

Remko Uijlenhoet

General remarks

This manuscript presents an analysis of (liquid) mountain precipitation (i.e. rainfall) for more than 400 river catchments (smaller than 2,500 km²) across different continents (Europe, the Americas, Asia), focusing on a large number of extreme, flood-producing rainfall events (more than 10,000 in total). The manuscript emphasises the lack of ground-based precipitation observations in mountainous regions, the bias (underestimation) in precipitation estimates derived from reanalyses (notably ERA5L) and their hydrological implications (through streamflow simulations using a distributed hydrological model, DCHM). The paper claims this bias is not only present in the mean but also in terms of the catchment-scale variability. The authors apply a method based on the catchment water balance and rainfall movement over the catchment (called Inverse Rainfall Correction, IRC) to compensate for the rainfall bias and artificially increase its spatial resolution. They also train an AI model (based on random forest regression) on catchment and rainfall attributes for Alpine events and apply it to events from the other continents. The authors report significant reductions in rainfall bias (with respect to rain gauge observations) and improvements in streamflow simulations (with respect to GRDC observations). Overall, the contribution of the paper is relevant for catchment hydro(meteoro)-logy and as such fits the scope of HESS. However, the organization of the paper can be improved, certain technical aspects require more detail and some of the claims can be toned down. Therefore, I recommend major revisions. Let me elaborate on this below.

Specific remarks

- L.11-13, Abstract, “This work provides a novel path to systematically enhance precipitation estimates globally using a general hybrid physics–AI framework, even for the top 1 % of extreme events”: Please provide a factual summary of the employed methods and main conclusions in the paper rather than general statements like this. Also, wouldn’t simply increasing the number of ground-based measurements in mountainous regions be a better recommendation to the hydrological community than “enhancing precipitation estimates globally using a general hybrid physics–AI framework”?
- L.63, L.88: How do you define the beginning, the end and hence the duration of these “extreme precipitation events”? Also, it seems that, in the remainder of your paper, you mainly focus on event total rainfall accumulations and leave the event durations out of your analyses. However, this could be a significant explanatory factor for bias and uncertainty in your further rainfall and streamflow analyses and it would add to the work you present if you would take it into consideration. Was there a specific reason for neglecting this aspect?
- L.85: Can you quote any results from previous studies for the performance of the Duke Coupled Hydrology Model (DCHM) for mountainous catchments smaller than 2,500 km²?
- L.90-91, “extreme rainfall produces extremely fast floods on the scale of a few hours to a few days, which are usually referred to as flash floods”: You call that “extremely fast”?

- L.95, “Calculation of rainfall uncertainty”: What do you mean with “rainfall uncertainty” in this context? Estimation uncertainty? But that is not really treated here.
- L.99-100, “IRC [...] produces improved rainfall estimates that are constrained by water budget closure at the basin scale using Lagrangian particle tracking coupled with a hydrological model”: Besides rainfall and streamflow, how are the other water budget components estimated, notably evaporation? Please provide additional details on how IRC works exactly.
- L.112-114, “Note that for basins in other mountain ranges (i.e., non-Alps basins), due to much lower spatial coverage of rain gauges, particularly in developing countries, rain gauge-only gridded QPE data at high resolutions are not available for comparison”: What does that imply for the quality of the ERA5L reanalyses in these poorly gauged areas?
- L.118, Fig. 2: Event rainfall totals up to 1000 mm! What are the corresponding event durations?
- L.128-130, “it is not practical to conduct a complete experiment of more than 10 000 events worldwide at these spatial scales with current computing infrastructure. The projected data storage would exceed 200 petabytes”: Do these computational constraints still exist in the age of AI?
- L.137-138, “flood simulations using APGD exhibit excessively large peak flows with errors greater than 140 %”: Is this not an indication that the hydrological model is not performing well?
- L.138, “This is because of the large spatial variability of rainfall in complex terrain”: Please explain how spatial variability in rainfall can lead to bias in streamflow simulations.
- L.146, caption Fig. 3, “Precipitation comparison at rain gauge locations”: So ERA5L-IRC overestimates rainfall with respect to gauges, but does a good job in hydrological simulations. How come?
- L.150, “at hourly resolution”: What is the corresponding spatial resolution?
- L.151-152, “MODIS dataset linearly interpolated to hourly resolution”: What was the native resolution of the MODIS data?
- L.157, “The total number of studied basins [...] is 423”: What is the repartitioning over the continents?
- L.165-167, “The focus of this paper is to demonstrate the underestimation of extreme rainfall in remote sensing products and show its impact on downstream tasks”: Is “the underestimation of extreme rainfall in remote sensing products” really “the focus of this paper”? You are using ERA5L reanalysis data rather than RS products directly.
- L.168, “2.5 Cross-regional model validation”: Does this section contain enough information to allow readers to reproduce the results?
- L.178, “where $do(x, y)$ is the distance from a basin pixel (x, y) to the basin outlet normalized by the maximum do in the basin, P ”: do is calculated via the drainage network I presume? Please refer to work by Jim Smith and others who have used this or comparable metrics for describing rainfall variation at catchment scale.
- L.209, “Rainfall uncertainty analysis suggests severe underestimation”: Bias is not uncertainty. Please explain.
- L.212, “thus roughly 10 %”: Fig. 1f seems to suggest 5% underestimation on average. Please check.
- L.226-228, “the spatial variability of extreme rainfall is much higher than what can be detected with current remote sensing and data assimilation techniques constrained by the

underpinning models and observing systems”: How do you know for sure if there are no data to underpin this?

- L.231, Fig. 4, X-axes (left panels) and Y-axes (right panels): replace “Unit” by “standard deviation”.
- L.244, “we estimate that it would require 13.6–54.4 million rain gauges in the Alps alone to capture or represent the heterogeneity of extreme rainfall, compared to its current number of roughly 8500”: How did you obtain this estimate?
- L.260, Fig. 5c: Even for ERA5L-IRC, more than half of the events have a KGE < 0.7 . What are the implications of this observation for the quality of the streamflow simulations?
- L.290, Fig. 6: Only 2 river catchments in the Himalayas and only 8 in SE Asia, out of a total of 423 (l.157). These are very small fractions.
- L.295-296, “Cross-regional validation with ~4500 events strongly supports that the spatiotemporal structure of orographic rainfall errors in ERA5L can be generalized and predicted”: Well, overall more than 80% of flood events seem to exhibit KGE values < 0.7 . In addition, the interest lies not so much in KGE values < 0 , but in the ones > 0.7 . Why extend x-axis to -3?
- L.308-309, “Even better performance would be expected if large enough data sets were available to enable training the AI rainfall error model conditional on specific weather regimes”: OK, but perhaps there are also other attributes that could be used in the training phase (related to geography and climate).
- L.313, “4 Outlook”: Is this an outlook or a discussion?
- L.316, “Using flood simulation KGE thresholds of 0 (useful)”: I would not call all KGE values above 0 “useful”.
- L.321-322, “However, incorporating flood information into rainfall estimation at these small scales (i.e., sub-km for extreme events) has never been done before in atmospheric science and hydrometeorology”: This is a very bold statement. For instance, what about the “doing hydrology backward” approach of Jim Kirchner that was attempted by several investigators?
- L.323, “5 Conclusions”: These are not conclusions; they represent an outlook at best. Please provide a point by point list of the main conclusions of this work.

Editorial remarks

- L.75, Fig. 1: Why present results way before you have introduced your methods and data? Please move figure panels reporting results to the Results section of your paper.
- L.102, “specifically”: Do you mean “namely”?
- L.107, “the total events studied amount to”: Replace with “the total **number of** events studied amounts to”.
- L.108, “and data outputs are approximately 186 TB”: Not relevant I think. Can be removed.
- L.118, Fig. 2: Figure not clear (blue dots cover red dots). Please provide statistics for both ERA5L and ERA5L-IRC, corresponding to both panels.
- L.143, Fig. 3: Which watershed is this?
- L.157, “10 791”: Already mentioned previously (l.107-108).
- L.160, “with its first version [...] published in 1995”: This information can be omitted if you refer to (Barros, 1995).
- L.164, “It is worth noting that this model is not calibrated”: Already mentioned before (l.105).

- L.206, “(Fig. 1e)”: Why refer back to Fig. 1 instead of including a figure with results in this section?
- L.215-219, “For the United States [...] in this manuscript”: These statements do not belong in the Results section of your paper.
- L.234, Fig. 4, “The x and 2x annotations”: No need to indicate these in the figure. The “roughly 2-fold difference in rainfall spatial variability between the original ERA5L and ERA5L-IRC” is already clear enough from left and right panels.
- L.242-246, “Based on our findings [...] in the foreseeable future”: Statements like these do not belong in the Results section. In addition, “(e.g., the 21st century)” can be omitted.
- L.324-325, “This work originates from a very different perspective than most rainfall research, and we believe it provides valuable insights from a fundamental physics point of view (i.e., water mass balance) into rainfall estimation”: A generic statement like this does not fit in the Conclusions. I suggest to shorten it to “This work employs the water balance for rainfall estimation”.
- L.330-334, “This paper sets a milestone in advancing optimal rainfall estimation in complex terrain and establishes a foundation for achieving high-quality rainfall forecasts, which in turn underpin closely related downstream hazard modeling and engineering applications (e.g., flood simulation, flood forecasting, landslide prediction, flood inundation, reservoir design, etc.). The reason is simple: errors in model input inevitably propagate to errors in model output”: Whether or not your paper “sets a milestone” or “establishes a foundation” is for others to judge. The final sentence is a truth known to all hydrologists.