Review hess-2024-375

General comments:

The authors present a method to improve the estimation of catchment-average effective precipitation from the ERA5 product by utilizing the information contained in stream flow data and a regional LSTM model.
To validate this interesting approach the authors model the runoff using this catchment-average effective precipitation as forcing and compare it to the runoff in the CAMELS data set. Averaged over all catchments contained in the CAMELS data set this approach improves the modelled runoff compared to using only ERA5 precipitation estimates as forcing.
The paper is well written and has a reasonable length. However, I think the authors could address the topic of "scale" more in depth. This starts at describing the used data sets in more detail, especially by mentioning their spatial and temporal resolution. Furthermore, the selection of the out-of-sample data sets as "proof-of-concept" is restricted to very small-scale basins. It is especially at this scale that daily ERA5 precipitation will most likely not perform well as forcing for a hydrological model due to its coarse spatial resolution.
The authors state that the LSTM model is estimating catchment-average precipitation amounts. This implies that the introduced approach might not perform equally well for differently-sized basins, when the runoff dynamics shift from surface-runoff to baseflow dominated basins. In my opinion the authors should elaborate more on this topic.

The study is interesting and introduces a promising approach to improve gridded precipitation products which is why I recommend its publication in NHESS after addressing the following comments.

Specific comments:

- The question about the feasibility of the introduced approach (as stated in the abstract) could be answered more clearly. While the authors show that their approach indeed improves the modelled discharge, they could elaborate more on how this method should be actually applied. E.g. a general improvement of products like ERA5: Does the method have potential to adjust the precipitation estimates of ERA5, e.g. as a pre-processing step? What could be general use-cases of the method?

- The authors state that the method could be useful to "reproduce small-scale high-impact events" (l.15-16) which also reflects in their choice of out-of-sample catchments. However, in my opinion it should be stated more clearly (e.g. in the

"Limitations") that using daily precipitation sums of products like ERA5 and EOBS should really be the last resort when trying to reproduce small-scale hydrological events (with or without LSTM correction). The selected flood events in four out-of-sample catchments were dominated by sub-daily rainfall and a rapid, probably also sub-daily, flood response that can just not be captured with daily sums and the spatial and temporal resolution of ERA5 and EOBS.

- In Figure 3 you are comparing ERA5 and EOBS precipitation estimates to the effective precipitation generated by the LSTM. Therefore, the lower amounts from the LSTM seem obvious. How useful is this comparison between two different types of precipitation?

- **Line 318**: *"this consistent underestimation also seems physically plausible"*. Do I understand correctly that the LSTM-model systematically underestimates the precipitation in order to generate less surface runoff, to compensate for lacking baseflow dynamics? The systematic underestimation of precipitation from the LSTM model (as seen in Fig. 3) seems problematic to me because then the inversely derived precipitation would be <u>only useful to model the discharge in surface-runoff dominated catchments.</u> The four out-of-sample catchments are all quite small: How would the LSTM-derived precipitation perform when forcing a hydrological model in a larger out-of-sample catchment that is dominated by base-flow dynamics? Please clarify this, because in my view this is a key point. You could also consider splitting the results of the two regional LSTM models into groups (by basin size) to investigate whether the underestimation of precipitation is more problematic for larger catchments when modelling the discharge.

- If you use "catchment-average effective precipitation" as forcing for HBV and CatFlow, what happens with the soil water dynamics, and interception? Do they get subtracted again from the effective precipitation input? Maybe the overall wording is misleading and the LSTM is not really returning effective precipitation?

- The authors mention that their approach could be useful in "data-scarce regions world wide". However, the uneven distribution or lack of stream gauges could also introduce a bias to the LSTM model.

- A very brief explanation of the used data sets in the data section would be helpful for the reader, even though most of these data sets are commonly used. The temporal and spatial resolution of these data sets has to be mentioned, as well as the way these data sets are derived (e.g. based on station data). The authors do this partially, e.g., in line 128, for the E-OBS dataset but I would recommend doing it in a more structured way, for instance in a table.

- CARAVAN
- ERA5
- E-OBS
- CAMELS

You could also consider mentioning MERRA and GLDAS already here. Additionally, you could consider splitting the “Data and Methods” section into “Data” and “Methods” to provide more clarity.

- Please describe your measures of goodness briefly and why chose to use them: mean wet days, spearman lag, 95th percentile. For mean wet days you should also mention the unit mm, otherwise this gets confusing. This also applies to Figure 3 where the unit “mm” is missing

- **Line 121**: Please write out the abbreviations for the five catchment static attributes.

- **Line 96**: “the model was provided with a 7-day lead time series for discharge” Don’t you also have to provide the 7-day lead time series of the other forcings? Can you explain why you decided for a lead time of 7 days? Is the lead time not catchment specific/scale dependent, or does 7 days cover everything?

Technical corrections:

Generally, the plots are not displayed nicely in the PDF. I am not sure if the reason is the processing by HESS or if you should increase the resolution.

Line 40: The reference “*Clerc-schwarzenbach et al.,2024*” should be written with a capital “S”: Clerc-Schwarzenbach

Line 45: The reference Manoj J et al. is missing a “.” after “J”.

Line 46: “*precipitation forcings data*”. This sentence seems a bit awkward. How about. “precipitation forcing data”?

Line 50: “*it is usually the rainstorm events occurring in poorly observed areas that lead to high impacts*”: This sentence could be misunderstood in a way that high-impact-rainstorms preferably happen in areas <u>because</u> they are poorly observed, while in fact the observation network is too sparse and the majority of high-impact-rainstorms is simply not observed.

Line 56-59: "*While the classical "forward rainfall-runoff generation problem" has received considerable attention over various decades (Montanari et al., 2013; Sivapalan et al., 2003), a smaller subset of studies (Brocca et al., 2013; Kirchner, 2009; Kretzschmar et al., 2014; Krier et al., 2012; Teuling et al., 2010) has investigated the feasibility of tackling the inverse problem more efficiently.*"

What do you mean with "more efficiently"? More efficient than what?

Line 59: Reference in the wrong format: Kirchner (Kirchner, 2009) should be Kirchner (2009).

Line 81: You could already make a reference to Table 2, to reveal more information about the out-of-sample catchments.

Line 90: References in wrong format. Parenthesis should just be around the year. E.g. Kratzert et al. (2018).

Line 93: Reference in wrong format. Should be "Loritz et al., 2024".

Line 104 and 105: Why are you using two different references for the HBV model?

Line 156: You already spelled out the abbreviation of the HBV model previously and do not need to do it again here.

Line 184: "underlying causes of precipitation". What do you mean by this? Maybe "causes" is the wrong word?

Line 190: I think you confuse "without_discharge" and "with_discharge" here. It looks like "with_discharge" has a steeper curve.

Figure 3: Please add the units, otherwise it is unclear whether it is [days] or [mm] for "mean (wet days)". You mention "mean wet days" before but explain the unit first in line 204.

Line 241: "*precipitation values revealed closer estimates to those reported*" -> I think it should be: precipitation values revealed estimates closer to those reported.

Table 2: The first column of the table is not correctly formatted.

Line 326-334: This paragraph seems unnecessarily long. If your main message is that products like ERA5 have a too coarse resolution to capture small-scale precipitation events you can shorten this. Instead, you could elaborate here on the potential of your method, improving such products.

Line 374: "*adds its own biases to the modelling exercise*" Can you explain this?

Figure S2 and S3: Minor detail: Maybe it would look better, if you put the north-arrow in some corner, so it is less prominent? In Figure 3 there is no legend, so you could at least mention the red triangles in the caption.

Figure S5: How did you derive the gridded precipitation? By interpolating the rain gauges? It is not recommended using the rainbow color map because it is not perceptually uniform and therefore not easy to distinguish for people with some kind of color blindness. See also the HESS guidelines, section "color schemes": https://www.hydrology-and-earth-system-sciences.net/submission.html#figurestables