



Can discharge be used to inversely correct precipitation?

Ashish Manoj J¹, Ralf Loritz¹, Hoshin Gupta², and Erwin Zehe¹

¹Chair of Hydrology, Institute of Water and Environment (IWU), Karlsruhe Institute of Technology,
76131 Karlsruhe, Germany

²Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson 85721, United States

Correspondence: Ashish Manoj J (ashish.jaseetha@kit.edu)

Received: 27 November 2024 – Discussion started: 18 December 2024

Revised: 20 October 2025 – Accepted: 21 October 2025 – Published: 11 November 2025

Abstract. This study explores the feasibility of using the information contained in observed streamflow measurements to inversely correct catchment-average precipitation time series provided by reanalysis products at the continental scale. We explore this possibility by training LSTM ensemble networks to inversely predict precipitation by using the streamflow of catchments as additional input. The first model uses discharge as an input feature along with other meteorological variables, while the second model uses only the meteorological predictors. Analysing the performance of both models showed that the discharge information not only led to an average improvement overall, but also resulted in a significant improvement (around 30 %) on days with precipitation amounts greater than 5 mm. An out-of-sample test showed that the inversely estimated precipitation is better able to reproduce small-scale, high-impact events that are poorly represented in the reanalysis product. Further, using the inversely generated precipitation time series for classical hydrological “forward” modeling resulted in improved estimates for streamflow and soil moisture. Given that the wealth of streamflow gauges around the world is currently underutilised for meteorological applications, our findings have significant implications for achieving better estimates of precipitation associated with high-impact flood events.

as they represent cause-and-effect in the context of system dynamics. Long-term experimental data from well-studied research catchments, and data from operational monitoring networks, have thus long been the cornerstone of the hydrological sciences (Tetzlaff et al., 2017). The relevance of observed data and research observatories cannot be overemphasised, particularly due to the invalidity of stationarity assumptions (Milly et al., 2008) in the face of anthropogenic climate change and its impacts on water-related hazards and availability.

As the availability and quality of observations crucially constrain the “realism” of a hydrological model and thus the accuracy of predictions, data scarcity impedes accurate modelling and inference of hydrological processes. Global reanalysis products (Muñoz-Sabater et al., 2021; Onogi et al., 2007; Rienecker et al., 2011) can potentially, if of sufficient quality, complement the few existing ground-based observations by offering a valuable alternative when exhaustive local observations are not available. Further, they play a pivotal role in hydro-climatic research (Alexopoulos et al., 2023; Gu et al., 2023), by providing a consistent, long-term view of the state of the global climate system via the assimilation of measurements and monitoring data into numerical weather models.

While previous studies (Essou et al., 2016; Tarek et al., 2020) have already shown the value of using reanalysis data as estimates for meteorological forcing data in regions with little or sparse ground-based weather station data, serious concerns about their quality remain when used in the context of hydrological modelling. The main issues include (Tarek et al., 2020) (i) regional variations in data quality and (ii) limited representation of local hydro-meteorological processes, with both of these impacting/biasing model structures and

1 Introduction

The performance of hydrological models has traditionally been constrained by the availability and quality of observations covering various aspects of the water cycle. Among those, precipitation and streamflow observations are pivotal,

simulated states and fluxes. Systematic biases are also critical obstacles to the broader applicability of such products (Clerc-Schwarzenbach et al., 2024). In the case of ERA5-Land, a component of the Copernicus Climate Change Service (C3S) provided by the European Centre for Medium Weather Forecasting (Muñoz-Sabater et al., 2021), there is a known tendency to significantly overestimate potential evapotranspiration (Clerc-Schwarzenbach et al., 2024; Kratzert et al., 2023; Xu et al., 2024). Deficiencies have also been documented in the representation of convective storms (Essou et al., 2016; Taszarek et al., 2021) with subsequent underestimation of precipitation magnitudes and intensities (Manoj J et al., 2024).

It is important to stress that “true” precipitation estimates are per default unknown at the catchment scale. We obtain estimates of them (with considerable uncertainty) by either interpolating data from stations in or surrounding the catchment or averaging gridded data from reanalysis/remote sensing products to the catchment scale. Such precipitation uncertainty is rarely considered when quantifying model output uncertainty; while studies are usually conducted to show how differences in simulated discharge can be as a consequence of changing precipitation input, they rarely look at how much improvement of the model performance would be possible by using different but plausible precipitation (Bárdossy et al., 2020, 2022).

Because precipitation forcing data plays a crucial role in rainfall–runoff modelling, several methods (Yumnam et al., 2022) have been suggested for correcting precipitation data. These range from the use of storm multipliers (Sun and Bertrand-Krajewski, 2013) to station-wise correction of data using a gauge-based precipitation network (Cornes et al., 2018). However, gauge-based methods require a sufficient number of weather stations (Agarwal et al., 2020), which is often not the case for most regions around the world. As seen from previous experience, the observation network is too sparse even in data rich regions, and the majority of high-impact rainstorms are simply not observed (Borga et al., 2008). This is particularly true for flash floods in response to convective storm activity (Manoj J et al., 2024; Meyer et al., 2022; Villinger et al., 2022) and well related to the classical “Predictions in Ungauged Basins – PUB problem” (Sivapalan et al., 2003). To overcome this problem, and in line with Kirchner’s (2009) work on “doing hydrology backwards”, this paper explores options for inverse estimation of precipitation using the information contained in observed streamflow. The goal is to determine whether inverse estimation at the catchment scale can refine precipitation estimates from reanalysis products, ensuring they are hydrologically consistent, especially for extreme events.

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. Kirchner (2009) reported an early and successful attempt to infer catchment average rainfall and evaporation time series from streamflow fluctuations and inspired several investigations examining the advantages and limitations of doing “hydrology backwards” in diverse catchments (Krier et al., 2012; Teuling et al., 2010). Although these studies have established a robust mathematical foundation for addressing the inverse hydrological problem, they were limited to smaller, well-monitored research catchments. This raises questions about the applicability of this approach to larger catchments as well as to smaller, non-experimental ones.

Note that inversions of the catchment water balance are inherently ill-posed, making it near impossible to find a unique solution (Bishop, 2006). Adopting the concept of micro- and macro-states from statistical mechanics (Zehe and Blöschl, 2004), we argue that the exact micro-state, i.e. the “true” space–time pattern of precipitation in the catchment, is neither uniquely identifiable nor observable. Yet, we conjecture that streamflow data (being an integral response from a potentially large and heterogeneous data) can reduce the uncertainty associated with this process, because it provides valuable information on antecedent precipitation and the current state of the catchment. As streamflow remains a non-linear convolution of the catchment-average precipitation, we propose that machine learning is well suited to this problem. Deep learning has recently fertilised almost all fields of the natural sciences and engineering, showing great promise in solving a wide range of inverse problems, especially those related to imaging (Ongie et al., 2020). It has also been argued that such models can provide meaningful and general benchmarks for hypothesis testing (Klotz et al., 2022; Nearing and Gupta, 2015) and afford powerful avenues for generalisation using large datasets (Loritz et al., 2024b).

The overall objective of this study is “do hydrology backwards” using regional-scale long short-term memory (LSTM) network ensemble models trained on large-scale hydrological datasets. While ERA5 Land (Table 1: Muñoz-Sabater et al., 2021) has well-documented issues in representing the driving precipitation estimates for specific event scales (Essou et al., 2016; Manoj J et al., 2024), recent studies (Bandhauer et al., 2022; Goteti and Famiglietti, 2024) have shown that they hold considerable promise to tackle the “Predictions in Ungauged Basins – PUB problem”. This makes it an ideal test candidate for an inverse correction using streamflow and observational precipitation estimates over the same region (E-OBS: Cornes et al., 2018). The underlying research question is, “How much information about the catchment-average precipitation is effectively encoded in the variability of the streamflow time series observed at the outlet?” To answer this question, we first look at the performance gain in using discharge for predicting precipitation by focusing on days with higher precipitation magnitudes and then investigate whether the approach can accurately replicate the spatial characteristics of the original observational dataset (by look-

ing at various time series measures) across European catchments for an unseen testing period. We then examine how the inverse model performs when moving to much smaller (50–200 km²; Table 2) out-of-sample catchments. Here, we compare (using the event runoff coefficients) LSTM-based inverse estimates during flood events to the original reanalysis product (ERA5 Land) and rain gauge-based observational estimates over the same region (E-OBS). Finally, we use a conceptual hydrological model (HBV: Bergström and Forsman, 1973) and a process-based model (CATFLOW: Zehe et al., 2001) to assess the quality of the precipitation estimates for forward modelling of streamflow and soil moisture dynamics, respectively.

2 Data and Methods

2.1 Model Configuration

LSTMs (Hochreiter, 1998) are a special type of recurrent neural network that makes use of cell states and so-called “gates” to control the information flow through the network. The LSTM model used in this study extends upon the work of Kratzert et al. (2018) and Acuña Espinoza et al. (2024). The LSTM architecture, which is commonly used for streamflow simulation in hydrology (Kratzert et al., 2018) uses a sequence of meteorological variables, such as precipitation and temperature as dynamic inputs, along with catchment attributes as static features, to predict the corresponding streamflow. In our setting, to establish an inverse model, we use the same general model architecture as in previous studies (Acuña Espinoza et al., 2024; Loritz et al., 2024b). The key difference is that future streamflow is now used along with other dynamic and static data as inputs (Table A1 in Appendix A) in order to estimate the precipitation forcings of the catchments. To account for the time lag between precipitation and discharge response observed at the catchment outlet, the model was provided with 7 d lead time series for discharge. We explored ranges of hyperparameter settings on a smaller subset of the training dataset to establish relatively stable hyperparameter configurations (Fig. S1 in the Supplement), finally setting them according to (Acuña Espinoza et al., 2024) with a reduced number (5) of training epochs. Table A2 in Appendix A indicates the values used for the LSTM network hyperparameters. Mean squared error was used as the training loss function. In accordance with standard practices in the deep learning community, we utilise an ensemble network for LSTM predictions. In all cases, three individual LSTM models (with different initialisation seeds) were trained, and we present the mean predictions for the remainder of this paper.

The codes for model building and training can be found online (Manoj J, 2025b). The LSTM was trained as a regional model (single network trained on all available catchments) based on the openly available datasets detailed in the

next section (Sect. 2.2). For forward hydrological modelling using the inversely-generated precipitation timeseries estimates, we use two hydrological models (Appendix B) – the lumped conceptual HBV model (Hydrologiska Byråns Vattenbalansavdelning: Bergström and Forsman, 1973) and the spatially distributed process-based CATFLOW model (Zehe et al., 2001).

2.2 Data sets

This study utilized the Caravan dataset (Kratzert et al., 2023) to investigate our hypothesis regarding the inverse identifiability of precipitation from information about discharge dynamics. We trained our model on European catchments from the GRDC-Caravan (Färber et al., 2023) community extension and the original Caravan dataset, which includes catchments from CAMELS-GB (Coxon et al., 2020). The Caravan dataset uses the ERA5 Land (Muñoz-Sabater et al., 2021) as the primary meteorological forcing, while the catchment attributes include data from HydroATLAS (Linke et al., 2019). The discharge data is tapped from relevant state and national authorities and is accessible as open datasets. The observational E-OBS precipitation product (v31.0 – Cornes et al., 2018), which uses the station network of the European Climate Assessment & Dataset (ECA&D) project, was used as the training target for the model runs. Figure S2 in the Supplement depicts the study catchments (1800 in total) in the training dataset.

We chose a training period of around 25 years between 1 October 1980 to 30 September 2005. Following the best practices in data-based modelling, the model was tested on an unseen testing period between 2006 and 2020 (2015 for CAMELS-GB catchments due to data unavailability). To investigate its generalizability across scales, we also tested the model on four catchments (Figs. S3 and S4 in the Supplement) that were not included in the original training set (Sect. 2.3.2). For the out-of-sample test, we made use of data from the Caravan Spain (Casado Rodríguez, 2023) and Caravan Switzerland (Höge et al., 2023) extensions, in addition to data from local data providers in Germany (Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg – LUBW) and Luxembourg (Nijzink et al., 2024). To validate the inversely generated precipitation (Sect. 2.3.3) during forward modeling, we conducted hydrological model simulations in the Elsenz Schwarzbach and Lippe catchments (Fig. S5 in the Supplement). Table 1 provides an overview of the main datasets used in this study, detailing their spatial and temporal resolutions, as well as their sources.

Table 1. Brief overview of the datasets used in this study, including their spatial and temporal resolution.

Dataset	Type and Source	Spatial Resolution	Temporal Resolution	Details
Caravan	Hydrometeorological dataset (Kratzert et al., 2023)	Catchment scale	Daily	Open community dataset that includes catchment forcing data and attributes along with streamflow.
GRDC-Caravan	Hydrometeorological dataset (Färber et al., 2023)	Catchment scale	Daily	Community extension to the Caravan dataset, incorporating data from the Global Runoff Data Centre (GRDC).
ERA5 – LAND	Reanalysis product (Muñoz-Sabater et al., 2021)	$0.1^\circ \times 0.1^\circ$	Hourly (aggregated to daily)	Reanalysis product produced by replaying the land component of ERA5 climate reanalysis
E-OBS	Gridded observational precipitation product (Cornes et al., 2018)	$0.25^\circ \times 0.25^\circ$	Daily	Interpolated observational precipitation product utilizing the station network from the European Climate Assessment & Dataset (ECA&D) project.
Caravan Spain	Hydrometeorological dataset (Casado Rodríguez, 2023)	Catchment scale	Daily	Community extension to the Caravan dataset, incorporating data from Spain.
Caravan Switzerland	Hydrometeorological dataset (Höge et al., 2023)	Catchment scale	Daily	Community extension to the Caravan dataset, incorporating data from CAMELS-CH catchments.
Caravan Germany	Hydrometeorological dataset (Dolich et al., 2025)	Catchment scale	Daily	Community extension to the Caravan dataset, incorporating data from CAMELS-DE catchments.
MERRA-2	Reanalysis product (Gelaro et al., 2017)	$0.625^\circ \times 0.5^\circ$	Hourly (aggregated to daily)	Global atmospheric reanalysis by NASA Global Modeling and Assimilation Office (GMAO) using the Goddard Earth Observing System Model (GEOS)
GLDAS-2.2	Reanalysis product (Li et al., 2019)	$0.25^\circ \times 0.25^\circ$	Daily	NASA Global Land Data Assimilation System model outputs with data assimilation for the Gravity Recovery and Climate Experiment (GRACE-DA)

2.3 Experimental Design

2.3.1 Exploring information about precipitation encoded in streamflow

To shed light on the value of discharge for inversely predicting precipitation, we conducted a virtual experiment (Fig. 1) in which two LSTM ensemble models (Tables A1 and A2 in Appendix A) were trained using the same catchments and training period. The first model (*without_discharge*) used only ERA5 Land meteorological time series (total_precipitation, air_temperature, solar_and_thermal_radiation) and static attributes (area, ele_mt_sav, frac_snow, pet_mm_syr; Kratzert et al., 2023), while the second model (*with_discharge*) included lagged discharge as an additional input variable. Both models were trained to predict daily

catchment average precipitation sums from the observational EOBS product. Therefore, we only deal with spatially averaged timeseries for precipitation, assuming that these values represent the actual precipitation over the entire catchment.

We then used both the trained regional-scale models (*with_discharge* and *without_discharge*) to predict the precipitation time series inversely for all the test catchments over the unseen testing period and evaluated (Appendix C) those using the mean wet day precipitation (MWD) – mm d^{-1} , 95th percentile limit (R95P) – mm d^{-1} , and Spearman autocorrelation values (SL) for each catchment, and then compared them to the values from ERA5 Land (the reanalysis product we want to improve) and E-OBS (observational product used as training target) at the continental scale.

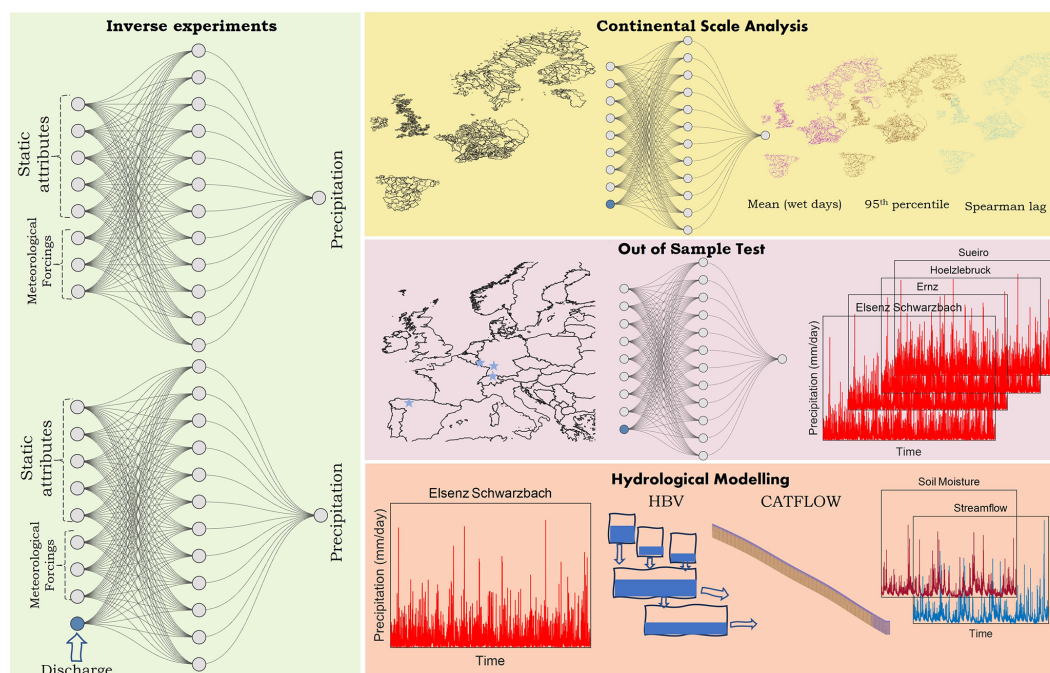


Figure 1. Schematic representation of our methodological approach. Each rectangular panel indicates different stages of our workflow. Initially, we train two LSTM ensemble networks to predict catchment average precipitation through inverse experiments (Sect. 2.3.1). The trained models are then utilized for a continental-scale analysis before being used for out-of-sample testing (Sect. 2.3.2). Finally, a validation exercise for the inversely generated precipitation is conducted using various hydrological models (Sect. 2.3.3).

2.3.2 Out of sample precipitation inversions and their quality

We further tested the feasibility of knowledge transfer to out-of-sample catchments and used the same regional-scale models (*with_discharge* and *without_discharge*) to inversely predict the intensity of driving rainstorms for selected flood events in four hydro-climatically diverse and much smaller catchments (not included in the original training dataset). These catchments (Table 2 and Figs. S3 and S4) were chosen based on the severity of the flooding and on the apparent inability of ERA5 Land forcings to accurately represent the storms that triggered the flood events.

2.3.3 The potential of inverted precipitation for forward modelling

To evaluate the value of generated precipitation data for forward modeling of streamflow, we calibrated the HBV conceptual hydrological model (Bergström and Forsman, 1973) over the Elsenz Schwarzbach (Manoj J et al., 2024) and Lippe (camelsde_DEA11130: Loritz et al., 2024a) catchments (Fig. S5) using both the original ERA5 Land and the *with_discharge* LSTM-generated precipitation timeseries and compared the evaluation period performance of both model versions (Table B1 in Appendix B). The HBV model (Appendix B) used in this paper requires precipitation (ERA5

Land/LSTM simulated), potential evapotranspiration, and air temperature as inputs. We follow the recommendations of Clerc-Schwarzenbach et al. (2024), similar to that of Loritz et al. (2024a), for the calculation of potential evapotranspiration, and use the temperature-based Hargreaves formula detailed by Adam et al. (2006).

Complementary to streamflow modelling, the performance of a hydrological model can also be judged by how well it replicates the catchment dynamics of a region. Soil moisture is a key variable controlling the partitioning of net radiation into sensible and latent heat (Seneviratne et al., 2010) or overland flow during a rainstorm (Zehe and Blöschl, 2004). We thus used each precipitation estimate (*with_discharge* LSTM and ERA5 Land) to run the process-based hillslope scale model CATFLOW (Appendix B), using a setup from Manoj J et al. (2024) used for uncalibrated predictions of local floods. Here, we focused on one of the headwater sub-catchments (Catchment W32 in Fig. S5) within the Elsenz Schwarzbach. The model simulated (Table B1) the period from 1 January 2008 to 31 December 2015 using each of the ERA5 Land and *with_discharge* LSTM precipitation estimates, and the corresponding spatially averaged soil moisture states were compared against several soil moisture re-analysis products (Table 1: due to the unavailability of observed data). These include (a) ERA5 Land: Muñoz-Sabater et al., 2021, (b) GLDAS (NASA Global Land Data Assimilation System, GLDAS-2.2 GRACE DA: Li et al., 2019)

Table 2. Attributes for the four catchments used for out-of-sample testing.

Catchment	Country	Area (km ²)	Mean precipitation (mm d ⁻¹)	Mean potential evapotranspiration (mm yr ⁻¹)	Mean elevation (m)
Elsenz-Schwarzbach	Germany	196.5	2.51	812.85	246.7
Ernz	Luxembourg	69.3	2.31	724.04	345.5
Sueiro	Spain	132.5	3.31	873	381
Hoelzlebruck	Germany	47.1	4.14	658	980

and (c) MERRA (Modern-Era Retrospective analysis for Research and Applications version 2 – `avg1_2d_lnd_Nx`; Gelaro et al., 2017).

3 Results

3.1 The information contained in streamflow about precipitation

Figure 2 shows violin plots displaying the pairwise difference in the mean performance of the two LSTM models (Fig. A1 in Appendix A) over the catchments ($n = 1800$) in the test dataset for varying precipitation amounts (All days, days with daily precipitation greater than 1 mm and days with daily precipitation greater than 5 mm). Each point denotes the difference in NSE (Appendix C) for individual catchments while making predictions using the *with_discharge* model compared to the *without_discharge* model. A marked shift towards higher positive differences indicates that the model “*with_discharge*” has higher NSE values than the model “*without_discharge*”. This holds true not only on average but also with respect to the best-performing catchments. The median NSE metric value (Nash and Sutcliffe, 1970) for the regional LSTM model (considering entire time series) across the study catchments is about 13 % higher when discharge is used as an additional predictor than when it is not. However, it is also observed that discharge information has worsened the performance in a few cases, likely due to the poor quality of streamflow data in these catchments. Analysing the performance improvement achieved by focusing on days with increasing precipitation amounts reveals that the gains are considerably greater on days with higher recorded precipitation (increase in median NSE value of about 29 % from 13 % as we look only at days with more than 5 mm precipitation). This largely answers our main research question and shows that the variability of discharge as measured in the catchment outlets holds enormous information about the driving storms over the entire catchment area. Consequently, we can utilise this information by applying a data-driven LSTM network. The information gain is naturally higher for more extreme precipitation events, as average streamflow conditions do not provide much information about the catchment scale precipitation.

3.2 Unraveling the Continental Scale Characteristics

To examine the characteristics of the simulated time series from the *with_discharge* and *without_discharge* models over the testing period in detail, we computed three time-series measures (Appendix C) namely mean wet day precipitation (MWD) – mm d⁻¹, 95th percentile limit (R95P) – mm d⁻¹, and Spearman autocorrelation values (SL) across all the catchments.

The continental-scale analysis reveals distinct patterns for the major European climatic regions. The spatial patterns for the mean wet day precipitation (Fig. 3d: MWD) obtained using the *with_discharge* LSTM model are well aligned to the ones from ERA5 Land (Fig. 3a) and EOBS (Fig. 3j). Higher daily average values are observed towards the Alps, the Carpathian Mountain ranges, and the coast of Norway, consistent with the climatology of these regions. In addition, we also see that the ERA5 Land largely matches the precipitation field’s characteristics (wet day mean and 95th percentile limit) as in the observational E-OBS product. This indicates that both products contain complementary information at such larger spatial scales.

For the 95th percentile of wet days (R95P), we again see a robust representation of the spatial differences, along with an underestimation of the magnitudes (Fig. 3b–k). The Spearman autocorrelation coefficient values (SL: Fig. 3c–l) indicate that while the models underestimate the mean and 95th percentile limits, they overestimate the autocorrelation (which indicates the persistence in the precipitation time series) compared to the ERA5 Land and EOBS time series.

Comparing the *with_discharge* and *without_discharge* models for MWD and R95P, we see that the addition of discharge information reduces the underestimation errors over the continental scale.

The higher autocorrelation values for the *with_discharge* (Fig. 3f) and *without_discharge* (Fig. 3i) may arise from model products incorporating catchment persistence, unlike the gridded observational E-OBS data. In the case of the *with_discharge* LSTM model, the higher values are likely due to the inclusion of strongly auto correlated streamflow data, which adds redundancy or a longer memory.

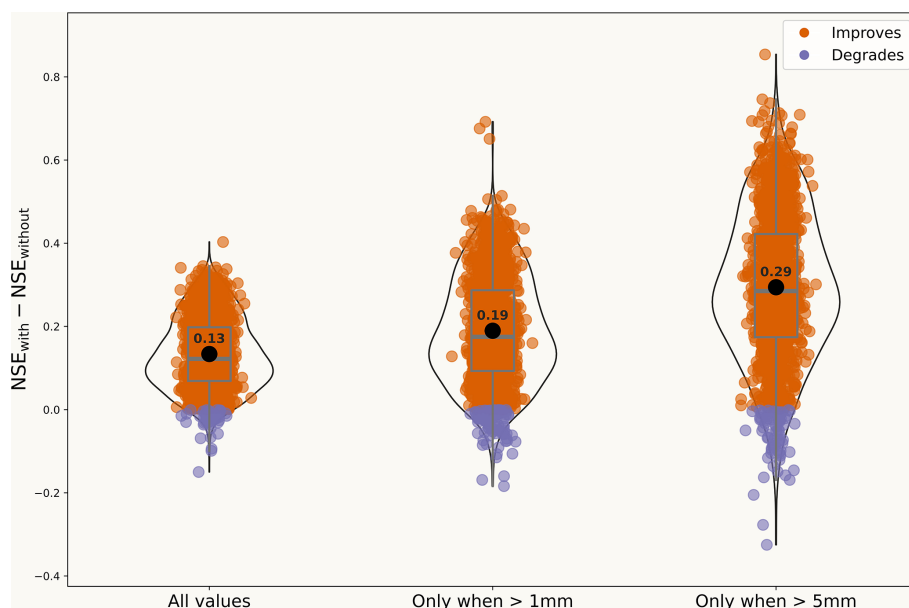


Figure 2. Comparison of performance gain for the *with_discharge* vs. *without_discharge* models in NSE for different precipitation amounts. The first violin plot illustrates the average improvement across all days in the testing period. The second and third plots display the mean performance gains over the catchments, specifically focusing on days where precipitation exceeded 1 and 5 mm, respectively.

3.3 Out of sample predictions

Figure 4 shows predicted event precipitation values over time for the four out-of-sample catchments. Again, we compare the inversely modelled values (*with_discharge* and *without_discharge*) to the ERA5 Land (the reanalysis product to be corrected) and the gauge-based E-OBS product (our training target). Table 3 lists the peak storm precipitation values reported by the different products along with the recorded flood values (both normalised to the catchment area in mm d^{-1}). Also shown are the storm runoff coefficients for the respective events based on the different precipitation estimates and discharge data.

Figure 4a represents the summer flood in June 2016 in the Elsenz Schwarzbach catchment in Germany. This annual flood event was triggered by a series of convective rainfall events caused by persistent atmospheric conditions in Germany during the summer of 2016. Localised rainfall totals exceeded 100 mm in some catchments (Bronstert et al., 2018), triggering widespread flash floods. Our previous work (Manoj J et al., 2024) indicated that the ERA5 Land reanalysis product could not accurately replicate the characteristics of the convective storm that caused this annual flood event over the Elsenz Schwarzbach catchment. The *with_discharge* LSTM simulated precipitation for this event was higher than the values reported by both ERA5 Land and the training target E-OBS, while the *without_discharge* model performed even worse than ERA5 Land.

A comparison of *with_discharge* LSTM-simulated precipitation values to radar estimates over the same region

(Manoj J et al., 2024) revealed the estimates to be closer than those reported by the observational E-OBS product. The runoff coefficient (Table 3) for the event also decreased from 48 % (ERA5 Land) to around 18 % (*with_discharge*), which is consistent with estimates from Manoj et al. (2024). The *with_discharge* LSTM model was also able to represent the second storm peak more accurately than ERA5 Land.

Next, the *with_discharge* model was used to estimate precipitation for another convective episode over the Ernzt Catchment in Luxembourg (Fig. 4b) in the summer of 2018. Once again, we observed that the model overestimated the peak precipitation compared to the observational E-OBS product used for training. However, the model benefited from integrating improved event timing information from ERA5 Land, which helped reduce timing errors compared to E-OBS. Essentially, the model combined information from both ERA5 Land and discharge to produce a storm estimate that was more consistent with the hydrology of the flood, taking into account both the volume and timing of the event, than the observational E-OBS product. In contrast, the *without_discharge* model again performed poorly for this event, resulting in an unrealistically high runoff coefficient of 4.34 (Table 3).

In the third catchment (Sueiro: camelses_1414 from Caravan Spain extension), the *with_discharge* estimate for storm forcing was higher than ERA5 Land and E-OBS (Fig. 4c). The corresponding runoff coefficients underline the reliability of the storm prediction from *with_discharge* (0.40) compared to E-OBS (0.79). For the Sueiro catchment (camelses_1414), the closest observational station is located

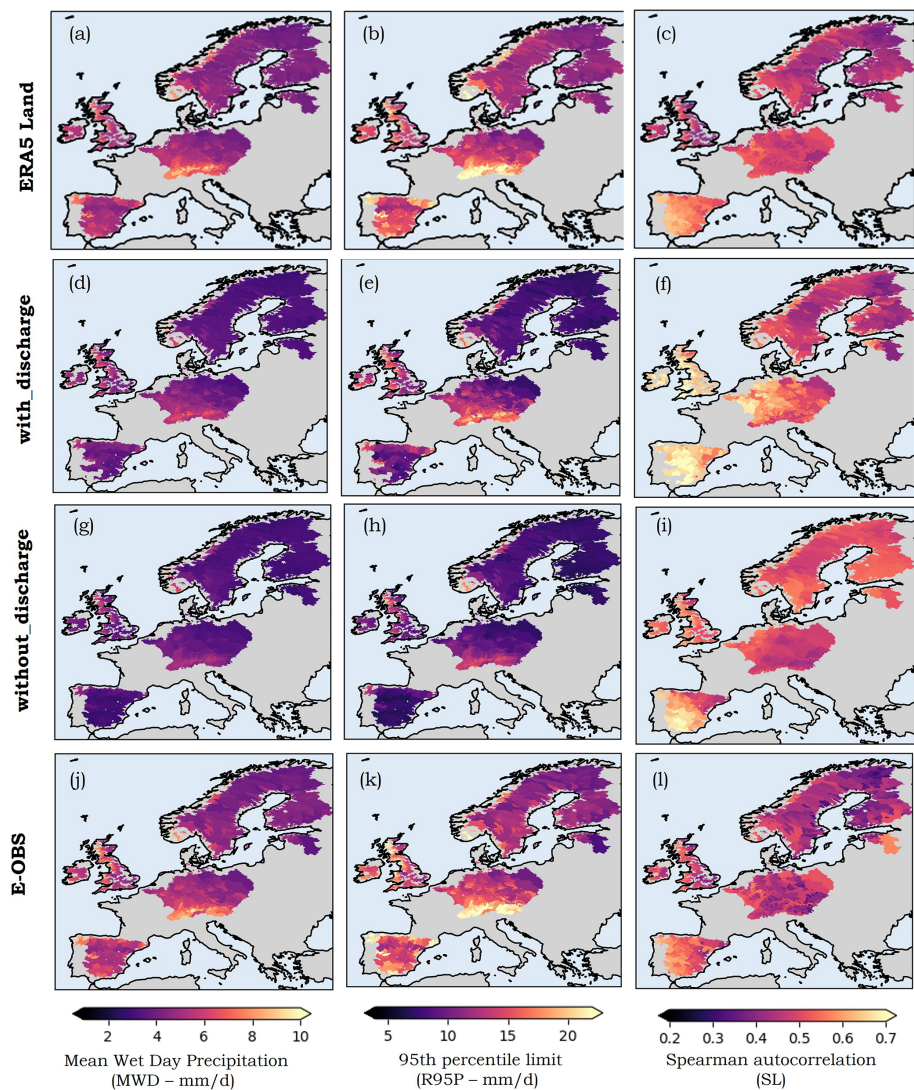


Figure 3. The spatial patterns of the different time series metrics (Appendix C) mean wet day precipitation (MWD) – mm d^{-1} , 95th percentile limit (R95P) – mm d^{-1} , and Spearman autocorrelation values (SL) over the study catchments for the different precipitation estimates – ERA5 Land (top row): (a–c), *with_discharge* LSTM model (second row): (d–f), *without_discharge* LSTM model (third row): (g–i) and E-OBS (bottom row): (j–l) from 2006 to 2020 (2015 for CAMELS-GB catchments).

Table 3. Event characteristics (storm volume and runoff coefficients) for the four out of sample catchments.

Event Characteristics		Elsenz-Schwarbach	Ernz	Sueiro	Hoelzlebruck
Precipitation (mm)	ERA5 Land	12.51	9.60	41.81	32.12
	<i>with_discharge</i>	32.79	42.75	58.53	50.85
	<i>without_discharge</i>	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
	<i>with_discharge</i>	0.18	0.63	0.40	0.38
	<i>without_discharge</i>	1.21	4.34	0.79	0.84
	E-OBS	0.30	0.52	0.79	0.43

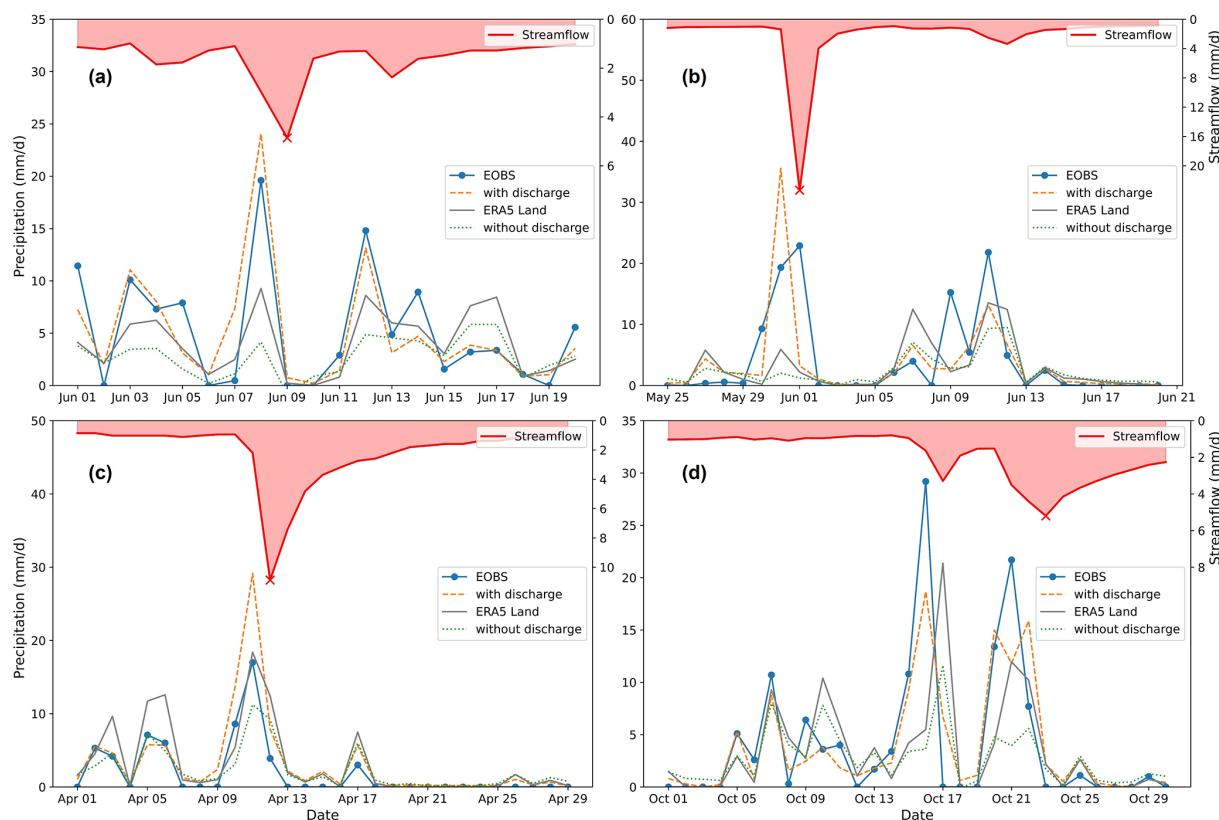


Figure 4. Precipitation estimates for flood events at four out-of-sample catchments: (a) Elsenz Schwarzbach, (b) Ernztal, (c) Sueiro, and (d) Hoelzlebruck. The red line represents the observed daily streamflow, with a cross marking the day of the flood event. The orange curve indicates the precipitation 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.

more than 60 km away (Fig. S4), explaining why the EOBS performs rather poorly in representing the driving forcings for the summer flood event.

In the Hoelzlebruck catchment (camelsch_4003 from Caravan Switzerland extension), two consecutive events occurred in October 2014. ERA5 Land was better than the *with_discharge* LSTM model in capturing the initial event magnitude, while the *with_discharge* model had better timing accuracy for the events (Fig. 4d). For the second event, which was the annual flood event, the *with_discharge* model, which incorporated streamflow information, was again able to reduce the relative errors in storm volume (Table 3). The *without_discharge* model showed the same timing error as ERA5 Land for the first storm; however, introducing discharge allowed the model to correct the timing bias.

3.4 Forward Hydrological Modelling

The precipitation estimates generated by the *with_discharge* LSTM model were then used to run classical hydrological models (HBV and CATFLOW; Table B1) in a forward manner. To address the question of performance in differently

sized basins, we run the conceptual HBV model in two catchments (Fig. S5) – Elsenz Schwarzbach (Fig. 5: 196.5 km²) and Lippe (Fig. 6: 3366.3 km²).

Figure 5d illustrates that the HBV model, which utilized the inverted precipitation estimates, performed better (NSE = 0.70) during the evaluation period over Elsenz Schwarzbach compared to the model driven by the ERA5 Land (NSE = 0.57). To gain a better understanding of the differences between the models, we visually examined the results for three individual flood events, as shown in Fig. 5a–c.

During the winter flood of December 2012 (23 December 2012, Fig. 5a), the model driven by ERA5 Land significantly underestimated both the peak and the volume of the flood event. When using *with_discharge*-simulated precipitation, the relative peak error decreased slightly. Similarly, the model runs using *with_discharge* precipitation more accurately captured the post-event conditions (28 December 2012). In the winter of 2015 (Fig. 5b), the model using *with_discharge* precipitation again demonstrated better performance. The model could more accurately represent the smaller flood peaks before the larger floods. This aligns with

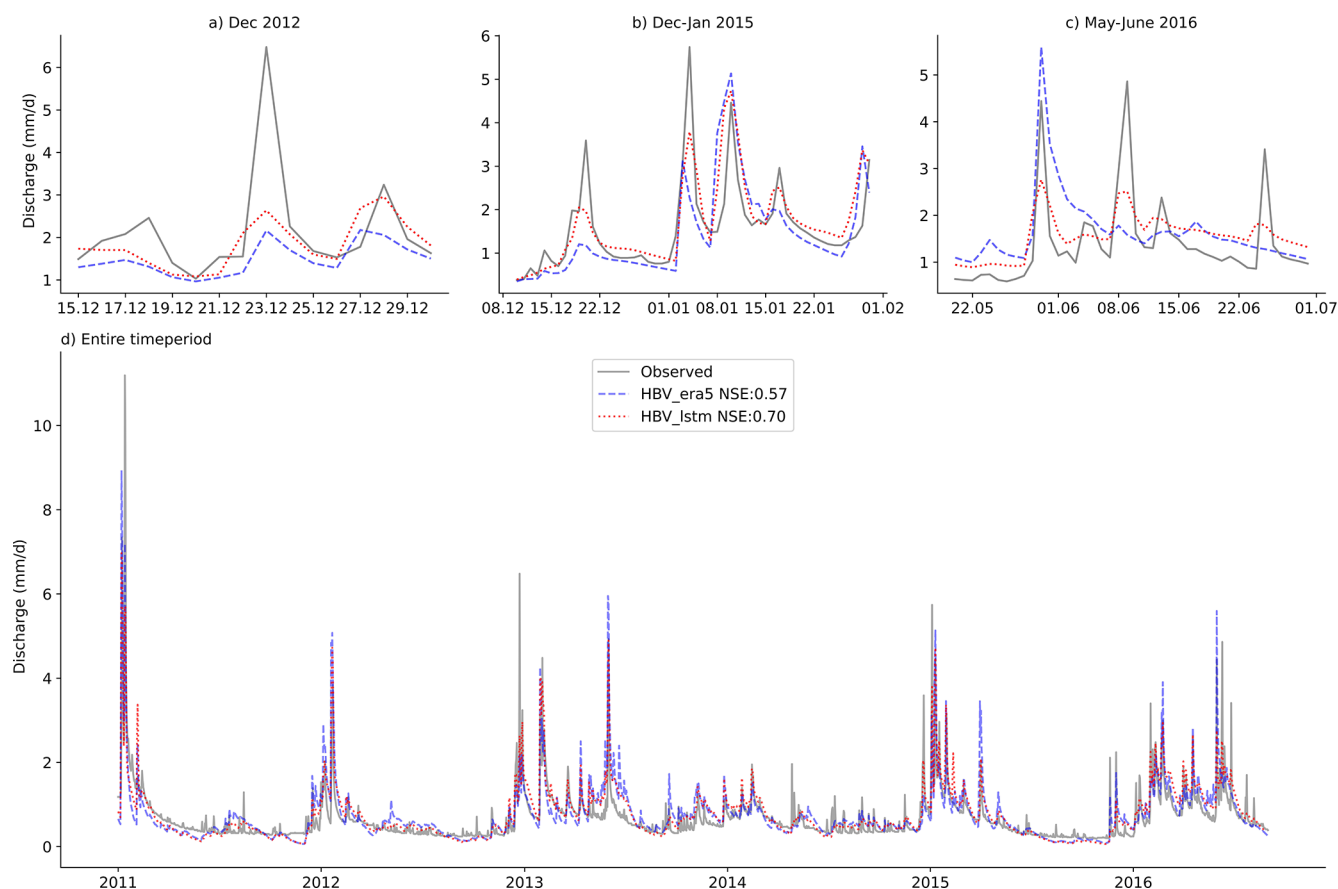


Figure 5. Observed (grey line) and simulated runoff (using the HBV model) at the Elsenz Schwarzbach catchment. The blue line denotes the streamflow simulated using the ERA5 Land precipitation product, while the red curve depicts the simulations using the inversely-estimate precipitation obtained using the *with_discharge* LSTM model. Moreover, three rainfall–runoff events are highlighted and displayed separately.

findings from other studies (Berghuijs et al., 2019; Manoj J et al., 2023) that emphasize the importance of initial conditions for floods across Europe.

During the convective summer storm event in June 2016 (Fig. 5c), neither model run successfully captured the flashy runoff response. Although the model that utilized ERA5 Land input predicted an earlier flood event in May 2016 with an overestimation bias, it did not accurately depict the dynamics of the annual flood event occurring a few days later. In contrast, the model with LSTM-generated precipitation (*with_discharge*) generally performed better in capturing both the magnitude and volume of the smaller storm peaks as well as the annual flood event on 8 June 2016.

For the larger Lippe catchment, we again saw improved mean performance for the run with inversely generated precipitation (Fig. 6c). For the winter flood of 2011 (Fig. 6a), the HBV model, which used inversely generated precipitation, better matched the observed streamflow dynamics, whereas the ERA5 Land run exhibited significant overestimation errors. The inversely generated precipitation estimates again improved HBV model performance for replicating the dis-

charge dynamics during the floods in December 2012 and February 2013 (Fig. 6b).

To understand the evolution of soil moisture dynamics while using the *with_discharge* LSTM-based precipitation estimates in physically based models, we conducted a hillslope-scale CATFLOW model simulation (Loritz et al., 2017; Manoj J et al., 2024) in one of the headwater catchments in Elsenz Schwarzbach (ERA5 Land vs. *with_discharge* LSTM). The pairwise correlation values, as shown in Fig. 7, indicate that the use of the LSTM-based precipitation estimates does not lead to a loss of information regarding soil moisture dynamics in the region. In fact, we observe a slight increase in correlation when comparing the inversely derived precipitation estimates (referred to as CATFLOW_lstm) to MERRA and GLDAS (Table 1), in contrast with the correlation obtained for the run with ERA5 Land (referred to as CATFLOW_era5). As expected, the correlation value for the ERA5 Land run is slightly higher when assessed against soil moisture from the same ERA5 Land dataset, which may be attributed to model biases arising from

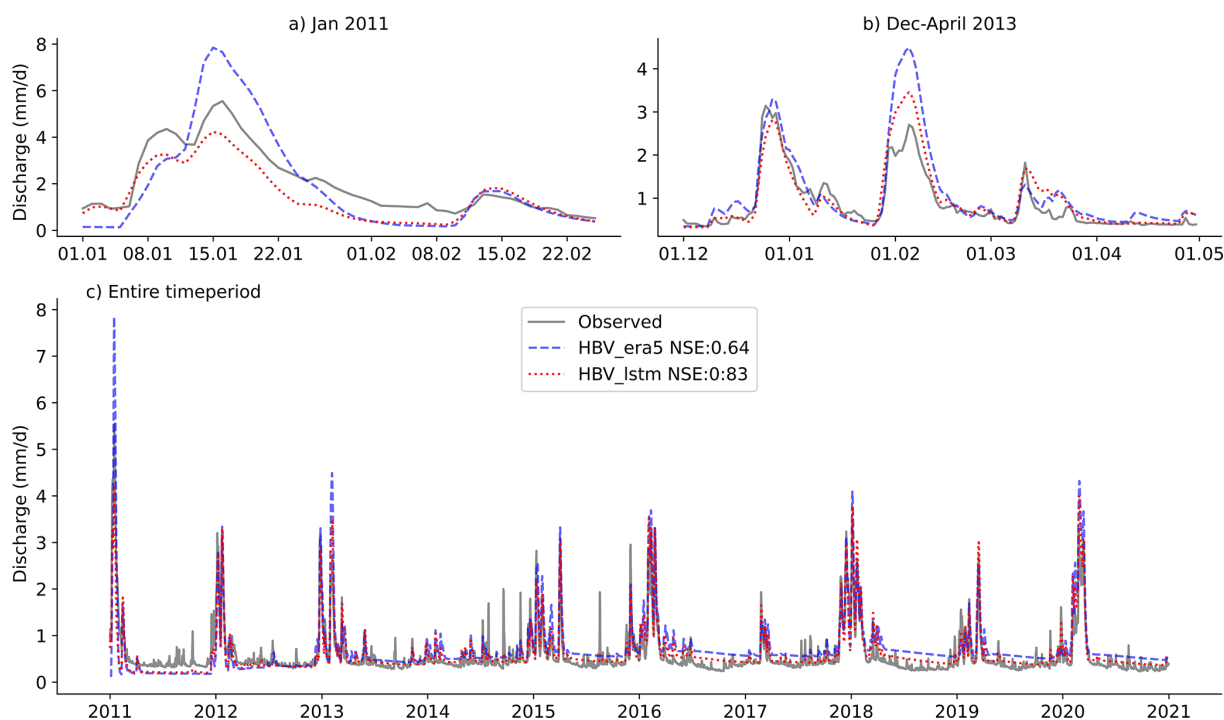


Figure 6. Observed and simulated runoff (using the HBV model) at the Lippe catchment. The blue line denotes the streamflow simulated using the ERA5 Land precipitation product, while the red curve depicts the simulations using the inversely-estimate precipitation obtained using the *with_discharge* LSTM model. Moreover, two rainfall–runoff events are highlighted and displayed separately.

using the same dataset for both precipitation and soil moisture.

4 Discussion

4.1 Improved precipitation estimation using discharge

Overall, our study reiterates that streamflow data can be exploited to obtain useful information about the nature of catchment-scale precipitation amounts: we can thus invert the cause using the effect as input to an LSTM. This is in line with, and steps beyond, previous studies (Brocca et al., 2013; Kirchner, 2009; Kretschmar et al., 2014; Krier et al., 2012; Teuling et al., 2010) that explored the possibility of doing hydrology backwards using experimental catchments. Here, we successfully expanded this idea to large samples, cutting across the wide range of hydro-climatic conditions that characterise Europe. We found a largely “normal” distribution of performance, with a few outliers, the latter indicating possible poor quality of discharge data.

Although ERA5 Land precipitation has known uncertainties, it provides continuous global spatial and temporal coverage, making it a useful training dataset. Our goal was not to generate a fully independent dataset but to improve the ERA5 Land precipitation estimates using the additional streamflow information. Reanalysis data, by definition, are a mix of observations and past short-range weather forecasts rerun with

modern weather forecasting models. Different data assimilation methods are then employed (Li et al., 2019). The inversion technique could be used as another final layer of post-processing (using the LSTM in this case) for the model outputs to ensure that the final product is more consistent with the variabilities observed in the discharge record.

One limitation of our approach is that the LSTM model tends to underestimate the timeseries measures (MWD and R95P) at the continental scale. The LSTM’s architecture is known to have a theoretical saturation limit, leading to the underestimation of some of the peak storm events. This so called “saturation problem” (Baste et al., 2025; Chen and Chang, 1996) implies that irrespective of the input series, the predicted values can never exceed a theoretical limit (which is established during the training phase). Furthermore, the LSTM model looks for recurrence in patterns and mean conditions. This means that it can indeed account for consistent baseflow dynamics (as also indicated by analysis over the larger Lippe catchment, Fig. 6). In extreme floods (Merz et al., 2021), the relative contributions of each component can vary significantly, depending on various factors such as the antecedent conditions of the catchment area. The model likely struggles to learn this variability while attempting to invert and obtain the driving precipitation values. Given the non-linear nature of the inverse problem, there are always multiple possible solutions. Since the model is trained to minimize the mean squared error (Gupta et al., 2009), it may

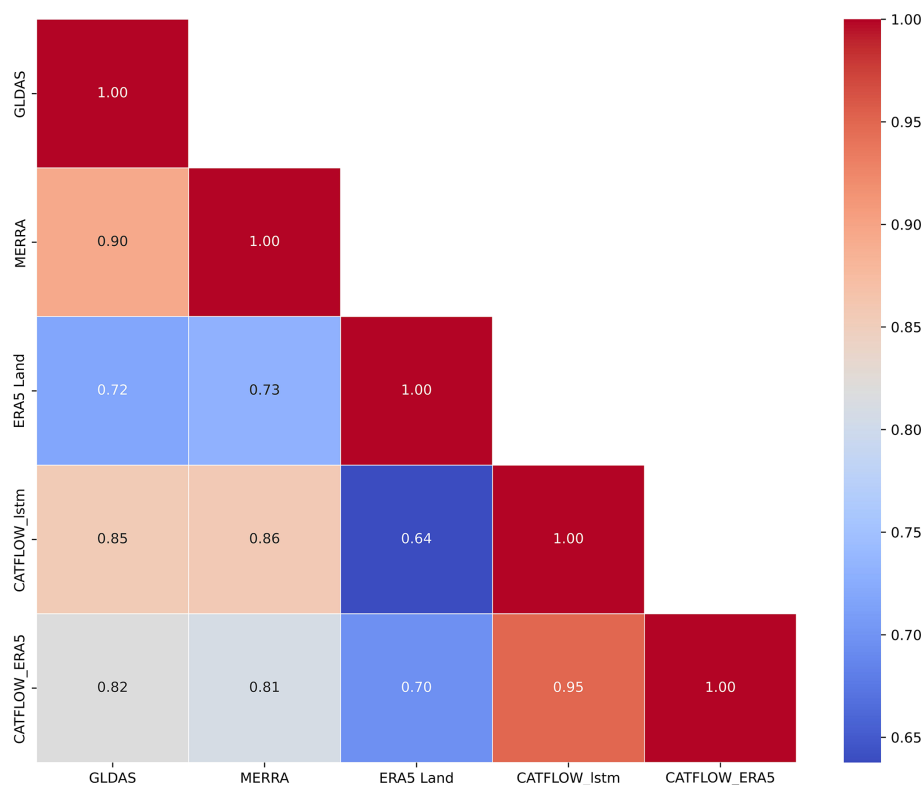


Figure 7. Correlation matrix plot illustrating the pairwise correlations between the different soil moisture estimates – GLDAS (NASA Global Land Data Assimilation System, GLDAS-2.2 GRACE DA: Li et al. (2019), MERRA (Modern-Era Retrospective analysis for Research and Applications version 2 – tavg1_2d_lnd_Nx: Global Modeling And Assimilation Office (2015), ERA5 Land: Muñoz-Sabater et al. (2021), CATFLOW_1stm: model run using inversely estimated precipitation estimate from the LSTM model and CATFLOW_ERA5: model run using precipitation estimate from ERA5 Land product.

also tend to consistently predict lower values (on peaks) to effectively reduce the average error during training.

It is also important to acknowledge that “true” precipitation estimates don’t exist at the catchment scale. We obtain estimates of forcing precipitation at such scales (with considerable uncertainty) by interpolating station data (e.g. EOBS) or averaging gridded data from reanalysis/remote sensing products (e.g. ERA5 Land).

In our out-of-sample simulations, we observed that the LSTM model, which included additional discharge information, overestimated the peak values reported by the observational product used for training. While such an overestimation is typically considered an artifact of imperfect model training and viewed as statistical white noise, we believe that the consistent overestimation of peaks in three out of the four catchments suggests that the LSTM model, trained globally on larger catchments with smaller observational uncertainties, is capable of learning the rainfall–runoff relationship and can adjust for observational errors at the out-of-sample sites. Although the model was not specifically trained on the timing characteristics of hydrological events, we found that it can still produce hydrologically consistent estimates for the time to peak for storms.

The performance comparison using the runoff coefficients was intended to provide insight into the feasibility of different precipitation estimates from a hydrological perspective. While we acknowledge the existence of even better regional products (e.g., HYRAS – German Weather Service) for some of the study catchments compared to the continental scale EOBS, we believe that these various products should not be viewed as independent of one another. Instead, they contain complementary information as they represent the same physical truth i.e. precipitation occurring over a catchment, albeit with different uncertainties and errors. Although the EOBS data is only available over Europe, the trained model could be transferred to similar hydroclimatic regions worldwide that have discharge information to correct the globally available ERA5 Land product.

4.2 Catchment as a functional unit

In the introduction, we argued that the catchment scale is crucial for improving our understanding of the factors that drive the water cycle and representing them more accurately in reanalysis products. Our findings across the four catchments highlight the benefit of using streamflow variations to rectify

precipitation estimates. By leveraging the generalisation capabilities of the data-driven LSTM model, we successfully transferred knowledge across different scales (Notably, only about 9 % of the catchments in our training dataset had areas smaller than 100 km²), indicating important implications for addressing the ever-evolving challenge of predictions in ungauged basins (PUB): Hrachowitz et al. (2013).

Although this approach can only be applied after the event has taken place, it has implications for generating coherent long-term statistical records for catchment forcings, which could be used for the design of small- to medium-purpose water resource projects. Employing daily precipitation sums from products like ERA5 Land and EOBS should ideally be a last resort for reproducing small-scale hydrological events, however, the scarcity of real-world data and the rarity of these events may sometimes necessitate a modelling decision to incorporate these coarser estimates. Using the streamflow fluctuations, it would be possible to identify localised rainfall cells or snowfall events that are poorly captured by traditional rain gauges (Kretzschmar et al., 2014). The approach also has potential for evaluating long-term rainfall estimates from Global Circulation Models for specific catchments using information about hydrological conditions (Fujihara et al., 2008).

While the LSTM-based precipitation estimates improved the representation of most events, there were still instances where the original ERA5 Land provided better accuracy for peak flood magnitudes (Fig. 5); this highlights the need for a blended approach that incorporates additional information rather than completely replacing one product with another. In regions around the world, the wealth of streamflow information remains underutilised in this aspect. For Germany alone (Loritz et al., 2024a), there are more than 1500 streamflow gauges, which represent a significantly higher representative area compared to precipitation stations.

The forward exercise using the HBV model showed that the precipitation estimates after inversion enhanced mean performance for streamflow simulation and helped improve the modelling of extreme individual floods. The ability to match the hydrograph differed between the different seasons. Compared to the storage-controlled winter floods (Dunne, 1978), summer floods in these regions are usually driven by Hortonian flow (Horton, 1932) in response to high-intensity rainfall during convective storms. Previous studies (Kirchner, 2009; Krier et al., 2012) have discussed such storage-controlled dynamics and their impact on the inversion problem.

Previous experiences at the event scale (Beauchamp et al., 2013; Zehe and Blöschl, 2004) have also shown that inferring the antecedent soil moisture conditions remains a key challenge for accurate and reliable flood simulations. By utilising the process-based CATFLOW model for soil moisture simulations in a small headwater catchment, we achieved high correlation values using the inverse precipitation estimate. This suggests that the approach can help represent the catch-

ment's overall water dynamics and has the potential for reliable flood design estimations at the event scale, particularly in data-scarce regions.

4.3 Limitations and Outlook

It is important to stress that, as for any data-driven study, the results of our work are contingent on the quality of the training dataset. While we are aware of better regional products for individual countries, ERA5 Land provides consistent global coverage, and a permissive data sharing policy makes it one of the obvious choices for a continental scale modelling exercise. To evaluate the applicability of the commonly used LSTM network architecture, we decided to use the same architecture previously employed in hydrological studies instead of creating an experimental design with modified individual layers and training functions for inverse modelling. It is evident that exploring the impacts of different loss functions and deep learning model architectures like transformers would help advance the methodology discussed in this paper. This approach could also shed light on best-suited algorithms for the problem but is beyond the scope of the present work. The choice of Mean Squared Error (MSE) as the training function and Nash Sutcliffe Efficiency (NSE) as a performance metric is motivated by its success and applications in the forward problem (streamflow prediction), but this adds its own biases to the modelling exercise. In the present work, we tried to overcome this issue by relying less on the evaluation measure (NSE) and placing greater emphasis on the hydrological feasibility of the predictions (using the runoff coefficient). Additionally, we tried to complement this by calculating various other time series metrics commonly used in hydrometeorological studies. The four events for out-of-sample tests across various catchments were chosen based on the severity of the floods and ERA5 Land's inability to capture the characteristics of the driving storms. The choice of the hydrological models and calibration period also adds uncertainty to the forward simulations.

Our approach opens up many perspectives for future research. Transfer learning to data-scarce regions could help address the challenge of highly uncertain precipitation estimates in smaller catchments without precipitation gauges, improving hydrological modeling and the representation of extreme events such as convective storms, which are crucial for designing flood defense measures. Additionally, the inversion technique could serve as a final post-processing layer for gridded reanalysis products, ensuring better consistency with discharge variability and enabling machine learning approaches to estimate spatial precipitation fields conditioned on discharge data (Bárdossy et al., 2020, 2022). Moreover, this methodology could be applied to reconstruct past floods by leveraging historical hydrological records, storm water level markings, and observational flood data (Bronstert et al., 2018; Seidel et al., 2009), providing valuable insights into the driving storms behind some of the devastating past flood

events. The workflow could also be expanded for the generation of new precipitation products, merging multiple different precipitation sources alongside the streamflow inversion.

5 Conclusions

Our main hypothesis was supported by the findings, which demonstrated that discharge has unused potential and can be inversely assimilated to adjust precipitation estimates derived from reanalysis products, while machine learning models are key to expanding this effort to large data sets spanning the scale of entire continents. As expected, the performance gain in using discharge information was significantly higher for days with increasing precipitation amounts. Insights from the out-of-sample catchments provided valuable information about the applicability of our method for estimating flood forcings and the generalizability of the model. Additionally, we have shown that the inversely estimated precipitation estimates can improve forward modelling of both streamflow and soil moisture dynamics, illustrating how the information gained can be integrated into existing modelling strategies.

Appendix A: LSTM configurations

Table A1 details the static and dynamic inputs used for setting up the *with_discharge* and *without_discharge* LSTM ensemble models. The hyperparameter settings for both models are shown in Table A2, while Fig. A1 provides the comparison results for both runs.

Table A1. Model configurations for the LSTM model runs.

Model	Inputs		Output
	Static Attributes	Dynamic Attributes	
<i>with_discharge</i>	<i>area</i> (area of catchment – km ²) <i>ele_mt_sav</i> (spatial mean elevation – m above sea level) <i>frac_snow</i> (fraction of precipitation falling as snow) <i>pet_mm_syr</i> (potential evapotranspiration annual mean – mm)	<i>total_precipitation_sum_era5</i> (precipitation daily sums – mm d ^{–1}) <i>temperature_2m_mean</i> (daily mean temperature – °C) <i>surface_net_solar_radiation_mean</i> (shortwave radiation – Wm ^{–2}) <i>surface_net_thermal_radiation_mean</i> (Net thermal radiation at the surface – Wm ^{–2}) <i>qobs_lead</i> (lead streamflow 7 d – mm d ^{–1})	<i>eobs_precipitation</i> (precipitation daily sums – mm d ^{–1})
<i>without_discharge</i>	<i>area</i> (area of catchment –km ²) <i>ele_mt_sav</i> (spatial mean elevation – m above sea level) <i>frac_snow</i> (fraction of precipitation falling as snow) <i>pet_mm_syr</i> (potential evapotranspiration annual mean – mm)	<i>total_precipitation_sum_era5</i> (precipitation daily sums – mm d ^{–1}) <i>temperature_2m_mean</i> (daily mean temperature – °C) <i>surface_net_solar_radiation_mean</i> (shortwave radiation — Wm ^{–2}) <i>surface_net_thermal_radiation_mean</i> (Net thermal radiation at the surface – Wm ^{–2})	<i>eobs_precipitation</i> (precipitation daily sums – mm d ^{–1})

Table A2. Hyperparameter settings for the LSTM models.

Hyperparameter	LSTM Network
Hidden Layer	1
Hidden cells	64
Batch size	256
Sequence length	365
Epochs	5
Drop out	0.4
Learning rate	0.001
Optimizer	Adam

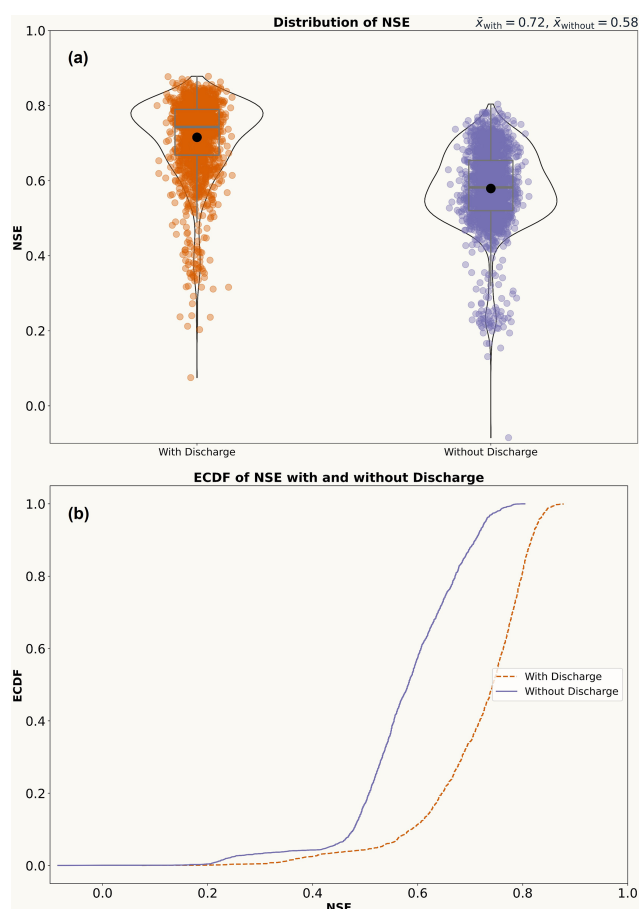


Figure A1. Comparison of the mean performance of the two regional scale LSTM models (*with_discharge* and *without_discharge*). **(a)** Top panel depicts violin plots with included boxplots showing the distribution of performance (quantified by comparing the LSTM model simulated precipitation series to the observational EOBS timeseries over the testing period: NSE) **(b)** Bottom panel displays cumulative distribution plots for the performance of the two models.

Appendix B: Hydrological Modelling

Hydrologiska Byråns Vattenbalansavdelning (HBV). The HBV model (Bergström and Forsman, 1973) is a so-called conceptual hydrological model that is used to simulate rainfall–runoff processes at the catchment scale. It makes use of different catchment water stores (storage elements, also referred to as buckets). Each storage element represents a certain compartment of a catchment (e.g. groundwater, surface water bodies, soil zone). The main input requirements include precipitation, temperature and potential evapotranspiration. The model has several empirical parameters that need to be calibrated during the model training phase. A more detailed description of the model architecture and set up can be found in the studies by Seibert (2005) and Loritz et al. (2024a).

CATFLOW. The physically based model CATFLOW for catchment water and solute dynamics was developed as part of the detailed process studies carried out from 1991–1996 in the Weiherbach catchment in South-West Germany (Zehe et al., 2001). The basic modeling unit is a 2D hillslope, discretized by curvilinear orthogonal coordinates in the vertical and downslope directions. Soil water dynamics within the hillslopes are characterized using the potential based form of the 2D Darcy–Richards equation. Overland flow is simulated using the diffusion wave approximation of the Saint-Venant equation and explicit upstreaming, in combination with the Gauckler–Manning–Strickler formula. A detailed model description with the workflow required for setting up the model can be found in Manoj J et al. (2024).

Table B1. Validation test cases for the hydrological models.

Model	HBV (Bergström and Forsman, 1973)		CATFLOW (Zehe et al., 2001)	
Catchment (Fig. S5)	Elsenz Schwarzbach (Manoj J et al., 2024)	Lippe (Loritz et al., 2024a)	Headwater catchment W32 in Elsenz Schwarzbach (Manoj J et al., 2024)	
Area (km ²)	196.5	3366.3	5.6	
Scenario	Using original ERA5 Land precipitation	Using original ERA5 Land precipitation	Using original ERA5 Land precipitation	Using inversely generated precipitation (with discharge)
Forcings	ERA5 Land precipitation, potential evapotranspiration (Hargreaves), air temperature (ERA5 Land)	ERA5 Land precipitation, potential evapotranspiration (Hargreaves), air temperature (ERA5 Land)	ERA5 Land precipitation, potential evapotranspiration (Hargreaves), air temperature (ERA5 Land)	ERA5 Land precipitation, potential evapotranspiration (ERA5 Land), plant and soil parameters from Manoj J et al. (2024)
Calibration period	1 Jan 2000–31 Dec 2010	1 Jan 1990–31 Dec 2010	Uncalibrated predictions 1 Jan 2008–31 Dec 2015	
Validation period	1 Jan 2011–31 Dec 2016	1 Jan 2011–31 Dec 2020	Soil moisture	
Outputs compared	Streamflow	Streamflow	0.64–0.86	
NSE	0.57	0.64	0.70–0.82	

Appendix C: Performance Metrics

Nash–Sutcliffe Efficiency (NSE). First proposed by Nash and Sutcliffe (1970), the Nash–Sutcliffe efficiency (NSE) is one of the most widely used similarity measures in hydrology for calibration, model comparison, and verification. It measures how well the simulated timeseries (y_{sim}) matches the observed values (y_{obs}).

$$\text{NSE} = 1 - \frac{\sum (y_{\text{obs}} - y_{\text{sim}})^2}{\sum (y_{\text{obs}} - \bar{y}_{\text{obs}})^2} \quad (\text{C1})$$

Values closer to 1 indicate excellent model performance (Moriassi et al., 2007), while NSE values near or below 0 suggest that the model, in fact, performs worse than simply using the mean of the observed values.

Mean Wet Day Precipitation (MWD) (mm d^{-1}). The Expert Team on Climate Change Detection and Indices (ETCCDI – World Climate Research program; 2021) recommends evaluating the intensity of precipitation on wet days (defined as a day with a minimum of 1 mm precipitation) to understand systematic over or underestimation of precipitation amounts. This metric (Simple Daily Intensity Index as per ETCCDI) is reported as the mean daily precipitation on days where precipitation > 1 mm. Let P_i be the daily precipitation amount on wet days, ($P_i > 1$ mm). If N represents the total number of wet days, then:

$$\text{MWD} = \frac{\sum_{i=1}^N P_i}{N} \quad (\text{C2})$$

95th Percentile Precipitation (R95P) (mm d^{-1}). This metric denotes the daily precipitation value at which 95 % of all daily values (again only considering rainy days) are lower (top 5 % events). This helps to assess the ability to capture extreme precipitation events. Let P_i be the daily precipitation amount on wet days, ($P_i > 1$ mm)

$$\text{R95P} = \text{Percentile}(\{P_i | P_i > 1 \text{ mm}\}, 95) \quad (\text{C3})$$

Spearman Rank Autocorrelation (SL). The Spearman Rank Autocorrelation measures the monotonic relationship between daily precipitation values and their values on the preceding day (1 d lag). It is computed using the ranked values of the precipitation time series. For a precipitation time-series (with total n observations) $P = \{P_1, P_2, \dots, P_n\}$ with $R(P_i)$ and $R(P_{i+1})$ being the ranks of the precipitation values at times t and $t + 1$,

$$\text{SL} = 1 - \frac{6 \sum_{t=1}^{n-1} (R(P_{t+1}) - R(P_t))^2}{n(n^2 - 1)} \quad (\text{C4})$$

This measure helps analyse persistence in precipitation patterns and whether the temporal structure of precipitation events are preserved.

Code availability. The codes used to conduct the LSTM analysis in this study are based on the publicly available HY²DL python library (<https://github.com/KIT-HYD/Hy2DL>, last access: 26 November 2024) and can be accessed at <https://doi.org/10.5281/zenodo.14161027> (Manoj J, 2025b). The code used to run the HBV models is available at <https://doi.org/10.5281/zenodo.15051966> (Manoj J, 2025a). The CATFLOW model and the setup used to run the experiment in this study are archived at <https://doi.org/10.5281/zenodo.10958813> (Manoj J, 2024).

Data availability. The Caravan dataset and related community extensions are publicly available at <https://doi.org/10.5281/zenodo.10968468> (Kratzert et al., 2023) and <https://github.com/kratzert/Caravan/discussions/10>, last access: 26 November 2024. We acknowledge the E-OBS dataset from the Copernicus Climate Change Service (C3S, <https://surfobs.climate.copernicus.eu>, last access: 26 November 2024) and the data providers in the ECA&D project (<https://www.ecad.eu>, last access: 26 November 2024). The datasets generated as part of this publication can be found at <https://doi.org/10.5281/zenodo.14161027> (Manoj J, 2025b) and <https://doi.org/10.5281/zenodo.15051966> (Manoj J, 2025a).

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/hess-29-6115-2025-supplement>.

Author contributions. AMJ designed the study and carried out all analysis and model simulations. Funding was acquired by EZ. The initial draft was prepared by AMJ, with all authors contributing to review and editing. RL, HG and EZ jointly supervised the work. All authors have read and agreed to the current version of the paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Hydrology and Earth System Sciences*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The authors acknowledge support by the federal state of Baden-Württemberg through bwHPC (High Performance Computing Cluster). Ashish Manoj J would like to thank Eduardo Acuña Espinoza for helpful discussions regarding the HY²DL python library for deep learning methods and Alexander Dolich for help in implementing the stgrid2area python package.

Financial support. This research has been supported by the Deutsche Forschungsgemeinschaft (German Research Foundation - DFG) via the project - Implementation of an InfraStructure for dAta-BasEd Learning in environmental sciences (ISABEL - grant no. 496155047).

The article processing charges for this open-access publication were covered by the Karlsruhe Institute of Technology (KIT).

Review statement. This paper was edited by Roger Moussa and reviewed by three anonymous referees.

References

- Acuña Espinoza, E., Loritz, R., Álvarez Chaves, M., Bäuerle, N., and Ehret, U.: To bucket or not to bucket? Analyzing the performance and interpretability of hybrid hydrological models with dynamic parameterization, *Hydrol. Earth Syst. Sci.*, 28, 2705–2719, <https://doi.org/10.5194/hess-28-2705-2024>, 2024.
- Adam, J. C., Clark, E. A., Lettenmaier, D. P. and Wood, E. F.: Correction of global precipitation products for orographic effects, *J. Climate*, 19, 15–38, <https://doi.org/10.1175/JCLI3604.1>, 2006.
- Agarwal, A., Marwan, N., Maheswaran, R., Ozturk, U., Kurths, J., and Merz, B.: Optimal design of hydrometric station networks based on complex network analysis, *Hydrol. Earth Syst. Sci.*, 24, 2235–2251, <https://doi.org/10.5194/hess-24-2235-2020>, 2020.
- Alexopoulos, M. J., Müller-Thomy, H., Nistahl, P., Šraj, M., and Bezak, N.: Validation of precipitation reanalysis products for rainfall-runoff modelling in Slovenia, *Hydrol. Earth Syst. Sci.*, 27, 2559–2578, <https://doi.org/10.5194/hess-27-2559-2023>, 2023.
- Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O. E., and Frei, C.: Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions, *Int. J. Climatol.*, 42, 727–747, <https://doi.org/10.1002/joc.7269>, 2022.
- Bárdossy, A., Anwar, F., and Seidel, J.: Hydrological Modelling in Data Sparse Environment: Inverse Modelling of a Historical Flood Event, *Water (Switzerland)*, 12, <https://doi.org/10.3390/w12113242>, 2020.
- Bárdossy, A., Kilsby, C., Birkinshaw, S., Wang, N., and Anwar, F.: Is Precipitation Responsible for the Most Hydrological Model Uncertainty?, *Front. Water*, 4, 1–17, <https://doi.org/10.3389/frwa.2022.836554>, 2022.
- Baste, S., Klotz, D., Espinoza, E. A., Bárdossy, A., and Loritz, R.: Unveiling the Limits of Deep Learning Models in Hydrological Extrapolation Tasks, *Hydrol. Earth Syst. Sci.*, 29, 5871–5891, <https://doi.org/10.5194/hess-29-5871-2025>, 2025.
- Beauchamp, J., Leconte, R., Trudel, M., and Brissette, F.: Estimation of the summer-fall PMP and PMF of a northern watershed under a changed climate, *Water Resour. Res.*, 49, 3852–3862, <https://doi.org/10.1002/wrcr.20336>, 2013.
- Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., and Kirchner, J. W.: The Relative Importance of Different Flood-Generating Mechanisms Across Europe, *Water Resour. Res.*, 55, 4582–4593, <https://doi.org/10.1029/2019WR024841>, 2019.
- Bergström, S. and Forsman, A.: Development of a Conceptual Deterministic Rainfall–Runoff Model., *Nord. Hydrol.*, 4, 147–170, <https://doi.org/10.2166/nh.1973.0012>, 1973.
- Bishop, C. M.: Pattern recognition and machine learning, Springer, New York, ISBN 978-0-387-31073-2, 2006.
- Borga, M., Gaume, E., Creutin, J. D., and Marchi, L.: Surveying flash floods: gauging the ungauged extremes, *Hydrol. Process.*, 22, 3883–3885, <https://doi.org/10.1002/hyp.7111>, 2008.1.
- Brocca, L., Moramarco, T., Melone, F., and Wagner, W.: A new method for rainfall estimation through soil moisture observations, *Geophys. Res. Lett.*, 40, 853–858, <https://doi.org/10.1002/grl.50173>, 2013.
- Bronstert, A., Agarwal, A., Boessenkool, B., Crisologo, I., Fischer, M., Heistermann, M., Köhn-Reich, L., López-Tarazón, J. A., Moran, T., Ozturk, U., Reinhardt-Imjela, C., and Wendi, D.: Forensic hydro-meteorological analysis of an extreme flash flood: The 2016-05-29 event in Braunschweig, SW Germany, *Sci. Total Environ.*, 630, 977–991, <https://doi.org/10.1016/j.scitotenv.2018.02.241>, 2018.
- Casado Rodríguez, J.: CAMELS-ES: Catchment Attributes and Meteorology for Large-Sample Studies – Spain, Zenodo [code], <https://doi.org/10.5281/zenodo.8428374>, 2023.
- Chen, C. T. and Chang, W. Der: A feedforward neural network with function shape autotuning, *Neural Networks*, 9, 627–641, [https://doi.org/10.1016/0893-6080\(96\)00006-8](https://doi.org/10.1016/0893-6080(96)00006-8), 1996.
- Clerc-Schwarzenbach, F. M., Selleri, G., Neri, M., Toth, E., van Meerveld, I., and Seibert, J.: Large-sample hydrology: a few camels or a whole caravan?, *Hydrology and Earth System Sciences*, 28, 4219–4237, <https://doi.org/10.5194/hess-28-4219-2024>, 2024.
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, *J. Geophys. Res. Atmos.*, 123, 9391–9409, <https://doi.org/10.1029/2017JD028200>, 2018.
- Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hanaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, *Earth Syst. Sci. Data*, 12, 2459–2483, <https://doi.org/10.5194/essd-12-2459-2020>, 2020.
- Dolich, A., Maharjan, A., Mälicke, M., Manoj J, A., and Loritz, R.: Caravan-DE: Caravan extension Germany – German dataset for large-sample hydrology, Zenodo [code], <https://doi.org/10.5281/zenodo.14755229>, 2025.
- Dunne, T.: Field studies of hillslope flow processes, in: *Hillslope Hydrology*, edited by: Kirkby, M. J., John Wiley & Sons, 227–293, ISBN 978-0-471-99510-4, 1978.
- Essou, G. R. C., Sabarly, F., Lucas-Picher, P., Brissette, F., and Poulin, A.: Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling?, *J. Hydrometeorol.*, 17, 1929–1950, <https://doi.org/10.1175/JHM-D-15-0138.1>, 2016.
- Färber, C., Plessow, H., Kratzert, F., Addor, N., Shalev, G., and Looser, U.: GRDC-Caravan: extending the original dataset with data from the Global Runoff Data Centre, Zenodo [code], <https://doi.org/10.5281/zenodo.10074416>, 2023.

- Fujihara, Y., Simonovic, S. P., Topaloglu, F., Tanaka, K., and Watanabe, T.: An inverse-modelling approach to assess the impacts of climate change in the Seyhan River basin, Turkey, *Hydrol. Sci. J.*, 53, 1121–1136, <https://doi.org/10.1623/hysj.53.6.1121>, 2008.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *J. Clim.*, 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- Global Modeling And Assimilation Office: MERRA-2 tavg1_2d_lnd_nx: 2d, 1-Hourly, Time-Averaged, Single-Level, Assimilation, Land Surface Diagnostics V5.12.4, <https://doi.org/10.5067/RKPH8K1Y1T>, 2015.
- Goteti, G. and Famiglietti, J.: Extent of gross underestimation of precipitation in India, *Hydrol. Earth Syst. Sci.*, 28, 3435–3455, <https://doi.org/10.5194/hess-28-3435-2024>, 2024.
- Gu, L., Yin, J., Wang, S., Chen, J., Qin, H., Yan, X., He, S., and Zhao, T.: How well do the multi-satellite and atmospheric reanalysis products perform in hydrological modelling, *J. Hydrol.*, 617, 128920, <https://doi.org/10.1016/j.jhydrol.2022.128920>, 2023.
- 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.
- Hochreiter, S.: The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions, *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, 06, 107–116, <https://doi.org/10.1142/S0218488598000094>, 1998.
- Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: Catchment attributes and hydro-meteorological time series for large-sample studies across hydrologic Switzerland, *Earth Syst. Sci. Data*, 15, 5755–5784, <https://doi.org/10.5194/essd-15-5755-2023>, 2023.
- Horton, R. E.: The role of infiltration in the hydrology cycle, *Eos Trans. AGU*, 14, 446–460, <https://doi.org/10.1029/TR014i001p00446>, 1932.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)–a review, *Hydrol. Sci. J.*, 58, 1198–1255, <https://doi.org/10.1080/02626667.2013.803183>, 2013.
- Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall–runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, 1–34, <https://doi.org/10.1029/2008WR006912>, 2009.
- Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., and Nearing, G.: Uncertainty estimation with deep learning for rainfall–runoff modeling, *Hydrol. Earth Syst. Sci.*, 26, 1673–1693, <https://doi.org/10.5194/hess-26-1673-2022>, 2022.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Hernegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrol. Earth Syst. Sci.*, 22, 6005–6022, <https://doi.org/10.5194/hess-22-6005-2018>, 2018.
- Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan – A global community dataset for large-sample hydrology, *Sci. Data*, 10, 61, <https://doi.org/10.1038/s41597-023-01975-w>, 2023.
- Kretzschmar, A., Tych, W., and Chappell, N. A.: Reversing hydrology: Estimation of sub-hourly rainfall time-series from streamflow, *Environ. Model. Softw.*, 60, 290–301, <https://doi.org/10.1016/j.envsoft.2014.06.017>, 2014.
- Krier, R., Matgen, P., Goergen, K., Pfister, L., Hoffmann, L., Kirchner, J. W., Uhlenbrook, S., and Savenije, H. H. G.: Inferring catchment precipitation by doing hydrology backward: A test in 24 small and mesoscale catchments in Luxembourg, *Water Resour. Res.*, 48, 1–15, <https://doi.org/10.1029/2011WR010657>, 2012.
- Li, B., Rodell, M., Kumar, S., Beaudoin, H. K., Getirana, A., Zaitchik, B. F., de Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I. B., Daira, D., Bila, M., de Lannoy, G., Mocko, D., Steele-Dunne, S. C., Save, H., and Bettadpur, S.: Global GRACE Data Assimilation for Groundwater and Drought Monitoring: Advances and Challenges, *Water Resour. Res.*, 55, 7564–7586, <https://doi.org/10.1029/2018WR024618>, 2019.
- Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., and Thieme, M.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, *Sci. Data*, 6, 283, <https://doi.org/10.1038/s41597-019-0300-6>, 2019.
- Loritz, R., Hassler, S. K., Jackisch, C., Allroggen, N., van Schaik, L., Wienhöfer, J., and Zehe, E.: Picturing and modeling catchments by representative hillslopes, *Hydrol. Earth Syst. Sci.*, 21, 1225–1249, <https://doi.org/10.5194/hess-21-1225-2017>, 2017.
- Loritz, R., Dolich, A., Acuña Espinoza, E., Ebeling, P., Guse, B., Götte, J., Hassler, S. K., Hauffe, C., Heidbüchel, I., Kiesel, J., Mälicke, M., Müller-Thomy, H., Stölzle, M., and Tarasova, L.: CAMELS-DE: hydro-meteorological time series and attributes for 1582 catchments in Germany, *Earth Syst. Sci. Data*, 16, 5625–5642, <https://doi.org/10.5194/essd-16-5625-2024>, 2024a.
- Loritz, R., Wu, C. H., Klotz, D., Gauch, M., Kratzert, F., and Bassiouni, M.: Generalizing Tree-Level Sap Flow Across the European Continent, *Geophys. Res. Lett.*, 51, <https://doi.org/10.1029/2023GL107350>, 2024b.
- Manoj J, A.: Simulation results of Manoj J et al. (2023), Zenodo [code], <https://doi.org/10.5281/zenodo.10958813>, 2024.
- Manoj J, A.: Ash-Manoj/Hy2DL_Caravan: Conceptual models for Manoj J et al. (2024), Zenodo [code], <https://doi.org/10.5281/zenodo.15051966>, 2025a.
- Manoj J, A.: Ash-Manoj/lstm_backward: LSTM models for Manoj J et al. (2024), Zenodo [code], <https://doi.org/10.5281/zenodo.14161027>, 2025b.

- Manoj J, A., Pérez Ciria, T., Chiogna, G., Salzmänn, N., and Agarwal, A.: Characterising the coincidence of soil moisture – precipitation extremes as a possible precursor to European floods, *J. Hydrol.*, 620, 129445, <https://doi.org/10.1016/j.jhydrol.2023.129445>, 2023.
- Manoj J, A., Loritz, R., Villinger, F., Mälicke, M., Koopaeidar, M., Göppert, H., and Zehe, E.: Toward Flash Flood Modeling Using Gradient Resolving Representative Hillslopes, *Water Resour. Res.*, 60, <https://doi.org/10.1029/2023WR036420>, 2024.
- Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U., and Macdonald, E.: Causes, impacts and patterns of disastrous river floods, *Nat. Rev. Earth Environ.*, 2, 592–609, <https://doi.org/10.1038/s43017-021-00195-3>, 2021.
- Meyer, J., Neuper, M., Mathias, L., Zehe, E., and Pfister, L.: Atmospheric conditions favouring extreme precipitation and flash floods in temperate regions of Europe, *Hydrol. Earth Syst. Sci.*, 26, 6163–6183, <https://doi.org/10.5194/hess-26-6163-2022>, 2022.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Climate change: Stationarity is dead: Whither water management?, *Science*, 319, 573–574, <https://doi.org/10.1126/science.1151915>, 2008.
- Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaeffli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: “Panta Rhei-Everything Flows”: Change in hydrology and society-The IAHS Scientific Decade 2013–2022, *Hydrol. Sci. J.*, 58, 1256–1275, <https://doi.org/10.1080/02626667.2013.809088>, 2013.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, *Trans. ASABE*, 50, 885–900, <https://doi.org/10.13031/2013.23153>, 2007.
- 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.*, 27, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Nearing, G. S. and Gupta, H. V.: The quantity and quality of information in hydrologic models, *Water Resour. Res.*, 51, 524–538, <https://doi.org/10.1002/2014WR015895>, 2015.
- Nijzink, J., Loritz, R., Gourdol, L., Zoccatelli, D., Iffly, J. F., and Pfister, L.: CAMELS-LUX: Highly Resolved Hydro-Meteorological and Atmospheric Data for Physiographically Characterized Catchments around Luxembourg, Zenodo [code], <https://doi.org/10.5281/zenodo.13846620>, 2024.
- Ongie, G., Jalal, A., Metzler, C. A., Baraniuk, R. G., Dimakis, A. G., and Willett, R.: Deep Learning Techniques for Inverse Problems in Imaging, *IEEE J. Sel. Areas Inf. Theory*, 1, 39–56, <https://doi.org/10.1109/jsait.2020.2991563>, 2020.
- Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R.: The JRA-25 Reanalysis, *J. Meteorol. Soc. Japan. Ser. II*, 85, 369–432, <https://doi.org/10.2151/jmsj.85.369>, 2007.
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications, *J. Clim.*, 24, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>, 2011.
- Seibert, J.: HBV light, HBV Light version 2 User’s Man., https://www.geo.uzh.ch/dam/jcr:c8afa73c-ac90-478e-a8c7-929eed7b1b62/HBV_manual_2005.pdf (last access: 25 November 2024), 2005.
- Seidel, J., Imbery, F., Dostal, P., Sudhaus, D., and Bürger, K.: Potential of historical meteorological and hydrological data for the reconstruction of historical flood events – the example of the 1882 flood in southwest Germany, *Nat. Hazards Earth Syst. Sci.*, 9, 175–183, <https://doi.org/10.5194/nhess-9-175-2009>, 2009.
- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, *Earth-Science Rev.*, 99, 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004>, 2010.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O’Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, 48, 857–880, <https://doi.org/10.1623/hysj.48.6.857.51421>, 2003.
- Sun, S. and Bertrand-Krajewski, J. L.: Separately accounting for uncertainties in rainfall and runoff: Calibration of event-based conceptual hydrological models in small urban catchments using Bayesian method, *Water Resour. Res.*, 49, 5381–5394, <https://doi.org/10.1002/wrcr.20444>, 2013.
- Tarek, M., Brissette, F. P., and Arsenaault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, *Hydrol. Earth Syst. Sci.*, 24, 2527–2544, <https://doi.org/10.5194/hess-24-2527-2020>, 2020.
- Taszarek, M., Allen, J. T., Marchio, M., and Brooks, H. E.: Global climatology and trends in convective environments from ERA5 and rawinsonde data, *npj Clim. Atmos. Sci.*, 4, 1–11, <https://doi.org/10.1038/s41612-021-00190-x>, 2021.
- Tetzlaff, D., Carey, S. K., McNamara, J. P., Laudon, H., and Soulsby, C.: The essential value of long-term experimental data for hydrology and water management, *Water Resour. Res.*, 53, 2598–2604, <https://doi.org/10.1002/2017WR020838>, 2017.
- Teuling, A. J., Lehner, I., Kirchner, J. W., and Seneviratne, S. I.: Catchments as simple dynamical systems: Experience from

- a Swiss prealpine catchment, *Water Resour. Res.*, 46, 1–15, <https://doi.org/10.1029/2009WR008777>, 2010.
- Villinger, F., Loritz, R., and Zehe, E.: Torrents in small rural Catchments and the Potential of physics-based Models for their Simulation, *Hydrol. und Wasserbewirtschaftung*, 66, 284–285, https://doi.org/10.5675/HyWa_2022.6_1, 2022.
- World Climate Research program (WCRP): Expert Team on Climate Change Detection and Indices (ETCCDI), <https://www.wcrp-climate.org/etccdi> (last access: 7 March 2025), 2021.
- Xu, C., Wang, W., Hu, Y., and Liu, Y.: Evaluation of ERA5, ERA5-Land, GLDAS-2.1, and GLEAM potential evapotranspiration data over mainland China, *J. Hydrol. Reg. Stud.*, 51, 101651, <https://doi.org/10.1016/j.ejrh.2023.101651>, 2024.
- Yumnam, K., Kumar Guntu, R., Rathinasamy, M., and Agarwal, A.: Quantile-based Bayesian Model Averaging approach towards merging of precipitation products, *J. Hydrol.*, 604, 127206, <https://doi.org/10.1016/j.jhydrol.2021.127206>, 2022.
- Zehe, E. and Blöschl, G.: Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, *Water Resour. Res.*, 40, <https://doi.org/10.1029/2003WR002869>, 2004.
- Zehe, E., Maurer, T., Ihringer, J., and Plate, E.: Modelling water flow and mass transport in a Loess catchment, *Phys. Chem. Earth, Part B*, 26, 487–507, [https://doi.org/10.1016/S0378-3774\(99\)00083-9](https://doi.org/10.1016/S0378-3774(99)00083-9), 2001.