https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

Reconstructing catchment precipitation with deep learning and a large network of streamflow gauges

Min Wang[1] and Haiyang Shi[2]

[1]Institute of Agricultural Resources and Environment, Xinjiang Academy of Agricultural Sciences, Urumqi 830091, Xinjiang, China
[2]Department of Land Surveying and Geospatial Science, The Hong Kong Polytechnic University, Hong Kong, China

Correspondence: Haiyang Shi (haiyang.shi@polyu.edu.hk, haiyang.shi9473@outlook.com)

Abstract. Global precipitation estimates are uncertain, particularly in sparsely gauged forested areas and mountainous terrain. Improving precipitation estimates in such regions is important for studies of the global water cycle. Previous studies in individual catchments have demonstrated the potential of inferring precipitation from streamflow. Using the Multi-Attribute Catchment Hydrometeorological (MACH) dataset, we shared the parameters of an entity-aware long short-term memory network (EA-LSTM) across catchments and used local weather and streamflow records to reconstruct daily areal precipitation for 906 catchments in the conterminous United States. Each reconstruction used observations through seven days after the target date, with no daily precipitation input. The reference precipitation and weather variables came from Daymet. With runoff included, adding relative humidity, vapour pressure deficit and downward shortwave radiation to temperature improved daily Kling–Gupta efficiency (KGE) in 895 catchments and raised its median from 0.626 to 0.771. Median KGE reached 0.834 and 0.845 for three- and five-day totals. The median fraction of reference events with a correctly identified peak date was 73.3%. Median amount bias on reference heavy days was −32.0%, indicating that the magnitude of extreme precipitation remains more difficult to reconstruct. Grouped by the relative size of the two surface-flow components, median exact peak-date success was 73.8% in infiltration-excess catchments and 67.5% in saturation-excess catchments. The median paired gain from adding runoff was 1.9 and 4.5 percentage points, respectively. These results show that learning across catchments can recover multi-day precipitation totals from streamflow and weather records. Future research could improve transfer across catchments and the representation of extreme precipitation amounts.

1 Introduction

Catchment water balances depend on how much precipitation falls over the upstream area. Interpreting a flood hydrograph also requires knowing when that precipitation fell and which day received the most rain. Both are difficult to establish from point measurements in sparsely gauged forests and complex mountain terrain (Henn et al., 2015; Sun et al., 2018). Gridded estimates extend the spatial coverage, but six products for the western United States showed their largest absolute differences in annual precipitation at high elevations (Henn et al., 2018). A stream gauge offers a complementary observation. Water

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

from different parts of the catchment converges at the outlet, where the hydrograph records the integrated response to upstream precipitation. Where discharge is reliably measured and human influences can be assessed, that record offers a way to revisit periods when precipitation observations are missing or uncertain (Kirchner, 2009; Herrnegger et al., 2015).

Studies have used streamflow both to infer precipitation and to correct an existing precipitation estimate. Kirchner (2009) exploited storage–discharge relationships to infer rainfall. Herrnegger et al. (2015) inverted an HBV-type model in two Austrian catchments and found higher correlations with station rainfall for daily than hourly totals. Henn et al. (2015) used Bayesian inference to estimate mountain basin precipitation under hydrological model uncertainty. Liao and Barros (2022) traced discharge residuals backward along simulated runoff pathways to correct existing precipitation estimates. Manoj J et al. (2025) combined subsequent discharge with ERA5-Land precipitation and other inputs in a regional LSTM, using E-OBS precipitation as the target. These studies establish that discharge carries information about precipitation. They also leave open which features of a storm can be recovered when daily precipitation estimates are withheld.

Which features can be recovered depends in part on what happens to rainfall before it reaches the gauge. Evapotranspiration, storage and routing can allow different precipitation sequences to produce similar outlet responses. Catchment state further changes the relationship. Storage deficits may weaken an event signal, whereas wet, connected pathways may respond rapidly. In two monitored catchments, Knapp et al. (2025) found stronger responses under wetter conditions without a corresponding increase in new-water fraction. Rapid response therefore need not mean rapid passage of new rainfall. Snow storage can also separate precipitation from subsequent water release.

Even in the absence of snow storage, a rapid discharge response can arise through different pathways. Rainfall intensity may exceed the soil's infiltration capacity, or rain may fall on already saturated ground connected to the channel. Both mechanisms can operate in the same storm (Hong et al., 2026). Subsurface flow is also common in forested catchments (McMillan et al., 2025), and climate and subsurface storage help explain spatial variation in event response (Tarasova et al., 2018). Thus, even where discharge typically peaks a day after rainfall, it remains an open question how much the outlet record helps identify the wettest day of a storm across different runoff-generation settings.

Answering that question requires comparing catchments under a common reconstruction method. Large-sample hydrological datasets link long weather and discharge records to catchment attributes (Addor et al., 2017). An entity-aware long short-term memory network (EA-LSTM) can learn from those records across catchments while allowing its response to vary with local attributes (Kratzert et al., 2019). Weather records also carry information about precipitation. Humidity and shortwave radiation can help identify wet conditions, and Daymet uses precipitation information when generating its vapour pressure and radiation fields (Thornton et al., 2021). Estimating what discharge contributes therefore requires comparing reconstructions with and without discharge at the same sites and dates, keeping the weather inputs fixed.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

We use records from the Multi-Attribute Catchment Hydrometeorological (MACH) dataset in the conterminous United States to ask which features of past precipitation can be recovered in a held-out catchment from local weather and streamflow. The EA-LSTM uses observations through seven days after each target date, with no daily precipitation input. We first examine how adding humidity and shortwave radiation to temperature affects reconstruction and how much discharge adds under each set of weather inputs. Across catchments with different snow influence and elevation, we assess precipitation totals at daily and multi-day scales, the date of the event peak and amounts on heavy precipitation days. Finally, we compare peak-date accuracy and the gain from discharge across surface-flow groups to determine where streamflow provides the most information about precipitation timing.

2 Data and methods

2.1 Catchments and data

MACH version 4 provides 1980–2023 records and attributes for 1,014 catchments in the conterminous United States (Sink and Brikowski, 2026). Catchment-mean Daymet V4 precipitation (mm d^{-1}), including rainfall and the water-equivalent amount of snowfall, was the reconstruction reference. USGS observed outlet streamflow was converted using catchment area to catchment-equivalent runoff depth (Q, mm d^{-1}), hereafter termed runoff. Temperature, vapour pressure and shortwave radiation came from Daymet (Thornton et al., 2021). Training eligibility required complete attributes and 1981–2000 coverage of ≥95% for runoff and ≥99% for temperature, precipitation and extended meteorological variables. The eligible pool comprised 855 catchments. Evaluation separately retained 906 catchments with common valid reference and prediction support on ≥90% of 2011–2022 days.

Catchments were classified using 1981–2000 snow water equivalent (SWE) and daily mean temperature. Rainfall-dominated catchments had ≤1% days with SWE >1 mm and ≤1% precipitation amount at mean temperature ≤0 °C. Among the remainder, ≤20% snow days and ≤10% cold precipitation defined low-to-moderate snow influence; others formed the greater snow-influence group. Mean elevation ≥1,500 m defined an overlapping subset.

2.2 Inputs and meteorological information

Four model configurations crossed basic meteorology (minimum and maximum temperature) or extended meteorology (adding relative humidity, vapour pressure deficit and shortwave radiation) with runoff present or absent (Table 1). All retained calendar terms and ten attributes: area, slope, elevation, silt, clay and sand fractions, forest fraction, mean annual precipitation, mean temperature and aridity. Climate attributes represented 1981–2000; forest fraction averaged 1985–2000. Aridity was potential evapotranspiration divided by precipitation, using GLEAM (Miralles et al., 2025), with one missing value supplied using potential evapotranspiration from TerraClimate (Abatzoglou et al., 2018).

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

Approximate relative humidity (RH) and vapour pressure deficit (VPD) were derived from catchment-mean daily variables (Allen et al., 1998), averaging saturation vapour pressures at minimum and maximum temperature. Shortwave radiation (SRAD) retained its daylight-mean definition in W m^{-2}. Daymet SRAD includes a precipitation-day adjustment, and its dry-climate vapour-pressure correction uses precipitation information (Thornton et al., 2021). These variables may therefore not be fully independent of precipitation data and are treated here as extended meteorological inputs.

Table 1. Input configurations for the fixed seven-day comparison.

Configuration	Meteorological variables	Observed runoff
Basic + Q	TMIN, TMAX	Included
Basic without Q	TMIN, TMAX	Excluded
Extended + Q	TMIN, TMAX, RH, VPD, SRAD	Included
Extended without Q	TMIN, TMAX, RH, VPD, SRAD	Excluded

2.3 EA-LSTM reconstruction and input comparisons

The EA-LSTM (Fig. 1) used static attributes to control its input gate (Kratzert et al., 2019). A chronological branch initialized its state with 365 antecedent days and then processed target dates in order. A separate branch processed seven subsequent days independently for each target, from day +7 to +1. Their combined representations produced nonnegative daily precipitation. The branches contained 64 and 32 units, with a 16-dimensional attribute representation. All configurations shared this architecture and identical complete input contexts, including runoff availability.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

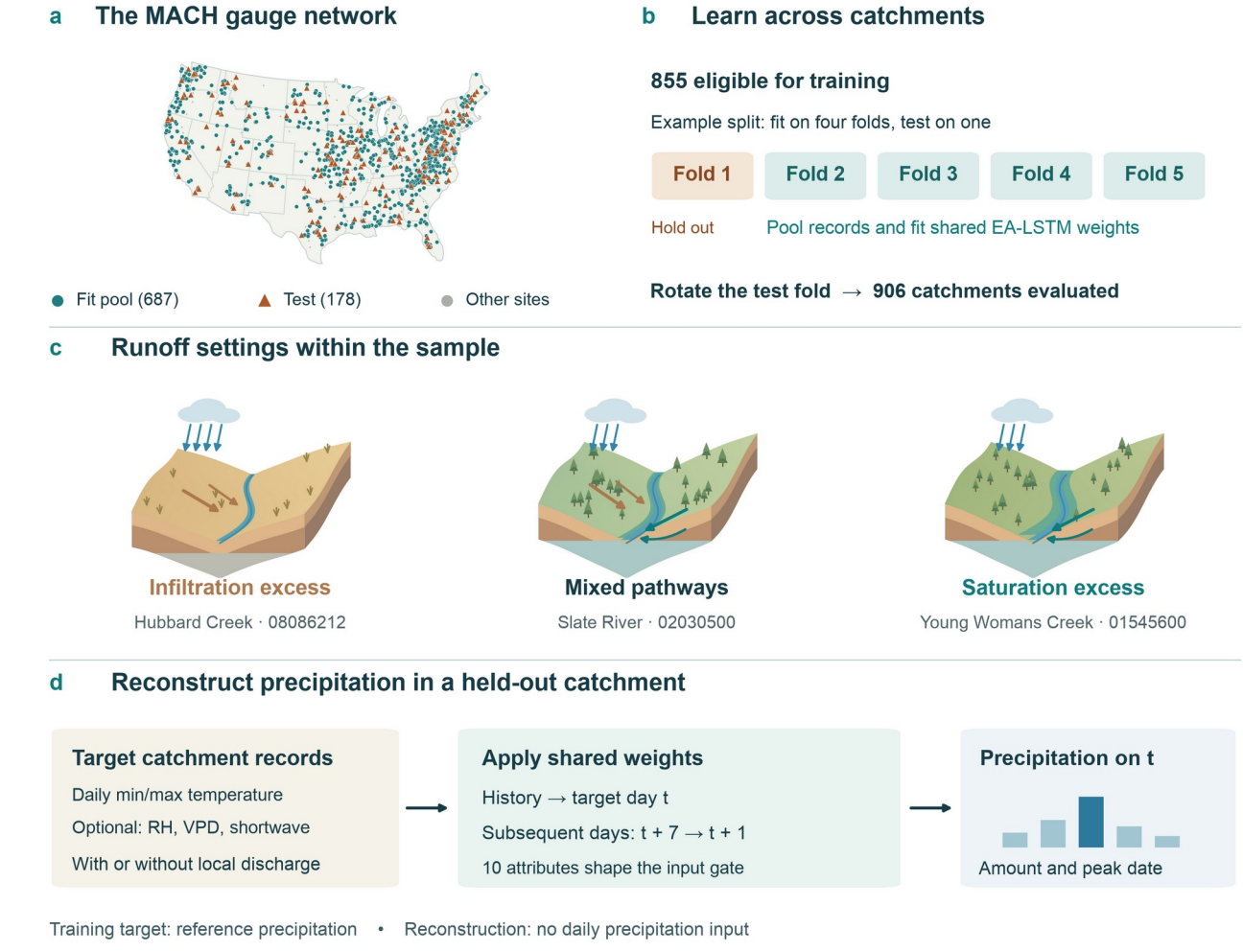


Figure 1. Learning across MACH catchments and reconstructing precipitation at a held-out gauge. (a, b) In the illustrated split, 687 catchments in folds 2–5 qualified for fitting and 178 in fold 1 for evaluation. Grey points are other gauges in this split. Across the five folds, 855 catchments qualified for training and 906 were evaluated. Base map boundaries are from Natural Earth (public domain). (c) Conceptual sections for three gauges illustrate infiltration-excess flow (brown), flow from saturated areas and lateral subsurface flow (green). (d) At a held-out gauge, the model combines shared weights with local attributes and observations to reconstruct target-day precipitation.

2.4 Model training, validation and evaluation

We used five-fold cross-validation. Fitting used 1981–1999 and model selection used 2001–2009 in the four training folds; evaluation used 2011–2022 in the withheld fold. Runoff was log-transformed; predictor standardization used training-fold statistics. Squared errors were normalized by training-catchment precipitation variance, with a minimum standard deviation of 0.1 mm d^{-1}. Each configuration was fitted three times with AdamW (learning rate 0.001; ≤100 epochs), selecting by median validation Kling–Gupta efficiency (KGE; Gupta et al., 2009).

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

Historical climatology assigned each date its 1981–2000 calendar-month mean daily precipitation, averaging complete monthly rates equally across years and requiring at least 15 years per month. Wet days were ≥1 mm d^{-1} and heavy days ≥20 mm d^{-1}. F1 assessed same-date occurrence as the harmonic mean of precision (the fraction of predicted positive days correctly identified) and recall (the fraction of reference positive days detected). Heavy-day amount bias used all reference heavy days. The exact peak-date success rate was the fraction of all complete reference events that were matched and had an identical reconstructed peak date; the peak-date success rate within ±1 day allowed a one-day difference. Metrics were calculated per fit and averaged within catchments only when all three values were defined, then summarized by catchment median and interquartile range. Events comprised consecutive wet days; events touching missing data or a period boundary were excluded. Each reference event could match at most one predicted event by temporal overlap. Peak-date scores required ten complete reference events and heavy-day metrics required ten heavy reference days.

2.5 Runoff pathways and observed response times

The MACH attributes TOT_IEOF and TOT_SATOF give TOPMODEL-based estimates of infiltration-excess and saturation-excess overland flow as percentages of total streamflow. We used the relative sizes of these two surface components to group catchments for evaluation. A positive estimate at least twice the other defined the infiltration-excess (IE) or saturation-excess (SE) group. Positive estimates within a factor of two defined the mixed group.

We estimated the rainfall-to-runoff peak lag from 1981–2000 reference rainfall and observed discharge. Isolated events comprised one to three consecutive wet days totalling at least 10 mm. During the three days before and seven days after each event, precipitation was less than 1 mm each day and at most 1 mm in total in each interval. From seven days before onset to seven days after the event, mean daily temperature exceeded 5 °C and snow water equivalent was at most 1 mm. We used the median flow in the three days before onset as the baseline and located the flow peak between onset and seven days after the event. When the peak exceeded baseline, lag was the interval from the first rainfall maximum to the first flow maximum. We summarised events within catchments with at least five isolated events, then weighted catchments equally.

Group comparisons for extended meteorology with runoff used catchment medians, with 95% intervals from 2000 resamples of connected catchment groups. Adjusted IE–SE mean differences came from a linear regression controlling for two-digit hydrological region, log area, log aridity, elevation, forest cover and cold-precipitation fraction, with standard errors clustered by connected group. We distinguished missed events from matched events assigned the wrong peak day, using all complete reference events as the denominator for both. Gains from runoff were paired within each catchment before summarising by group. For IE and SE separately, we calculated the across-catchment Spearman correlation between historical flow rise per unit of rainfall and the later gain in peak-date success.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

3 Results

3.1 Meteorological information improves reconstruction across most catchments

The common evaluation comprised 906 catchments: 102 rainfall-dominated, 298 with low-to-moderate snow influence and 506 with greater snow influence; 113 also belonged to the overlapping high-elevation subset. With runoff included, median daily KGE increased from 0.626 under basic meteorology to 0.771 under extended meteorology. Median Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) increased from 0.585 to 0.747, while root mean square error (RMSE) decreased from 4.619 to 3.557 mm d^{-1} (Table 2). Adding the derived humidity and radiation variables improved daily precipitation reconstruction across most catchments (Figs. 2 and 3). The within-catchment KGE gain had a median of 0.134 and an interquartile range of 0.087–0.217; 895 of 906 catchments improved. Thus, the overall improvement reflected widespread gains rather than being dominated by a few catchments with large improvements.

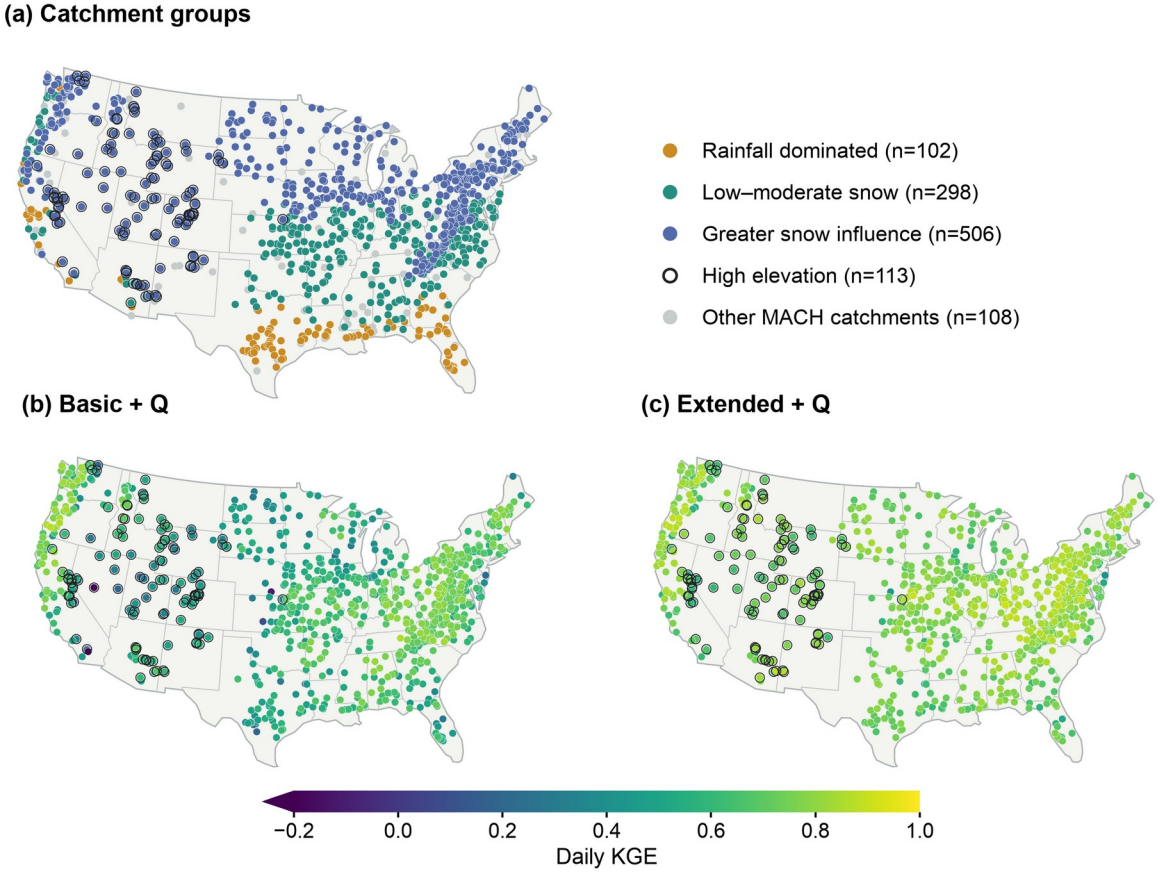
Daily KGE medians increased from 0.544 to 0.723 in rainfall-dominated catchments, from 0.661 to 0.779 in the low-to-moderate group, and from 0.621 to 0.774 in the greater-snow group. The greater-snow group consequently approached the low-to-moderate group in daily KGE. High-elevation catchments showed a larger meteorological benefit: their median daily KGE increased from 0.497 to 0.730, with a paired median gain of 0.250 and improvement in 112 of 113 catchments.

The models without runoff suggested that the extended meteorological inputs already contained some precipitation-related information: median daily KGE reached 0.640 even without runoff observations. Adding runoff further improved overall reconstruction. Its paired median KGE contribution was 0.299 with basic meteorology and 0.134 with extended meteorology; the latter was positive in 816 of 906 catchments. At high elevation, however, the additional runoff contribution under extended meteorology was only 0.022, positive in 74 of 113 catchments. A relatively high reconstruction score in this subset therefore did not imply a comparably strong constraint from streamflow. Meteorological variables and runoff observations contributed differently across the study area.

Table 2. Overall reconstruction performance.

Configuration	Daily KGE	Daily NSE	Daily RMSE	Wet-day F1
Basic without Q	0.309	0.229	6.433	0.641
Basic + Q	0.626	0.585	4.619	0.671
Extended without Q	0.640	0.548	4.944	0.903
Extended + Q	0.771	0.747	3.557	0.894

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



165

Figure 2. Study catchments and spatial reconstruction performance. (a) Evaluation groups and other MACH gauges; rings identify the overlapping 113-catchment high-elevation subset (≥1500 m). (b,c) Daily KGE with basic and extended meteorology, both including runoff.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

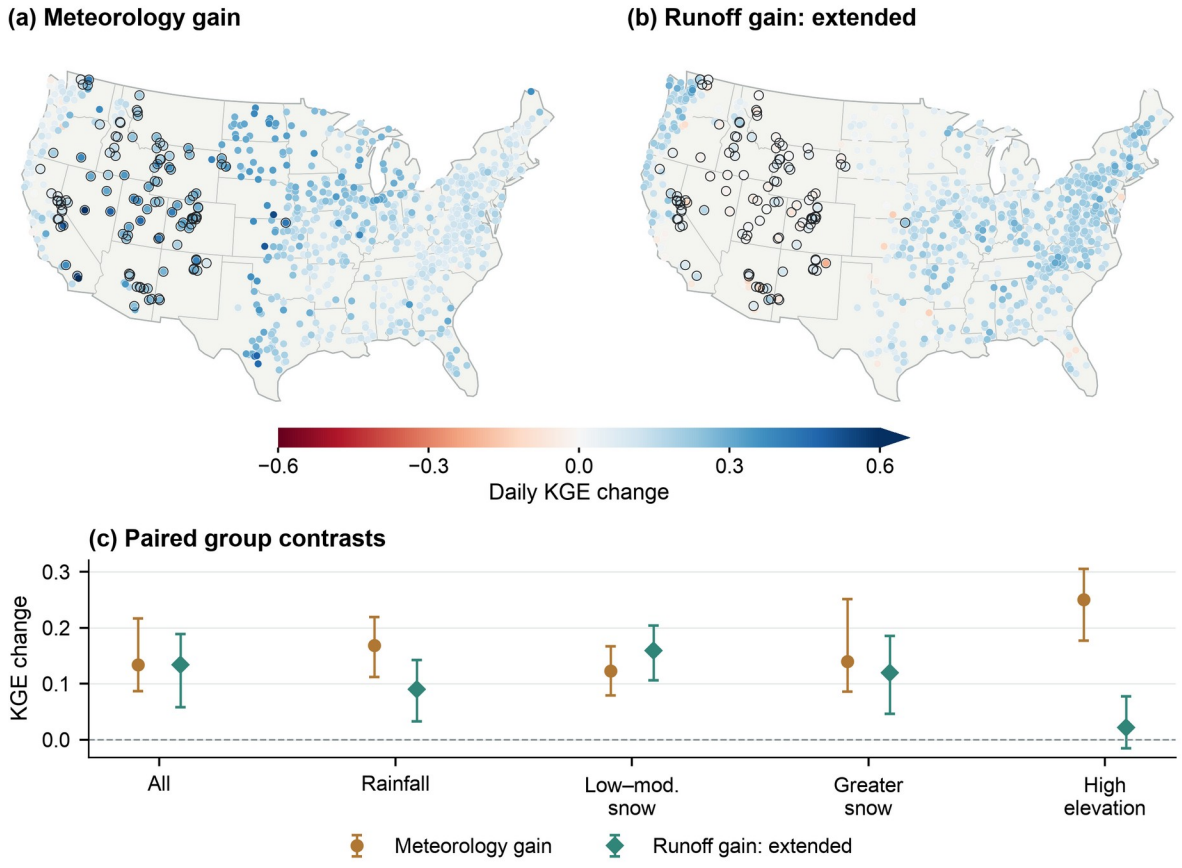


Figure 3. Additional predictive value of meteorological information and runoff. (a) Daily KGE change from extended minus basic meteorology with runoff retained. (b) Daily KGE change from adding runoff under extended meteorology. (c) Medians and interquartile ranges of these within-catchment paired changes across reporting groups. Rings identify high-elevation gauges.

3.2 Differences in accuracy across temporal scales

We also evaluated agreement between reconstructed and reference precipitation across temporal scales by accumulating the daily results into 3-, 5-, 7- and 10-day totals. The largest change occurred between daily and three-day estimates (Fig. 4). For extended meteorology with runoff, median KGE at 1, 3, 5, 7 and 10 days was 0.771, 0.834, 0.845, 0.849 and 0.854. The median paired three-day gain over daily scores calculated on the same retained dates was 0.057, positive in 883 catchments. Extending accumulation beyond five days produced comparatively small additional gains.

For extended meteorology with runoff, median mean square error (MSE) improvement relative to historical climatology was positive at every scale: 74.3% for daily estimates, 81.7% at three days, 82.7% at five days and 83.1% at ten days. At ten days, the extended model retained a paired median KGE advantage of 0.049 over the basic model. Accumulation can partly cancel opposing daily errors. Longer accumulation was less consistently beneficial at high elevation. Extended-model

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

median KGE rose to 0.781 at seven days and then declined to 0.775 at ten days; MSE improvement was 73.4% at five days and 72.1% at ten days. These results indicate that aggregating daily estimates over three to seven days may improve accuracy to some extent, with little additional gain at longer temporal scales.

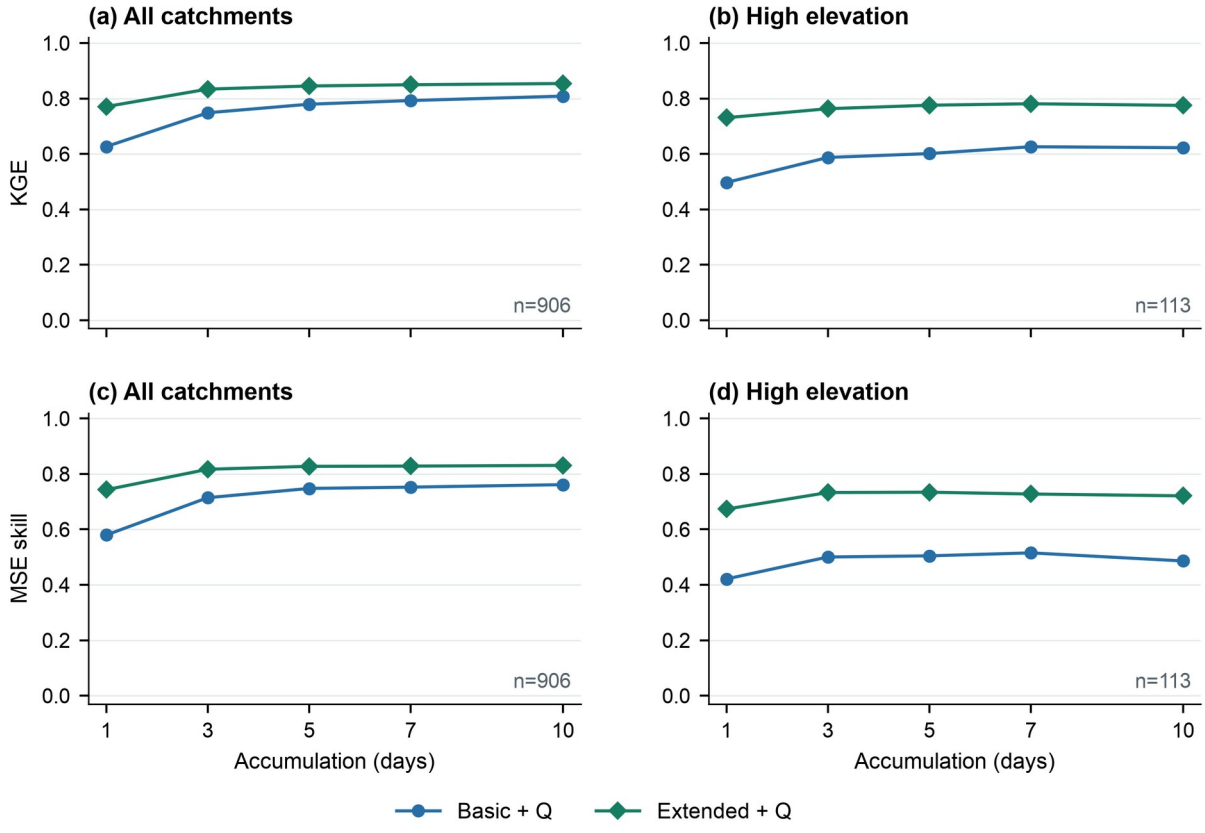


Figure 4. Precipitation reconstruction across accumulation scales. Catchment-median KGE (a,b) and median MSE improvement relative to 1981–2000 historical climatology (c,d) for the models with runoff under basic and extended meteorology.

3.3 Precipitation event identification and heavy precipitation

Compared with basic meteorological inputs, extended inputs improved wet-day identification and event peak-date success rates (Fig. 5). Median wet-day F1 increased from 0.671 to 0.894, median precision increased from 0.550 to 0.922, and median recall changed from 0.884 to 0.872, indicating that the improvement in wet-day identification mainly reflected fewer false detections. The median exact peak-date success rate increased from 34.4% to 73.3%, and the median peak-date success rate within ±1 day increased from 53.2% to 84.6%. The paired gain in exact peak-date success rate had a median of 37.4 percentage points and was positive in every catchment evaluated for this metric. These results indicate that adding the

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

extended meteorological variables to basic meteorology and runoff can locate precipitation dates more accurately, possibly because precipitation days typically have high RH and low VPD.

However, improved event peak-date reconstruction did not remove the underestimation of heavy-day amounts. With runoff retained, among 903 catchments meeting the heavy-day sample requirement, median heavy-day amount bias on reference days with precipitation ≥20 mm d^{-1} was −39.1% under basic meteorology and −32.0% under extended meteorology, indicating a relatively modest improvement. Across 111 eligible catchments in the high-elevation subset, median heavy-day amount bias for Extended + Q remained −51.8%; within this subset, the paired heavy-day F1 gain from extended over basic meteorology had a median of zero. Good daily agreement and high event peak-date success rates could therefore coexist with substantial underestimation of heavy-day amounts.

Runoff still provided additional predictive value for heavy-day identification under extended meteorology. Median wet-day F1 was slightly higher without runoff than with it (0.903 versus 0.894), but adding runoff increased median heavy-day F1 from 0.203 to 0.583, with a paired median gain of 0.295. Thus, adding runoff did not improve every indicator, but it markedly improved heavy-day identification.

The representative process sequences illustrate why timing and amount require separate evaluation (Fig. 6). In catchment 08086212, the 23 April 2019 precipitation peak was followed by a local runoff maximum on 24 April; both models nevertheless underestimated that 60.6 mm d^{-1} peak. In high-elevation catchment 09107000, the extended model reconstructed 31.4 mm over 4–6 October 2011, close to the 32.9 mm reference and above the basic model's 22.4 mm, yet both placed the maximum within that three-day episode a day early. This illustrates the remaining challenge of locating the peak date accurately during precipitation episodes lasting several days.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

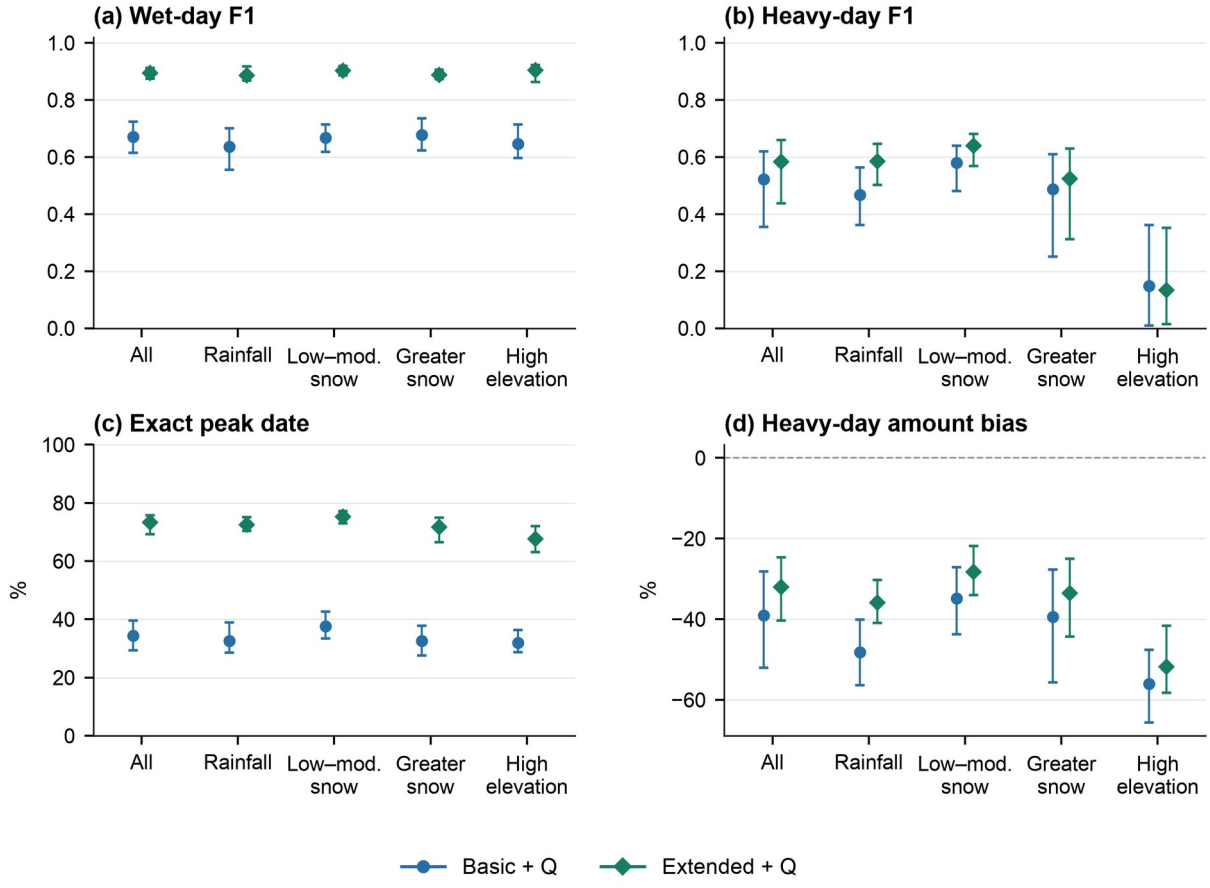


Figure 5. Wet-day, heavy-day and event peak-date reconstruction. Basic + Q and Extended + Q are compared. Metrics are first averaged across three fits within each catchment; points and bars show catchment medians and interquartile ranges within each reporting group. (a,b) F1 for wet and heavy days (≥1 and ≥20 mm d^{-1}, respectively). (c) Exact peak-date success rate, with all complete reference events in the denominator, including reference events left unmatched by one-to-one overlap matching. (d) Heavy-day amount bias on reference heavy days; negative values indicate underestimation.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

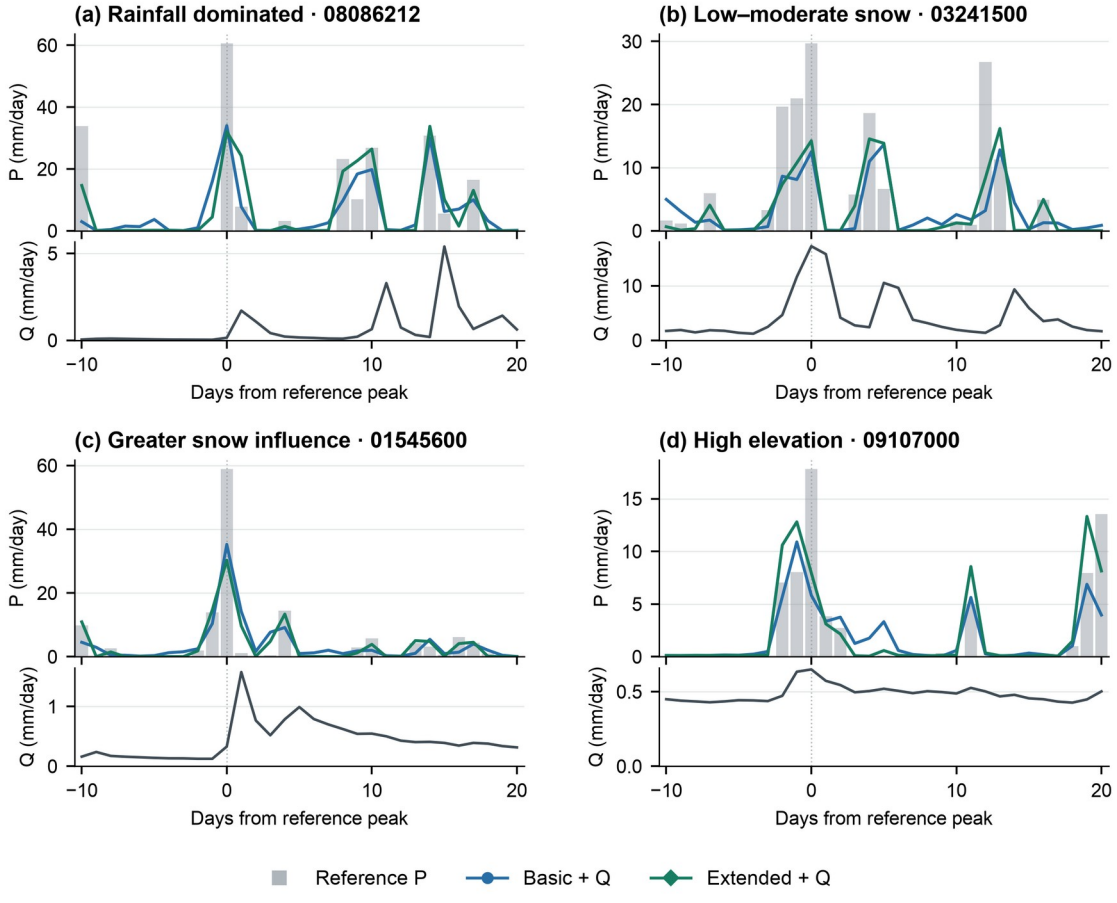


Figure 6. Representative precipitation and runoff sequences. The upper part of each panel shows reference precipitation bars and Basic + Q and Extended + Q reconstructions; the lower part shows observed runoff. A catchment with Basic + Q daily KGE near its reporting-group median is selected for each group, followed by a complete year with annual precipitation near that catchment's median. The three-day period with the largest reference precipitation total in that year is shown with its surrounding sequence, from ten days before to twenty days after the wettest reference day within that period.

3.4 Precipitation peak dates and runoff response across surface-flow groups

The mixed and infiltration-excess (IE) groups had the highest median peak-date success, whereas discharge added more to peak dating in the saturation-excess (SE) group than in IE (Fig. 7b, d). The comparison covers 359 IE, 347 mixed and 188 SE catchments, with 12 of the 906 gauges unclassified (Fig. 7a). IE catchments were common in the central United States. Compared with IE, SE catchments tended to be wetter and more forested (median aridity index 0.88 versus 1.21 and forest cover 69% versus 15%).

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

The main difficulty in SE was locating the wettest day within an event that the model had already detected. With extended meteorology and discharge, median exact peak-date success was 67.5% in SE, 73.8% in IE and 74.4% in mixed catchments (Fig. 7d). Comparing catchment means allowed us to separate two sources of the 6.6 percentage point IE–SE difference. Errors in dating the peak of a matched event accounted for 5.6 points, and missed events accounted for 1.0 point. The mean difference narrowed to 2.4 points (95% interval 1.1–3.8) after adjustment for region and catchment attributes.

With extended meteorology and no discharge, median exact peak-date success was 64.9% in SE and 71.6% in IE. Adding discharge yielded a median within-catchment gain of 4.5 percentage points in SE and 1.9 in IE (Fig. 7b). Thus the SE group gained more information about timing from the outlet gauge even though its final peak-date accuracy was lower.

The historical storm records showed no longer median rainfall-to-flow lag in SE. For isolated warm storms in 1981–2000 with a measurable rise above baseline, the median lag from rainfall peak to discharge peak was one day in both groups, with an interquartile range of one to two days (329 IE and 129 SE catchments). Response magnitude, however, varied with the gain in peak-date accuracy within each group. Catchments with larger historical discharge rises per unit of rain tended to gain more from discharge when dating precipitation peaks in 2011–2022 (Spearman correlations of 0.33 in IE and 0.44 in SE).

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

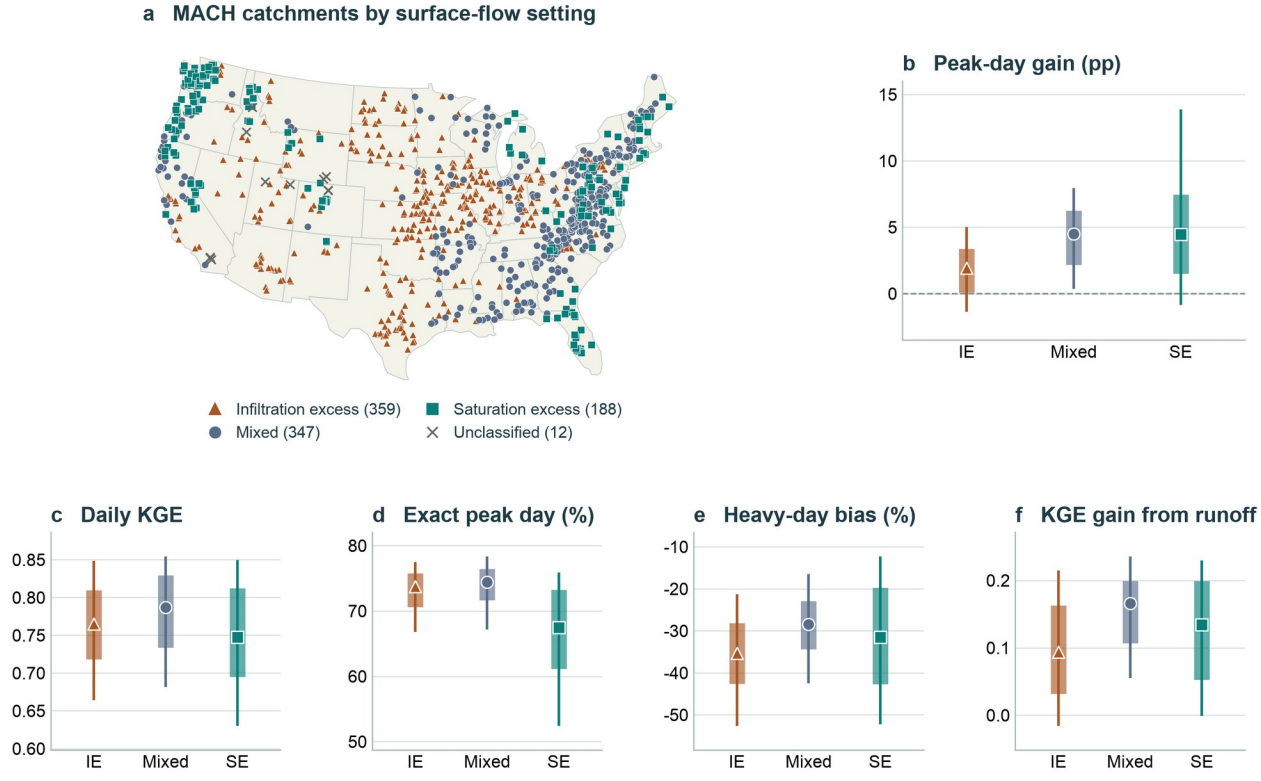


Figure 7. Reconstruction by mapped surface-flow setting. (a) Outlets of all 906 evaluation catchments. IE and SE denote relatively greater infiltration-excess and saturation-excess surface flow, respectively. Grey crosses indicate unclassified gauges. (b, f) Within-catchment gains from adding runoff to extended meteorology in exact peak-date success (percentage points, pp) and daily KGE. (c–e) Daily KGE, exact peak-date success and amount bias on reference heavy days for extended meteorology with runoff. All scores are for 2011–2022. Symbols show catchment medians. Thick and thin lines span the 25th–75th and 10th–90th percentiles.

4 Discussion

4.1 Implications

This study shows that EA-LSTM can combine runoff observations and meteorological variables to learn statistical relationships between precipitation and runoff responses from large-sample hydrological data and use them for precipitation reconstruction. Comparisons under two meteorological settings show how the value of runoff changes with the accompanying information. Common input availability and evaluation dates make these differences interpretable as conditional predictive gains.

A stronger runoff signal need not identify either the amount or arrival time of new precipitation uniquely. Storage deficits, delayed release and responses to neighbouring events can alter the relationship between precipitation and outlet discharge. In two monitored catchments, Knapp et al. (2025) found stronger hydrological responses under wetter conditions without a

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

corresponding increase in new-water fraction. Mobilisation of stored water can therefore transmit a recent rainfall signal without requiring the new rainfall itself to reach the outlet quickly. These processes offer possible explanations for differences among catchments.

The contribution of runoff to peak-date identification also varied with surface-flow setting. Among isolated warm storms that produced a discharge rise, the median peak lag was one day in both the IE and SE groups. The groups differed chiefly in identifying the wettest day within a detected wet spell. Without discharge, peak-date success was lower in SE, and this group gained more when discharge was added. The SE catchments had a wetter climate. Under wet conditions, hillslopes or saturated areas near the stream may become better connected to the channel, giving the outlet rise more information about the peak day (Dunne and Black, 1970; Penna et al., 2011). This interpretation is consistent with the association between larger rises and larger peak-date gains within each group. Establishing the role of wetness in the difference between groups would require soil moisture and runoff-generation observations for the same storms.

The distinction between detection and amount also matters when interpreting improvement. Wet-day F1 treats days as classifications rather than weighting them by precipitation volume. Peak-date success includes missed events, whereas heavy-day bias assesses reconstructed amounts on fixed reference days. Higher detection scores can therefore coexist with persistent amount deficits. In one semi-arid watershed, Dastjerdi and Nasseri (2026) found that soil-moisture/runoff-based inversions could improve quantitative agreement relative to gridded precipitation products while retaining lower precipitation-detection probability. This comparison reinforces the need to assess amounts and occurrence separately rather than interpreting either as a complete measure of reconstruction quality.

Temporal aggregation provides a further perspective. Kretzschmar et al. (2016) showed that the temporal resolution of inferred effective rainfall depends on catchment dynamics, rainfall regime and model fit. Our comparisons indicate that aggregation from daily to three- to seven-day scales can partly cancel opposing daily errors and improve reconstruction accuracy. However, a three-day amount close to the precipitation reference still accompanied a displaced peak (Fig. 6), and a distinct heavy-precipitation pulse was reconstructed with a lower maximum. Better accumulated agreement therefore does not demonstrate recovery of the daily sequence, highlighting the greater difficulty of accurate daily reconstruction than reconstruction at three- to seven-day scales.

Heavy precipitation is a more demanding application. Runoff markedly improved identification of heavy days under extended meteorology, but substantial amount deficits remained. Additional bias assessment is therefore needed before using these reconstructions directly in analyses sensitive to extreme precipitation quantities. High-elevation catchments merit particular scrutiny: they benefited strongly from the meteorological extension but retained large heavy-day deficits, weak additional runoff gains and nonmonotonic accumulation benefits.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

Observational requirements may also limit applicability. Models with runoff need local discharge records, meteorological covariates and data for constructing static climate attributes; the reliability and accuracy of these data matter. Catchment attributes may also need to better represent hydrological processes and response characteristics. Otherwise, the hydrological relationships learned by deep learning may be inaccurate and their transferability across catchments limited.

4.2 Data dependence and remaining limitations

One limitation of the extended meteorological inputs is that their generation already uses precipitation information. Shortwave radiation generation includes a precipitation-day adjustment, and the vapour-pressure correction for dry climates also uses precipitation information (Thornton et al., 2021). The extended inputs may therefore already carry some information about precipitation days. ERA5-Land is also indirectly affected by observation assimilation through ERA5 atmospheric forcing (Muñoz-Sabater et al., 2021), but this dependence is not equivalent to the direct use of precipitation information in Daymet variable generation. A deterministic model provides one reconstruction rather than multiple precipitation sequences and their ranges consistent with the inputs. Probabilistic or ensemble reconstruction may offer opportunities, particularly for quantifying extreme precipitation intensity. MACH also includes catchments with potential human influences; regulation and other human alterations may weaken or reshape the relationship between precipitation and observed runoff. Finally, this study uses one architecture and a fixed subsequent-observation window. Future studies can assess robustness over new periods and with independent precipitation references.

5 Conclusions

This study evaluated precipitation reconstruction using deep learning with large-sample runoff observations and meteorological data. Across 906 MACH catchments, extending the meteorological inputs of an EA-LSTM with a fixed seven-day subsequent-observation window improved daily precipitation reconstruction and event identification. Median daily KGE reached 0.771 and median exact peak-date success reached 73.3%. Runoff supplied less additional daily KGE under extended meteorology but remained valuable for identifying heavy days. Compared with daily results, temporal aggregation over three to seven days achieved higher accuracy, with limited gains from longer periods. Heavy-day amounts remained underestimated. Adding discharge improved peak-date success more in the saturation-excess group, while overall peak-date success was higher in the infiltration-excess group. The study demonstrates the potential of combining deep learning with large-sample hydrological and meteorological data for precipitation reconstruction.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

Code and data availability

MACH version 4 is available from Zenodo (Sink, 2026, https://doi.org/10.5281/zenodo.18686475). The study code is available from the corresponding author on request.

Author contributions

Min Wang: Conceptualization, Formal analysis, Writing – original draft. Haiyang Shi: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This work was supported by the Xinjiang Academy of Agricultural Sciences (grant no. xjnyywdzx-2026001-53).

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, Sci. Data, 5, 170191, https://doi.org/10.1038/sdata.2017.191, 2018.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration – Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization of the United Nations, Rome, 1998.

Dastjerdi, P. A. and Nasseri, M.: Comparing novel backward hydrological models for watershed-scale precipitation estimation: an evaluation of inverted PDM and Kirchner-hybrid structures, Sci. Rep., 16, 14265, https://doi.org/10.1038/s41598-026-42647-0, 2026.

Dunne, T. and Black, R. D.: Partial area contributions to storm runoff in a small New England watershed, Water Resour. Res., 6, 1296–1311, https://doi.org/10.1029/WR006i005p01296, 1970.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Henn, B., Clark, M. P., Kavetski, D., and Lundquist, J. D.: Estimating mountain basin-mean precipitation from streamflow using Bayesian inference, Water Resour. Res., 51, 8012–8033, https://doi.org/10.1002/2014wr016736, 2015.

Henn, B., Newman, A. J., Livneh, B., Daly, C., and Lundquist, J. D.: An assessment of differences in gridded precipitation datasets in complex terrain, J. Hydrol., 556, 1205–1219, https://doi.org/10.1016/j.jhydrol.2017.03.008, 2018.

Herrnegger, M., Nachtnebel, H. P., and Schulz, K.: From runoff to rainfall: inverse rainfall–runoff modelling in a high temporal resolution, Hydrol. Earth Syst. Sci., 19, 4619–4639, https://doi.org/10.5194/hess-19-4619-2015, 2015.

Hong, Y., Abeshu, G. W., Li, H.-Y., Wang, D., Zhao, M., Wild, T., Blöschl, G., and Leung, L. R.: A unified scheme for modeling saturation and infiltration excess runoff, Hydrol. Earth Syst. Sci., 30, 4293–4303, https://doi.org/10.5194/hess-30-4293-2026, 2026.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45, W02429, https://doi.org/10.1029/2008wr006912, 2009.

Knapp, J. L. A., Berghuijs, W. R., Floriancic, M. G., and Kirchner, J. W.: Catchment hydrological response and transport are affected differently by precipitation intensity and antecedent wetness, Hydrol. Earth Syst. Sci., 29, 3673–3685, https://doi.org/10.5194/hess-29-3673-2025, 2025.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019.

Kretzschmar, A., Tych, W., Chappell, N. A., and Beven, K. J.: Reversing hydrology: quantifying the temporal aggregation effect of catchment rainfall estimation using sub-hourly data, Hydrol. Res., 47, 630–645, https://doi.org/10.2166/nh.2015.076, 2016.

Liao, M. and Barros, A. P.: Toward optimal rainfall – Hydrologic QPE correction in headwater basins, Remote Sens. Environ., 279, 113107, https://doi.org/10.1016/j.rse.2022.113107, 2022.

Manoj J, A., Loritz, R., Gupta, H., and Zehe, E.: Can discharge be used to inversely correct precipitation?, Hydrol. Earth Syst. Sci., 29, 6115–6135, https://doi.org/10.5194/hess-29-6115-2025, 2025.

McMillan, H., Araki, R., Bolotin, L., Kim, D.-H., Coxon, G., Clark, M., and Seibert, J.: Global patterns in observed hydrologic processes, Nat. Water, 3, 497–506, https://doi.org/10.1038/s44221-025-00407-w, 2025.

Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C., and Haghdoost, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, Sci. Data, 12, 416, https://doi.org/10.1038/s41597-025-04610-y, 2025.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Penna, D., Tromp-van Meerveld, H. J., Gobbi, A., Borga, M., and Dalla Fontana, G.: The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment, Hydrol. Earth Syst. Sci., 15, 689–702, https://doi.org/10.5194/hess-15-689-2011, 2011.

https://doi.org/10.5194/egusphere-2026-6840
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Sink, K. and Brikowski, T.: MACH: A Multi-Attribute Catchment Hydrometeorological dataset, Sci. Data, 13, 816, https://doi.org/10.1038/s41597-026-07162-x, 2026.

Sink, K.: MACH: A Multi-Attribute Catchment Hydrometeorological dataset (version 4.0), Zenodo [data set], https://doi.org/10.5281/zenodo.18686475, 2026.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. L.: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons, Rev. Geophys., 56, 79–107, https://doi.org/10.1002/2017rg000574, 2018.

Tarasova, L., Basso, S., Poncelet, C., and Merz, R.: Exploring controls on rainfall-runoff events: 2. Regional patterns and spatial controls of event characteristics in Germany, Water Resour. Res., 54, 7688–7710, https://doi.org/10.1029/2018WR022588, 2018.

Thornton, P. E., Shrestha, R., Thornton, M., Kao, S.-C., Wei, Y., and Wilson, B. E.: Gridded daily weather data for North America with comprehensive uncertainty quantification, Sci. Data, 8, 190, https://doi.org/10.1038/s41597-021-00973-0, 2021.