5. In this way, the LSTM could deviate from the ERA5 because (a) ERA5 is not capturing precip as it actually occurred or (b) the LSTM is not performing well. Unless I am mistaken it seems hard to disentangle these, and the observation gage-based "E-OBS" product seems important here and could be better described. For example, Figure 2 shows that the LSTM "with discharge" better replicates ERA5 precipitation than the LSTM "without" – and it is assumed that this better replication is a good thing. Meanwhile later figures illustrate differences in ERA5, LSTM, and E-OBS regarding specific events, but the LSTM "without" discharge is dropped. In general, it seems useful if E-OBS, ERA5, and both LSTM estimates could be compared up front to more clearly establish differences between them, i.e. what is currently done just between the LSTM models and ERA5 in Figure 2. As far as E-OBS, a few more details on that data might be beneficial especially in the events selected for Figure 5 and associated discussion. For example, what is the proximity of a gage to the specific study catchments?

We agree with the reviewer that some additional information is required for a better understanding of the results. Some of the relevant points we will discuss are:

A. "True" precipitation estimates are not known at the catchment scale. We obtain estimates of them (with considerable uncertainty) by either interpolating station data (EOBS) or averaging gridded data from reanalysis/remote sensing products (ERA5 Land). Our aim was to generate a precipitation time series (estimate) that is more "consistent" with the dynamics captured in the discharge record.
B. The comparison between the models *with_discharge* and *without_discharge* had two primary objectives. First, we aimed to determine whether the discharge values contained useful information about the corresponding precipitation. Second, we sought to evaluate if the LSTM model could effectively capture this non-linear relationship. The *without_discharge* served as a benchmark for evaluating the information gained from including discharge data. To avoid redundancy, we chose not to include these runs in the spatial maps. In response to the reviewer's comment, we conducted another comparison (Figure 1) of the model *without_discharge* for one of the out-of-sample catchments and again found that its performance was inferior to the model that incorporated discharge information.
C. More information regarding the EOBS observational product for the out of sample test will be provided in the revised draft. Figure 2 shows the proximity of observational stations (used for deriving EOBS) to the four specific catchments considered in this study.

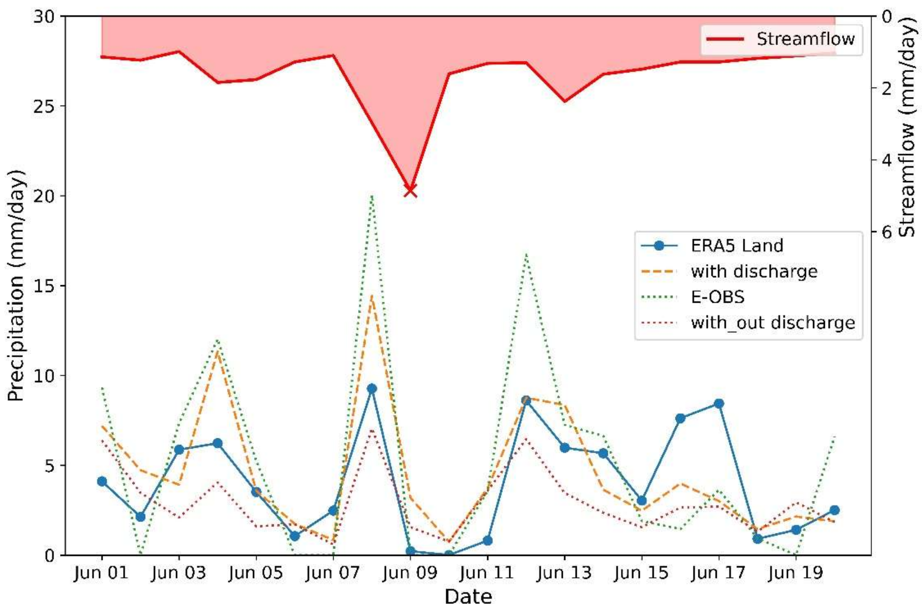

Figure 1 Precipitation estimates for the flood event on June 8, 2016, at the Elsenz Schwarzbach. The red line represents the observed daily streamflow, with a cross marking the day of the flood. The orange curve illustrates the precipitation amount predicted by the with_discharge LSTM model, while the dotted red line represents the without_discharge model. The blue line depicts the original ERA5 Land time series, and the green line shows the estimate from the gauge-based E-OBS product.

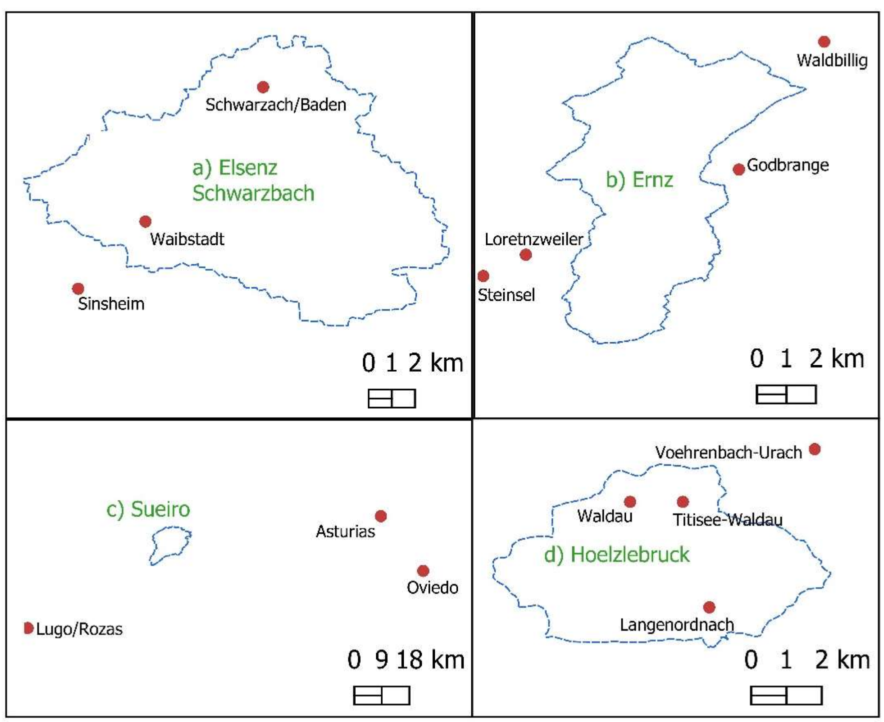

Figure 2 Spatial maps showing the proximity of observational stations (used for deriving the EOBS gridded product) to the four out of sample catchments considered in the present study.

I can imagine that this method might be more effective for smaller catchments, and less effective for very large catchments where the effect of P on Q is more lagged and smoothed. For a very large catchment, estimating a single time-series of P based on Q seems like it would be trying to "average" multiple ERA-5 or gage-based grid cells. Some details on the spatial characteristics of the study catchments might be useful here, especially relative to the scale of gridded precipitation forcing. For example, is there any meaningful trend in model behavior (for any model) with catchment area, or are all study catchments relatively smaller in scale than any precipitation input that would be used?

To address the question of performance in differently sized basins (also asked by Reviewer 1), we will supplement the existing analysis by an evaluation of the methodology over larger, previously unseen test catchments from the recently published Caravan CAMELS-DE (Loritz et al., 2024) dataset. The results (Figure 3: camelsde_DEA11130) again indicate a reduction in relative errors in peak discharge and flood volume for such large catchments (>3000 km^2), using the inversely generated precipitation.

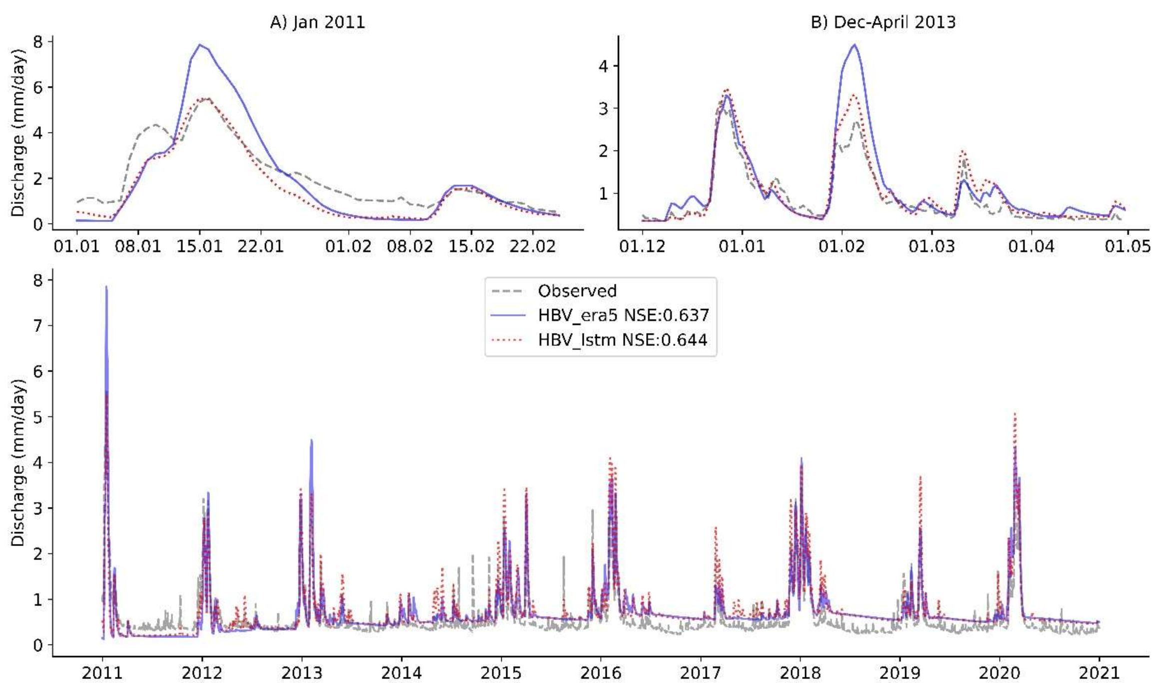

Figure 3 Observed and simulated runoff (using the HBV model) at the Lippe catchment (camelsde_DEA11130). The blue line denotes the streamflow simulated using the ERA5 Land precipitation product, while the red curve depicts the simulations using the inversely-estimate precipitation obtained using the regional with_discharge LSTM model. Moreover, two rainfall-runoff events are highlighted and displayed.

Although the out-of-sample catchments we selected are relatively small, our approach using the LSTM model—trained on much larger catchments—showed skill at adjusting the (under) estimated precipitation values for these events at such smaller scales. This is noteworthy given that ERA5 Land precipitation typically performs poorly at this scale; only about 9% of the catchments in our training dataset had areas smaller than 100 km^2. This finding demonstrates the ability of our inversion methodology and spatial transfer learning (He et al., 2011) to effectively leverage knowledge from data-rich, well-represented regions to make predictions in data-scarce areas.

With all the different models and products, a table or two might be useful – for example listing the properties of precipitation datasets, models and references, flow data. This could be linked to Figure 1 which gives the flow of the study.

We thank the reviewer for his suggestion. A new Table 1 will be added to the Data section detailing the spatial and temporal resolution of all the data sources used in the study.

Figures: figure captions could all be expanded or improved. For example, Figure 1needs a more descriptive caption that addresses the content of each panel and the connections. As it is, it does not really describe the flow of the study and could be a lot more useful to the reader. Otherwise, figures with (a), (b) (c) should be more clearly labeled as such in the captions, and figures like Figure 7 with no panels should not have any references to panels (a), (b), (c), etc. It is also a bit hard to compare the spatial images in Figure 3 because of the grey shading in the top 2 figures but not in the E-OBS panels (so it would be nice if the same masking could be applied to all of these maps). Finally in figures and text it should be made specific that when "LSTM" is mentioned in text or a caption that it is specified as one model or the other ("with" or "without" discharge).

We will implement the necessary changes in the revised draft. The captions for Figure 1 will be expanded to include detailed methodological steps that link to different sections of the paper. We will also correct the captions for Figure 7. For Figure 3, the grey area indicates the regions where catchment data is not included in this study. The EOBS was presented in its full spatial extent, as it is an interpolated product. Additionally, we will clarify the use of the *with_discharge* and *without_discharge* models throughout the manuscript.

References

He, Y., Bárdossy, A., Zehe, E., 2011. A review of regionalisation for continuous streamflow simulation. Hydrol. Earth Syst. Sci. 15, 3539–3553. https://doi.org/10.5194/hess-15-3539-2011

Loritz, R., Dolich, A., Acuña Espinoza, E., Ebeling, P., Guse, B., Götte, J., Hassler, S., Hauffe, C., Heidbüchel, I., Kiesel, J., Mälicke, M., Müller-Thomy, H., Stölzle, M., Tarasova, L., 2024. CAMELS-DE: hydro-meteorological time series and attributes for 1555 catchments in Germany. Earth Syst. Dyn. Discuss. 1–30.