https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Learning from dense meteorological networks: a regional LSTM model for streamflow simulation

Inmaculada C. González-Planet[1], Daniel Klotz[2], Carmelo Juez[1]

[1]Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Zaragoza, Spain.
[2]Interdisciplinary Transformation University Austria, Linz, Austria.

Correspondence to: Inmaculada C. González-Planet (ic.planet@ipe.csic.es)

Abstract. Long Short-Term Memory (LSTM) networks have become the current state-of-the-art for streamflow modelling due to their ability to capture non-linear relationships and long-term temporal dependencies. In regional LSTM applications, meteorological forcings are typically aggregated at the catchment scale, which removes spatial variability and may limit the representation of key hydrological processes, particularly in large and heterogeneous basins. This study departs from this conventional practice by proposing a regional LSTM model that directly exploits the full spatial structure of raw meteorological observations without any prior spatial aggregation. The model integrates dynamic and static information from more than 700 meteorological stations across the northern Ebro River Basin (Spain) to simulate streamflow at 31 catchments. Missing data are handled using a masked mean approach, enabling direct use of observations. Compared to a benchmark regional LSTM based on catchment-averaged forcings, the proposed approach improves overall performance, and yields substantial reductions in error for medium- and high-flow conditions. These results indicate that preserving spatial meteorological variability enhances the representation of hydrological processes associated with runoff generation and peak-flow response. To further investigate the learned relationships, explainable artificial intelligence (XAI) techniques are applied. The analyses reveal physically consistent spatial and temporal attribution patterns, showing that the model dynamically adjusts the relevance of meteorological stations depending on hydrological conditions. This study demonstrates that spatial aggregation of meteorological inputs can limit their ability to capture hydrological dynamics in complex basins. The proposed framework provides a flexible and data-driven approach for exploiting dense observational networks in regional streamflow modelling.

1 Introduction

Physically-based models have progressively become more realistic over the past decades due to technological advances and the continuous expansion of hydrometeorological observations at a global scale. These advances include not only improvements in observation systems, from early satellite missions to high-frequency measurements supported by radar and lidar techniques(McCabe et al., 2017), but also significant progress in computational capabilities. In particular, the widespread availability of high-performance computing and graphics processing units (GPUs) has enabled the simulation of increasingly complex hydrological processes at finer spatial and temporal resolutions (Juez et al., 2016). At the local scale, hydrological monitoring networks, including rain gauges and streamflow gauges, have also gained importance and continue to expand (Holzmann, 2018).

Despite the increasing availability of hydrological data, physically-based models still present several limitations, including high computational costs, particularly in a high definition of complex and heterogeneous catchments (Beven, 2012). Additionally, climate variability and change can alter entire flow regimes in ways that are difficult to represent through predefined process-based formulations (Bourdin et al., 2012; Patterson et al., 2022). These limitations, together with the increasing availability of hydrometeorological observations and the need to represent highly nonlinear and potentially non-stationary catchment responses, have encouraged the exploration of complementary data-driven modelling approaches.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Artificial neural networks (ANNs) have gained increasing attention since their introduction to hydrology (French et al., 1992), due to their ability to represent highly nonlinear relationships and complex system dynamics. The inherently non-stationary nature of hydrological processes (Juez et al., 2022, 2023) motivated the use of recurrent neural networks (RNNs), which are specifically designed to capture temporal dependencies in sequential data. However, classical RNNs are limited by vanishing and exploding gradient problems, which hinder their ability to learn long-term dependencies (Hochreiter, 1998).

Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber (1997), were developed to address these limitations and have become one of the most widely used deep learning architectures for streamflow modelling. LSTMs retains information over long time series thanks to its memory cells and gating mechanisms. Their ability to capture long-term temporal dependencies has led to superior performance compared to traditional RNNs and, in some cases, physically based models Kratzert et al. (2018) were the first to apply LSTMs for runoff simulations using meteorological observations. Their study demonstrated that LSTMs consistently outperformed traditional RNNs and, at several gauging stations, achieve better predictive performance than the benchmark physically-based model. A major jump in the applicability of LSTMs to streamflow prediction was, however, the transition from single-basin models to large-sample and regional modelling approaches, improving generalization and robustness (Kratzert et al., 2024; Ma et al., 2021).

As of today, most regional LSTM applications rely on catchment-averaged meteorological forcings, which inherently remove spatial variability. This simplification may be particularly limiting in large and heterogeneous basins, where meteorological conditions exhibit strong spatial gradients and localized responses that are relevant for runoff generation. As a result, spatial aggregation may blur important hydrometeorological signals and thus reduce the ability of models to capture key processes.

Several recent studies have attempted to better represent spatial variability using more complex deep learning architectures. For instance, Anderson and Radić (2022) encoded spatially discretized meteorological inputs using a convolutional LSTM framework, while Sun et al. (2022), and Acosta et al. (2026) used graph neural networks to represent spatial dependencies between locations. Other approaches have focused on improving the representation of meteorological inputs by integrating multiple forcing datasets within a single modelling framework (Kratzert et al., 2021). However, these methods often impose predefined spatial structures or rely on gridded datasets, and may not fully exploit the information available in dense observational networks.

Furthermore, meteorological observations are often affected by missing data, heterogeneous availability and measurement inconsistencies across stations. Gauch et al. (2025) proposed several strategies to address this issue, including masked mean imputation, which augments input data with a binary masking scheme to explicitly account for missing values while preserving temporal structure. This approach is particularly relevant when working with dense observational networks, where missing data are frequent and unevenly distributed across stations.

Despite these advances, the potential of directly using dense networks of raw meteorological observations in regional LSTM frameworks remains insufficiently explored. In particular, most existing approaches still rely on some form of spatial aggregation or predefined spatial representation, limiting the ability of models to learn spatial relationships in a fully data-driven manner.

While preserving the full spatial structure of meteorological observations may improve predictive performance, it also increases model complexity and makes the learned spatial relationships more difficult to interpret. In particular, when information from hundreds of meteorological stations is provided as input, it becomes important to understand which stations and meteorological signals contribute most strongly to streamflow predictions. As deep learning models are often referred to as black-box models (Beven, 2012; Hosseini et al., 2025; Yu et al., 2025), interpreting these relationships becomes particularly challenging.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Explainable AI (XAI) methods have been developed to improve model transparency by identifying the predictors and patterns that drive model outputs (Guidotti et al., 2018). Questions such as which inputs are more relevant for a given prediction and whether the results are physically meaningful can be addressed using XAI methods. The choice of an XAI method depends on both the underlying model and the type of explanation sought. XAI methods can be intrinsic (ante-hoc) or post-hoc, global or local, model-agnostic or model-specific (Theissler et al., 2022).

Accordingly, this study raises two research questions: i) does preserving the spatial structure of meteorological observations improve the performance of regional LSTM models compared to catchment-averaged meteorological inputs? and ii) Which meteorological signals are most influential for streamflow prediction across different catchments?

To answer these questions, this study proposes a regional LSTM model that directly exploits the full spatial structure of raw meteorological observations without any prior spatial aggregation. The model integrates information from more than 700 meteorological stations across the northern Ebro River Basin (> 40,000 km²) to simulate streamflow at 31 gauging stations. Missing data are handled using the masked mean imputation approach of Gauch et al. (2025), enabling the direct use of heterogeneous observational records.

A core contribution of this work is the direct, data-driven use of the full meteorological station network within a regional LSTM, without imposing predefined spatial structures. Instead of prescribing how spatial information should be aggregated, the model is allowed to automatically identify the most informative stations for each gauging location. This results in an adaptive and flexible spatial representation of meteorological information, enabling the effective use of dense and heterogeneous observational data within a regional LSTM framework. The proposed approach is evaluated against the regional LSTM model previously applied to the same region by González-Planet and Juez (2025), and based on catchment-averaged climatic information. In addition, explainable artificial intelligence (XAI) techniques are used to quantify the relative importance of individual meteorological stations for each catchment, providing insights into the spatial and temporal distribution of relevant information and the spatial and temporal patterns learned by the model.

2 Data

The Ebro River Basin is located in north-eastern Spain and is the second largest basin of the Iberian Peninsula. It covers an area of 85,611 km2. The Pyrenees form its northern boundary, while the eastern Cantabrian Range defines its north-western limit. The region of interest in this study corresponds to the northern part of the basin, covering approximately 42,550 km2.

The Pyrenees play a key hydrological role, as nearly 90 % of the total runoff in the Ebro River Basin originates from this mountainous region (Batalla et al., 2004). They present a complex topography, climate, and geological characteristics. The mountainous areas of the Pyrenees, with mean elevations of approximately 2600 m.a.s.l. receive around 2000 mm/year of annual precipitation, and are characterized by cold and humid conditions. Precipitation events are often high-intensity and torrential. In contrast, the Ebro valley, located at approximately 200 m.a.s.l., is influenced by a Mediterranean climate, characterized by warmer and drier conditions, with mean annual precipitation of approximately 350 mm/year (Mariano et al., 2023). Oceanic influences are also evident in the north-west, which is characterized by humid conditions. However, precipitation events there are generally less intense and more persistent than those observed in the mountainous regions.

The strong climatic and topographic contrasts across the study area result in diverse hydrological regimes (Juez et al., 2022). Catchments located in the Pyrenean headwaters are typically characterized by fast and pronounced streamflow responses (Viviroli et al., 2011) , whereas downstream areas exhibit smoother and less variable flow dynamics.

This study considers 31 catchments across the northern part of the Ebro River Basin, representing a wide range of climatological, atmospheric, and hydrological conditions (Figure 1). All selected catchments exhibit natural flow regimes, with

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

no significant anthropogenic influences such as reservoirs, flow regulation, or channelization. The main characteristics of each catchment are summarized in Appendix A.

The dataset comprises 31 gauged catchments. In this study, we discuss four catchments pars pro toto in detail. The chosen four catchments are geographically, climatically, and hydrologically distinct, allowing the attribution patterns to be analysed under a wide range of conditions (Ega-Andosilla, Segre-Seo de Urgel, Veral-Zuriza, and Jerea-Virués gauging stations). July and December results were selected to represent contrasting hydrological conditions, corresponding to dry/warm and wet/cold periods, respectively.

Ega-Andosilla station (A003) drains one of the largest catchments in the study area (1313 km2; Appendix A). The catchment originates in the Cantabrian Range and exhibits a predominantly pluvial streamflow regime. As the gauging station is located near the Ebro River in the Ebro Valley, the catchment is strongly influenced by a Mediterranean climate.

Segre-Seo de Urgel station (A023) drains a catchment slightly smaller than that of Ega-Andosilla, with a comparable mean streamflow (Appendix A). Located in the northeastern part of the study area, it exhibits a pluvio-nival flow regime, with streamflow largely governed by runoff generated in the eastern Pyrenees.

Veral-Zuriza station (A080) drains one of the smallest catchments in the study area, with a mean streamflow of 1.43 m³/s (Appendix A). The gauging station is located at high elevation (1187 m a.s.l.) in the central Pyrenees and exhibits a pluvio-nival flow regime. Among the four selected catchments, it is the most strongly influenced by mountainous climatic conditions.

Jerea-Virués station (A166) presents a pluvial streamflow regime and is located in the western part of the study area. The catchment is relatively small, with a mean streamflow of 4.7 m³/s (Appendix A). Originating in the Cantabrian Range, it is subject to a stronger Atlantic influence than the other selected catchments.

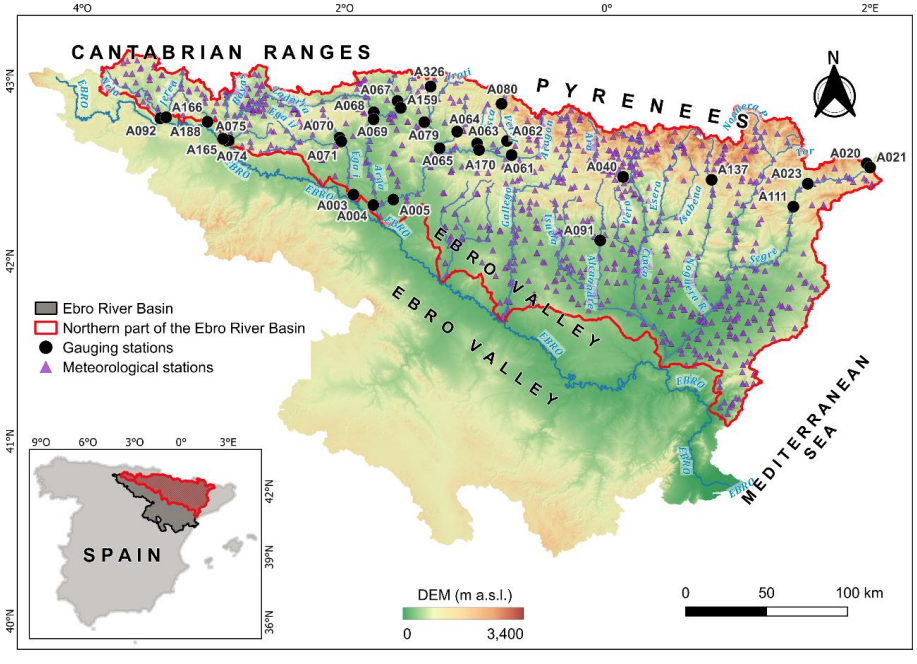


Figure 1: Location of the Ebro River Basin (dark grey) within Spain and the study area (red line). Black dots indicate the gauging stations, and purple triangles represent the meteorological stations.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Both streamflow and meteorological data are available at a daily temporal resolution. Streamflow was used as the target variable, while meteorological observations were used as predictors. Daily streamflow data were collected from the Ebro Water Authorities (CHE) and the Centre for Studies and Experimentation of Public Works (CEDEX). Meteorological data were provided by the Spanish State Meteorological Agency (AEMET).

In contrast to previous studies which uses catchment-averaged meteorological forcings (Lees et al., 2021; González-Planet and Juez, 2025), this work leverages the full extent of the available meteorological network within the Ebro River Basin. Stations with no data during the study period (i.e., those consisting entirely of missing values) were excluded, resulting in a dataset of 777 meteorological stations. We used three meteorological variables as input for each station: precipitation, maximum temperature, and minimum temperature. No pre-processing is required prior to model input, as the model was designed to directly handle missing data.

The study period spans from October 1989 to October 2022. Prior to this period, the abandonment of cultivated fields between 1950 and 1960 led to significant changes in land use. Land use is considered relatively stable from the 1980s onwards (Juez et al., 2022, 2023). The training, validation and test periods are the same used by González-Planet and Juez (2025). That is, the training period is from October 1989 to September 2012, the validation period from October 2012 to September 2017, and the evaluation period from October 2017 to September 2022). Hence, we can compare the proposed regional LSTM model with a prior benchmark of results.

3 Methods

This study proposes a regional model with a LSTM network for daily streamflow simulation using all available meteorological observations within the study area as dynamic inputs. In addition, spatial information from both meteorological and gauging stations was incorporated as static inputs to learn not only temporal but also spatial relationships between regional meteorological information and streamflow responses. This approach allows the model to be trained across multiple catchments simultaneously and learn spatial patterns related to catchment location (Kratzert et al., 2019b).

To train the regional model using raw meteorological station observations, missing values had to be addressed. Instead of applying manual or semi-manual pre-processing — which may introduce approximations and inconsistencies — we adopted the masked mean method proposed by Gauch et al. (2025). This approach allows the direct use of meteorological station data as model input (including stations that measure only a subset of variables, e.g. precipitation only or temperature only).

Missing values are first replaced with a fixed value (zero in this study), while an additional binary mask records whether each observation was originally available or missing. This mask enables the model to distinguish between true observations and imputed values. The input data from each station and time step are then processed by a multilayer perceptron (MLP), which transforms the observations into a vector representation of equal dimension for all stations. The representations corresponding to valid observations are subsequently averaged across stations at each time step, whereas those associated with missing observations are excluded using the mask (Gauch et al., 2025). We set the number of neurons in the MLP to 10 (e in Figure 2b).

For the input we incorporate static meteorological features (latitude, longitude and elevation) together with the dynamic input data (precipitation, and minimum and maximum temperature). Similarly, we concatenate static gauging station features (catchment area, latitude, longitude and elevation) with the masked mean output. The resulting spatio-temporal embedding is then processed by a standard LSTM layer. A simplified overview of the model workflow is illustrated in Figure 2a. See Figure 2b for implementation details (masking, tensor dimensions, embedding module).

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Since joint inference with all 777 meteorological stations was computationally prohibitive, the model processes 20 meteorological stations for a given sample (k in Figure 2b). This number was selected after preliminary experiments, which indicated that 20 stations provided a good balance between predictive performance and computational cost.

During training, the stations are randomly sampled for each training sample — allowing the regional LSTM model to progressively learn from the full set of available meteorological stations over successive epochs. However, during evaluation the model requires a fixed number of 20 meteorological stations as input. Identifying the subset that yields the best predictive performance becomes an optimization problem.

Because the number of possible 20-station combinations is extremely large (approximately 1038), exhaustive search is infeasible. Therefore, we deployed a genetic algorithm to efficiently explore the space of candidate subsets by iteratively selecting, crossing, and mutating high-performing combinations. Because the genetic algorithm is stochastic, it was initialized with different random seeds. Appendix B summarizes the results obtained from multiple initializations of the genetic algorithm. Since the identified station subsets and model performances were broadly consistent across seeds, we present the results from one representative run in the main text. The best subsets were then analysed using explainable artificial intelligence (XAI) methods, specifically Integrated Gradients (IG; Sundararajan et al., 2017) and GradientSHAP (Lundberg and Lee, 2017), to investigate which meteorological stations contributed most strongly to streamflow prediction at each gauging station.

The selected 20 meteorological station subsets are interpreted as computationally feasible, high-information representations of the dense meteorological network, rather than as a replacement of the original full-network training strategy.

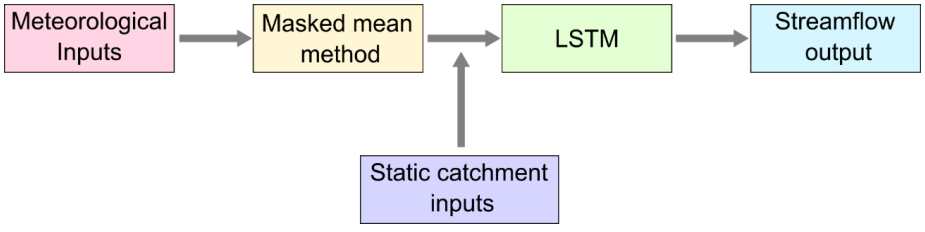


Figure 2a: Simplified overview of the model workflow.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

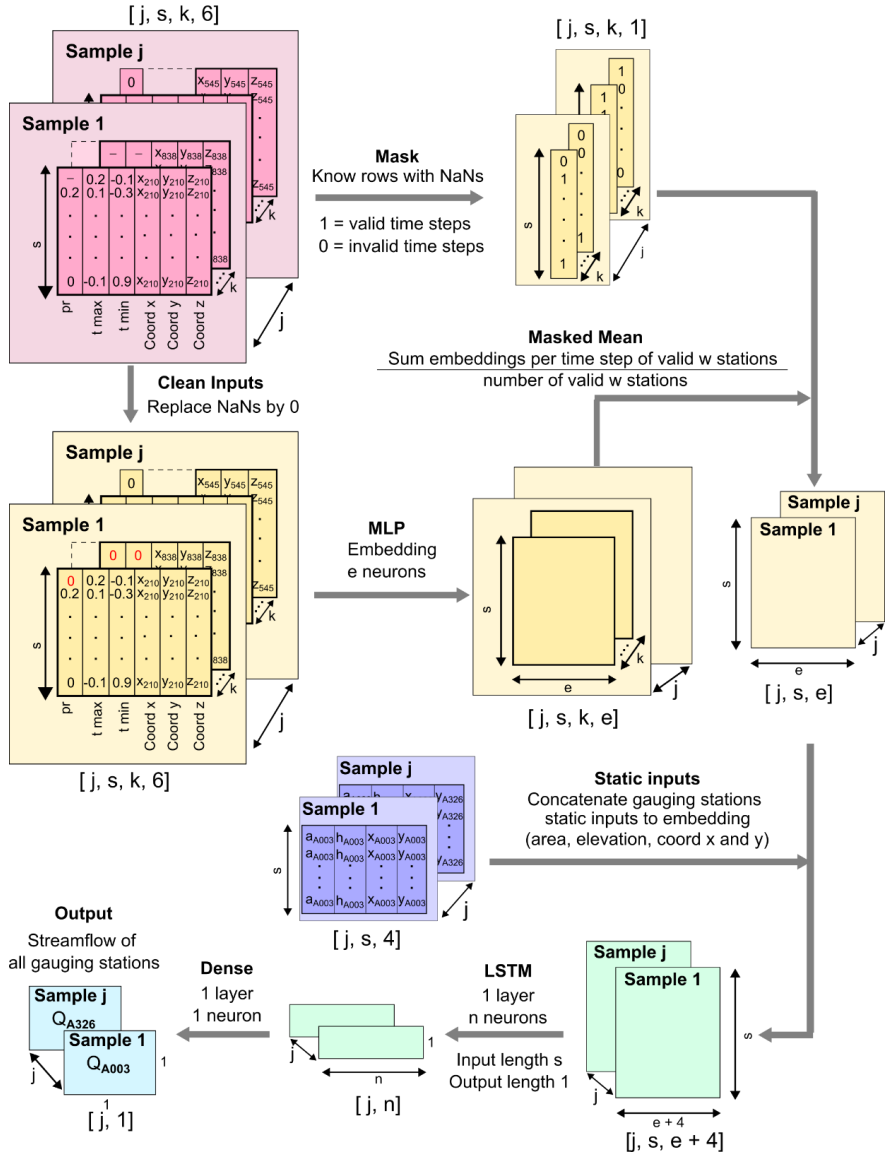


Figure 2b: Detailed workflow of the proposed regional LSTM model. For each training sample (j), a random subset of k meteorological stations is selected from the complete monitoring network. Dynamic meteorological variables are combined with the corresponding station static features and processed through a masked-mean embedding layer (implemented as an MLP with e neurons) to handle missing values and variable station availability. The resulting embeddings are concatenated with static catchment attributes and passed to the LSTM network with n neurons to predict streamflow at the target gauging station. The input sequence length is denoted by s. Pink boxes indicate the dynamic and static inputs related to meteorological stations, yellow boxes represent the processing steps before the LSTM, purple boxes correspond to static catchment features, green boxes denote the LSTM network, and blue boxes represent the model output.

3.1 LSTM network

Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) have become one of the most widely used deep learning architectures for streamflow simulation and currently represent the state of the art in many hydrological applications (Kratzert et al., 2019b, 2024). A detailed description of the regional LSTM architecture adopted in this study is provided in González-Planet and Juez (2025).

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

The LSTM in this study contains 64 neurons (n in Figure 2b) without dropout and is followed by a dense output layer with a linear activation function and a single neuron to produce the simulated streamflow. The input sequence length was set to 365 days (s in Figure 2b). Model training was performed using the Root Mean Squared Error (RMSE) as the loss function and the Adam optimizer. A learning-rate decay strategy was adopted, with learning rates of 10^{-3} during the first 10 epochs, 10^{-4} during the following 10 epochs, and 10^{-5} thereafter. The number of training epochs was determined using an early stopping criterion, whereby training is terminated when no further improvement in model performance is observed.

3.2 Genetic algorithm

A genetic algorithm (GA; Holland, 1992) is a heuristic search algorithm inspired by the natural selection process in biological evolution. A GA operates through successive generations, where each generation consists of a population of candidate solutions. At each generation, candidate solutions from an elite set are preserved and carried over to the next generation. The remaining population is generated through selection of parents, followed by crossover and mutation to produce offspring.

We used a population size of 100 candidate solutions. The number of generations depends on the convergence process, as an early stopping criterion was applied to finish the algorithm when no further improvement is observed. The mutation rate was set to 0.25 and the elite size to 10. Each candidate solution represents a combination of 20 meteorological stations. The performance of each candidate solution was evaluated using the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970) metric, which was used to identify the best-performing combinations.

The genetic algorithm is not part of the LSTM model itself, but a post-training optimization step required to identify an optimal subset of meteorological stations under computational constraints mentioned before (Sect. 4).

3.3 XAI techniques

We use Integrated Gradients (Sundararajan et al., 2017; Kratzert et al., 2019a) and GradientSHAP (Lundberg and Lee, 2017) to interpret the regional model. Both are post-hoc, local, model-specific attribution methods that assign a relevance score to each input feature. For time series applications, attribution scores are computed for every input variable at every time step (Theissler et al., 2022), allowing the contribution of meteorological information to be analysed in both space and time.

IG computes feature attributions by integrating gradients along a straight-line path between a reference baseline and the actual input. This approach is widely used for interpreting differentiable deep learning models and attributes a model prediction to its input features. However, the resulting attributions are sensitive to the choice of baseline, and the straight-line interpolation may pass through unrealistic regions of the input space.

GradientSHAP combines concepts from Shapley value theory with gradient-based attribution methods. It computes feature attributions by averaging gradients obtained from multiple stochastic interpolations between the baseline and the input, providing an approximation of Shapley values. Similar to IG, GradientSHAP s sensitive to the choice of baseline. In addition, the stochastic sampling procedure introduces variability in the attribution estimates, meaning that explanations may differ slightly across runs. Furthermore, the method relies on assumptions regarding feature independence that may not be fully satisfied in practice.

XAI techniques are not necessarily equivalent and may produce inconsistent attributions depending on the formulation and sensitivity to input transformations (Kindermans et al., 2017; Sundararajan et al., 2017). Given the methodological assumptions and limitations associated with individual attribution methods, both IG and GradientSHAP were applied to assess the robustness and consistency of the interpretation results.

Because both methods rely on a reference input, the baseline, its choice can strongly influence attribution values. To ensure comparability between both techniques, the same baseline was used in both analyses (Feng et al., 2022) . The baseline was

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

defined using a precipitation value corresponding to the physical zero expressed in the standardized space, and with maximum and minimum temperatures corresponding to the physical mean, also expressed in the standardized space, i.e. the zero value after standardization.

The use of two explainable AI methods provides a robustness check. Consistent spatial and temporal attribution patterns increase confidence regarding those relationships that drive the model behaviour.

3.4 Metrics

Model performance was evaluated using the Nash–Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), one of the most widely used metrics in hydrology. NSE ranges from -∞ to 1, with a value of 1 indicating a perfect agreement between observed and simulated streamflow. As a global performance metric, NSE provides an overall assessment of model accuracy over the study period.

Because NSE may mask substantial errors under specific flow conditions, particularly during extreme events, the normalized RMSE associated with selected return periods (nRMSETx) proposed by González-Planet and Juez (2025) was also employed. Return periods were estimated based on the 33-year length of the available observation period. In this study, nRMSE values were calculated for return periods of 0.5, 2, and 10 years. The metric ranges from 0 to ∞, with lower values indicating better performance.

Peak timing was additionally evaluated to assess the temporal agreement between observed and simulated streamflow peaks. We define peak events as those exceeding the 90th percentile of the observed streamflow at each catchment. The peak timing error was then computed following the methodology of Kratzert et al., (2021) considering the maximum simulated streamflow within a ±3-day window around each observed peak.

4 Results

4.1 Regional LSTM model performance

Table 1: Performance metrics of the proposed regional LSTM model ("Dist."), which uses spatially distributed raw meteorological station data as inputs, and the benchmark model ("Avg."), which uses catchment-averaged meteorological forcings. Values correspond to the mean and standard deviation across the 31 gauging stations.

	NSE		$nRMSE_{T0.5}$		$nRMSE_{T2}$		$nRMSE_{T10}$		Peak-Timing*	
	Dist.	Avg.	Dist.	Avg.	Dist.	Avg.	Dist.	Avg.	Dist.	Avg.
Mean	0.695	0.640	0.530	0.600	0.442	0.669	0.491	0.778	0.205	0.439
Deviation	0.118	0.151	0.210	0.210	0.158	0.122	0.185	0.096	0.155	0.257

* The mean and standard deviation reported for the peak-timing metric were computed using the absolute time-lag values to avoid cancellation between positive and negative timing errors.

Table 1 summarizes the performance metrics of the proposed regional LSTM model ("Dist."), which uses spatially distributed raw meteorological station inputs, and the benchmark model of González-Planet and Juez (2025), ("Avg."), which uses catchment-averaged meteorological forcings. Appendix B provides the corresponding results for each individual catchment.

In terms of NSE, the proposed model shows a higher mean value across catchments (0.695) than the benchmark (0.640), indicating a better overall agreement between observed and simulated streamflow. The standard deviation is slightly smaller in the proposed model (0.118), reflecting less variability in performance among catchments. This increase in mean NSE indicates an overall improvement in regional model performance when spatially distributed raw meteorological observations are used instead of catchment-averaged forcings.

For low-flow conditions (nRMSET0.5), both models present comparable performance, although the proposed model achieves lower mean error values. The differences become more pronounced for medium-flow (nRMSET2) and high-flow conditions (nRMSET10), where the proposed model achieves lower error values than the benchmark. This suggests that the main benefit

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

of using spatially distributed raw meteorological inputs is not limited to an improvement in overall fit (i.e., in the NSE value), but it is especially evident in the representation of medium- and high-flow events.

The detailed results presented in Appendix B reveal a clear relationship between catchments showing overall improvements in NSE and those exhibiting reductions in the error metrics associated with the 0.5-year return period. For medium- and high-flow conditions, improvements are even more widespread, with nearly all catchments showing lower error values in the proposed model. In contrast, a reduction in overall model performance, as measured by NSE, does not necessarily imply a poorer representation of peak flows. For example, Urederra-Eraul gauging station (A070) shows a slightly lower NSE value in the proposed model, while exhibiting a substantial improvement in the simulation of 10-year return period peak flows, with the corresponding error metric decreasing from 0.618 in the benchmark model to 0.244 in the proposed model. Similar improvements in the representation of extreme flows can also be observed at Ayuda-Berantevilla (A075), Bayas-Miranda (A165), and Jerea-Virués (A166) stations, among others. These results indicate that changes in NSE do not always reflect changes in the model skill for reproducing high-flow events.

Regarding the peak timing metric, the proposed model also achieves better performance across the analysed catchments. The benchmark model exhibits a mean absolute peak timing error across all catchments of 0.439 days (approximately 10.5 hours), while the proposed model reduces this value to 0.205 days (approximately 5 hours), representing a substantial improvement in the temporal representation of peak-flow events. Appendix B reports the individual peak timing values for each catchment. Positive time lag values indicate that the observed streamflow peaks occur before the simulated ones, whereas negative values indicate that the simulated peaks occur earlier then the observations.

Overall, the results are consistent across metrics and flow conditions, with the clearest gains appearing under medium- and high-flow scenarios. This pattern supports the idea that the spatially distributed representation of meteorological inputs provides relevant information for capturing the hydrometeorological conditions associated with runoff generation and peak-flow response.

4.2 Spatial interpretability of the regional LSTM

Figure 3a shows the IG contribution of each meteorological station selected by the genetic algorithm for the Ega-Andosilla station (A003; top) and the Segre-Seo de Urgel station (A023; bottom).

The selected meteorological stations of Ega–Andosilla station (A003) are spatially distributed across the region, covering the eastern and western Pyrenees, the Cantabrian Range, and the Ebro Valley. In December, precipitation shows the highest positive contributions, particularly from stations located in the Cantabrian Range and the western Pyrenees. In contrast, precipitation contributions are negligible in July across all stations. This seasonal contrast is physically consistent with the pluvial regime of the catchment. During wet seasons (including December), headwater regions become the primary source of streamflow generation, whereas in summer (including July), the semiarid conditions of the Ebro River Basin lead the model to effectively down weight the precipitation signal. Maximum temperature shows negative contribution in both months, and the strongest negative contributions concentrated in the Ebro Valley and the Cantabrian Range. This highlights the model ability to capture the influence of high temperatures in the water balance. Higher temperatures increase evapotranspiration demand, reducing the water available for streamflow generation. The selection of stations in the Ebro Valley is physically coherent, as this region records some of the highest temperatures in the basin. The Ebro Valley appears to be particularly informative region for capturing temperature-driven water losses.

The selected meteorological stations for the Segre-Seo de Urgel station (A023) are concentrated in the central and eastern Pyrenees, as well as in the eastern Ebro Valley. In July, the dominant variable is maximum temperature, with the strongest positive contributions from stations located in the Pyrenees. This result is closely linked to the pluvio-nival regime of the

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

catchment, as higher temperatures drive snowmelt processes, providing additional water input to streamflow during warm seasons. At the same time, the model captures evapotranspiration losses associated with high temperatures, reflected in the negative contributions of stations located in the Ebro Valley. In December, precipitation shows the strongest positive contributions, particularly from stations in the Pyrenees, as physically expected. Maximum temperature shows negative contributions, consistent with the hydrological behaviour of the catchment. In the Segre river sub-basin, to which this station belongs, the surrounding mountain ranges limit precipitation inputs due to a rain shadow effect (Grouillet et al., 2016), making snow processes an important component of the winter water balance. Under cold winter conditions, a fraction of precipitation is stored as snow in the Pyrenees, which justifies the negative contribution of the temperature.

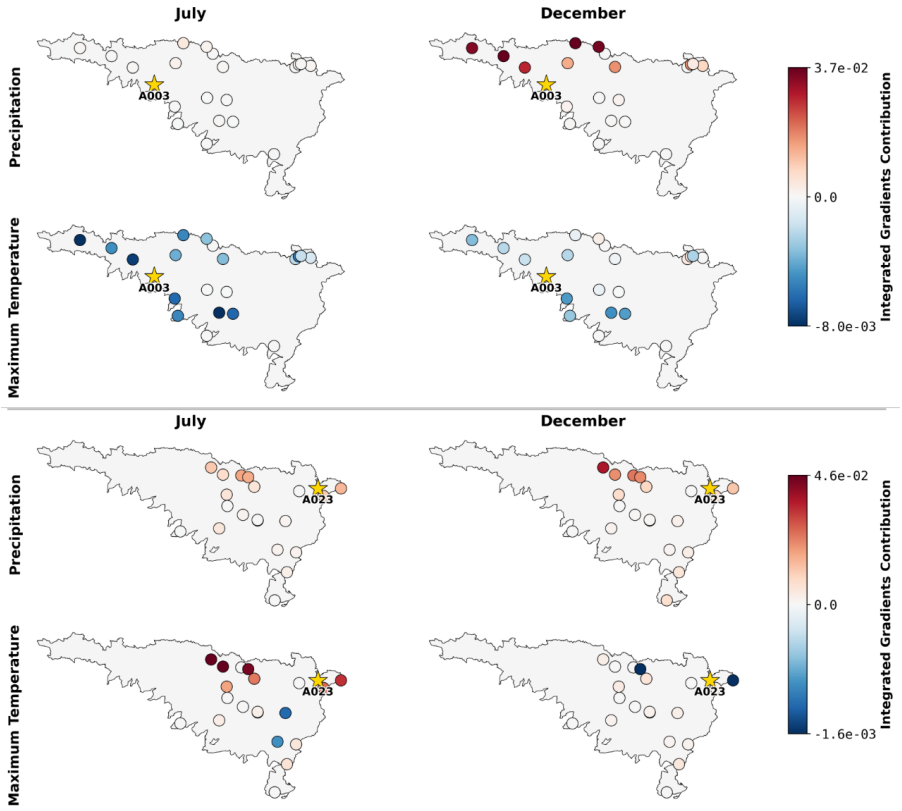


Figure 3a: IG contributions per variable of each meteorological station for the Ega-Andosilla station (A003; top) and Segre-Seo de Urgel (A023; bottom) during July and December. Yellow star indicates the location of the analysed gauging station. Circles indicates the location of each meteorological station and colours are associated with the contribution values according to the colour bar.

Figure 3b shows the IG contribution of each meteorological station selected by the genetic algorithm for the Veral-Zuriza station (A080; top) and the Jerea-Virués station (A166; bottom).

The selected meteorological stations for Veral-Zuriza station (A080) include some near the gauging station in the eastern Pyrenees and Cantabrian Range, which show the most relevant contributions, as well as stations in the Ebro Valley. Precipitation contribution is strongest in winter (December), particularly from the nearest meteorological stations, and diminishes substantially in summer (July) due to the reduced precipitation typical of this season. This pattern confirms the model's ability to capture the spatial and temporal structure of precipitation across the basin. Regarding maximum temperature, the behaviour is similar to that of Segre-Seo de Urgel, as both catchments follow a pluvio-nival regime. In summer, positive

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

contributions are detected from Pyrenean stations, consistent with snowmelt-driven discharge, while negative contributions dominate in winter, reflecting snowpack accumulation processes. Additionally, stations located in the Ebro Valley show positive contributions in summer, consistent with evapotranspiration losses driven by high valley temperatures.

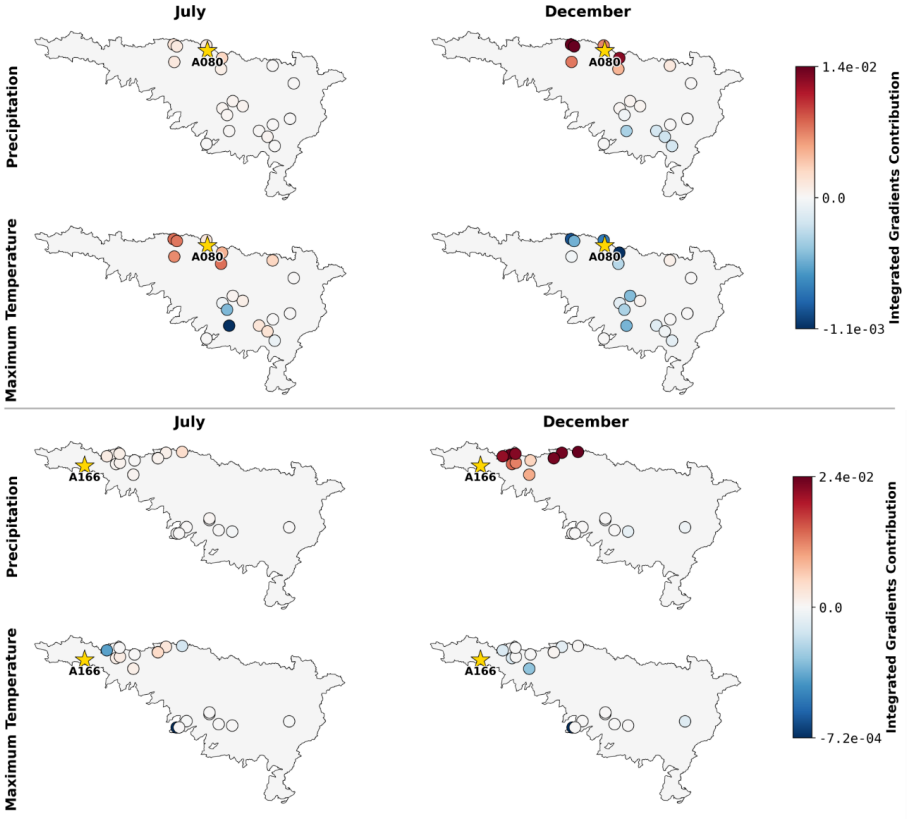


Figure 3b: IG contributions per variable of each meteorological station for the Veral-Zuriza station (A080; top) and the Jerea-Virués station (A166; bottom) during July and December. Yellow star indicates the location of the analysed gauging station. Circles indicates the location of each meteorological station and colours are associated with the contribution values according to the colour bar.

The meteorological stations selected for the Jerea-Virués station (A166) by the genetic algorithm are mostly concentrated in the western and mountainous regions, near the gauging station, with some additional stations located in the Ebro Valley. Precipitation contributes positively to streamflow in both months, with the strongest contributions from stations located in mountainous regions such as the Cantabrian Range and the western Pyrenees. Stations located in the south-eastern part of the basin show negligible contributions. The signal in December is considerably stronger than in July, consistent with the pluvial regime of the catchment and the seasonal precipitation patterns of the western part of the study area. Maximum temperature shows spatially contrasted results in July. Stations located in mountainous regions present slightly positive contributions, whereas stations in the Ebro Valley show negative contributions. This spatial pattern may reflect a marginal snowmelt component during warm seasons at high elevations, combined with evapotranspiration losses captured by lower-elevation stations. In December, temperature contributions from mountainous stations are negligible, which is physically coherent with a purely pluvial regime in which no snow component is present. Lower-elevation stations, however, maintain negative temperature contributions in December, consistent with evapotranspiration losses occurring even during winter months.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Taken together, these results show that the meteorological stations selected by the genetic algorithm are physically coherent with respect to geographical location, seasonal hydrological functioning, and catchment regime. The attribution patterns are also consistent with the role of precipitation and temperature under pluvial and pluvio-nival conditions, suggesting that the model captures meaningful spatial signals rather than arbitrary statistical relationships.

The same analysis was performed using the Gradient SHAP method. To assess the consistency between attribution approaches, Figure 4 reports the Pearson correlation between SHAP and IG values across all catchments and seeds. Correlation is computed between station-wise contribution values derived from both methods. Results show consistently high agreement, with Pearson correlation coefficients exceeding 0.9 in all cases, indicating strong methodological consistency in the inferred spatial attribution patterns. This robustness suggests that the identified importance of meteorological stations is largely independent of the attribution technique used.

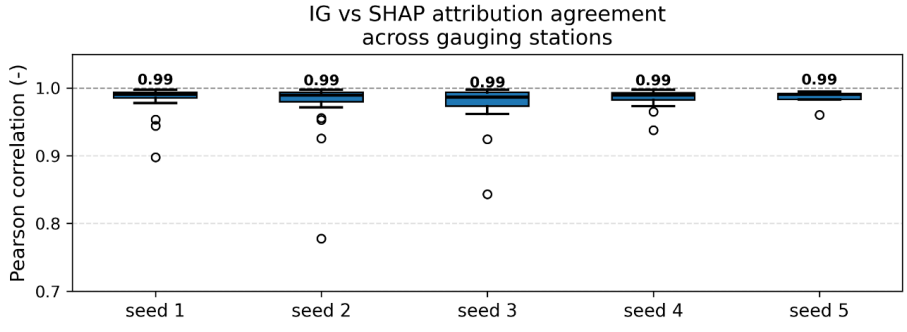


Figure 4: Pearson correlation between SHAP and IG attributions across all catchments and seeds. The horizontal line corresponds to the highest correlation (1).

5 Discussion

The proposed regional LSTM model takes advantage of the dense and spatially distributed meteorological monitoring network available within the Ebro River Basin. Unlike previous approaches based on catchment-averaged meteorological forcings, the direct use of raw observations preserves the spatial variability of meteorological conditions that is otherwise lost through aggregation.

The Ebro River Basin represents a particularly challenging environment for regional streamflow simulation due to its pronounced topographic and climatic heterogeneity. Elevation ranges from the Pyrenees to the Ebro Valley, producing strong gradients in precipitation, temperature, snow accumulation, and evapotranspiration over relatively short distances. Such heterogeneity increases the difficulty of representing meteorological forcing through catchment-averaged values, as these may smooth localized processes that are relevant for streamflow generation.

Despite methodological differences across studies, previous applications based on CAMELS-GB (Lees et al., 2021) and CAMELS-US (Wilbrand et al., 2023) have reported higher global performance using catchment-averaged meteorological forcings in comparison with the global performance obtained under the benchmark conditions. In contrast, the improvement obtained in this study over the benchmark model indicates that preserving the spatial structure of meteorological observations can provide substantial benefits in geographically heterogeneous basins such as the Ebro. These results suggest that predefined spatial aggregation may not be necessary for regional deep learning models and can limit their ability to represent heterogeneous hydrological processes.

The advantage of the proposed approach is further supported by the exceptionally dense and spatial distributed meteorological monitoring network in this study area. The northern part of the Ebro River Basin contains 777 meteorological stations

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

distributed over approximately 42,000 km2, corresponding to roughly one station every 54 km2. This observational density provides an excellent spatial coverage to exploit localized meteorological information.

Our results indicate that preserving the spatial structure of the meteorological observations substantially improves the simulation of medium- and high-flow events. This improvement is particularly relevant given the difficulty of LSTM networks in accurately reproducing peak flows (Xie et al., 2021; Frame et al., 2022; Acuña Espinoza et al., 2025). Rather than relying on a fixed representation of meteorological inputs, the model dynamically adjusts the contribution of individual meteorological stations according to the prevailing hydrological situation (e.g., wet and dry seasons). This mirrors the findings of Kratzert et al. (2021), who found that the LSTM dynamically leverages different meteorological products depending on the region and situation. This flexibility enables the model to better represent peak flows, as reflected by the lower normalized RMSE values for medium and high return period flows (Table 1).

Our interpretability analysis suggests that the selected stations and variables are hydrologically consistent with the dominant processes governing each catchment (Figures 3a, 3b), i.e. precipitation, snow accumulation, snowmelt processes or evapotranspiration losses. These findings suggest that dense meteorological monitoring networks provide valuable spatial information that is partially lost when meteorological forcings are aggregated at the catchment scale.

We also observed that different initializations of the genetic algorithm lead to distinct subsets of used meteorological stations that all exhibit very similar predictive performance across catchments (Table B1). This indicates that the information contained within the meteorological network is highly redundant for predicting streamflow – multiple combinations of stations are able to capture the dominant spatial and temporal variability relevant for streamflow simulation. To assess how independent these alternative solutions actually are, we computed the pairwise Pearson correlation between the streamflow predictions obtained from each pair of seeds, for every gauging station, and compared the NSE of the seed-ensemble mean against the mean NSE of the individual seeds. Predictions from different seeds were highly correlated (mean Pearson's $r = 0.98$ across all seed pairs and stations). Averaging predictions across seeds improved NSE only marginally (mean gain = 0.008), confirming that the different station combinations do not provide complementary, quasi-independent information that an ensemble could exploit. Rather, they behave as near-interchangeable solutions that converge to the same underlying streamflow response. This behaviour is a direct manifestation of the equifinality problem (Beven, 2006): multiple parameterizations — in this case, multiple subsets of input stations — can produce equally acceptable representations of catchment behaviour, so that the "best" combination identified by the optimization is better understood as one of many equally valid solutions than as a uniquely determined optimum.

IG and GradientSHAP show highly consistent attribution patterns (with Pearson's $r > 0.9$). This agreement, together with the hydrologically coherent spatial patterns identified across catchments and seasons, increases our confidence in the robustness of the inferred relationships.

The subselection of 20 meteorological stations strikes a balance between memory and computational effort and the information processed by the LSTM. However, it also represents the main limitation of our study. Although the model is exposed to all available meteorological stations during training, we were not able to simultaneously exploit and study the complete meteorological network during evaluation. Future developments in GPU memory capacity or more memory-efficient neural network architectures may enable the direct use of the complete meteorological network during evaluation.

Although the proposed model is not currently applicable to ungauged catchments, it can be extended to previously unseen gauged catchments. This provides an opportunity to evaluate the generalization capability of the model under conditions where streamflow observations are available but were not used during training.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Another potential limit of our approach is that is very likely sensitive to the density and spatial distribution of meteorological stations. That is, we expect that the high performance observed in this study is partly enabled by the exceptionally dense observational network available in the Ebro River Basin. In future work we will therefore use the available data of the Ebro River Basin to study. The sensitivity of the regional LSTM to different meteorological network densities and spatial configurations remains an important direction for future work.

6 Conclusions

This study shows that preserving the spatial structure of raw meteorological observations can improve regional streamflow simulation in geographically heterogeneous basins.

To demonstrate that we incorporate the spatial information from both meteorological and gauging stations to learn spatial relationships between regional meteorological information and streamflow responses, we analyse the resulting regional LSTM. Our analyses show that the model dynamically adapts the contribution of individual meteorological stations according to the prevailing hydrological conditions and is able to identify physically consistent patterns associated with pluvial and pluvio-nival catchments.

We also examined how performative different subsets of meteorological stations are with regard to prediction quality, and were able to show that different subsets exhibit similar predictive capabilities.

Future work should evaluate its applicability in regions with different hydroclimatic conditions and varying observational network densities and explore direct use of complete meteorological networks.

Code and data availability

Streamflow observations are available under request in the CEDEX gauging network (CEDEX, 2026) and the CHE SAIH monitoring system (CHE, 2026). Meteorological observations were obtained and are available under request in the Spanish State Meteorological Agency (AEMET, 2026). The datasets are maintained by the corresponding data providers.

The code used to train the regional LSTM model, perform the genetic algorithm optimization, and conduct the explainable artificial intelligence analyses are publicized in the CSIC SACO repository (https://saco.csic.es/s/WJF2dQmGSME4eec).

Author contributions

Inmaculada C. González-Planet implemented the regional LSTM model, conducted the experiments, analysed the results, and led the manuscript preparation. Daniel Klotz conceived and co-designed the regional model and provided methodological guidance throughout the study. Carmelo Juez supervised the research from a hydrological perspective and provided overall scientific guidance. D. Klotz supervised the machine learning aspects of the study. All authors contributed to the interpretation of the results, reviewed the manuscript, and approved the final version.

Competing interests

Daniel Klotz is a member of the editorial board of Hydrology and Earth System Sciences.

Acknowledgements

This work is funded by the European Research Council (ERC) through the Horizon Europe 2021 Starting Grant program under REA grant agreement number 101039181-SED@HEAD. The authors are grateful to the Ebro Water Authorities (CHE) and the Centre for Public Works Studies and Experimentation (CEDEX) for access to the hydrological data. We also thank the Laboratory for Climatology and Climate Services (LCSC) and the State Meteorological Agency of Spain (AEMET) for providing climatological and atmospheric data.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix A

Table A1: Characteristics of the selected gauging stations.

ID	SUB-BASIN NAME	STATION NAME	Latitude	Longitude	Drainage Area (km^2)	Elevation gauging station (m.a.s.l.)	Mean Q (m^3/s)	Q_{10} (m^3/s)	Q_{90} (m^3/s)
A003	Ega	Ega-Andosilla	42°22'26.7"N	1°56'56.1"W	1313.24	304	11.03	1.17	25.81
A004	Arga	Arga-Funes	42°18'53.3"N	1°48'04.5"W	2932.14	284	41.21	7.20	97.60
A005	Irati	Aragón-Caparroso	42°20'35.8"N	1°39'02.0"W	5483.95	302	38.17	6.60	87.88
A020	Segre	Carol-Puigcerdá	42°26'48.4"N	1°54'31.6"E	153.00	1170	2.76	0.35	7.13
A021	Segre	Segre-Puigcerdá	42°25'24.8"N	1°55'59.5"E	311.00	1125	2.91	0.61	6.52
A023	Segre	Segre-Seo de Urgel	42°21'08.7"N	1°27'35.1"E	933.42	670	12.73	2.51	28.19
A040	Cinca	Ara-Boltaña	42°26'10.3"N	0° 4'41.79"E	622.85	570	14.83	2.50	32.70
A061	Upper Aragón	Subordán-Javierregay	42°34'36.6"N	0°45'16.0"W	350.61	628	9.15	0.74	21.14
A062	Upper Aragón	Veral-Biniés	42°39'21.0"N	0°47'9.8"W	163.01	650	3.67	0.18	9.37
A063	Upper Aragón	Esca-Sigüés	42°38'41.9"N	1° 0'43.9"W	498.22	520	8.17	0.70	21.52
A064	Irati	Salazar-Aspurz	42°42'51.3"N	1°09'37.0"W	397.74	547	7.79	0.31	21.16
A065	Irati	Irati-Liédena	42°37'25.3"N	1°17'41.6"W	1561.12	412	27.37	4.50	61.30
A067	Arga	Ulzama-Olave	42°53'26.0"N	1°36'24.0"W	231.14	465	5.40	0.28	13.42
A068	Arga	Araquil-Asiáin	42°49'47.8"N	1°47'14.3"W	901.40	390	21.00	1.50	51.50
A069	Arga	Arga-Echauri	42°47'26.9"N	1°47'23.7"W	1858.27	378	35.75	3.50	90.50
A070	Ega	Urederra-Eraul	42°41'30.4"N	2°02'54.9"W	182.20	438	5.52	0.10	14.56
A071	Ega	Ega-Estella	42°40'16.3"N	2°02'09.3"W	808.13	425	10.15	1.44	23.50
A074	Zadorra	Zadorra-Arce	42°40'38.5"N	2°53'48.4"W	1343.81	450	11.50	2.04	23.85
A075	Zadorra	Ayuda-Berantevilla	42°40'50.7"N	2°53'02.7"W	307.62	460	2.58	0.22	6.02
A079	Irati	Erro-Urroz Villa	42°46'13.6"N	1°24'24.0"W	199.61	483	3.39	0.13	9.01
A080	Upper Aragón	Veral-Zuriza	42°51'45.4"N	0°49'13.1"W	46.79	1187	1.43	0.14	3.48
A091	Alcanadre	Alcanadre-Lascellas	42° 5'14.7"N	0° 6'34.8"W	499.76	390	3.73	0.54	8.67
A092	Nela	Nela-Trespaderne	42°48'08.8"N	3°24'06.1"W	1088.34	540	15.58	1.12	38.33
A111	Segre	Segre-Orgaña	42°13'45.9"N	1°20'35.9"E	2078.95	542	24.45	8.20	50.80
A137	Noguera Ribagorzana	Noguera R.-Pont de Suert	42°24'00.1"N	0°44'28.3"E	567.46	824	13.99	5.80	25.60
A159	Arga	Arga-Huarte	42°51'05.5"N	1°35'01.4"W	174.62	453	4.25	0.45	10.43
A165	Bayas	Bayas-Miranda	42°41'20.5"N	2°55'40.9"W	310.06	460	5.10	0.04	13.33
A166	Jerea	Jerea-Virués	42°48'29.0"N	3°21'26.4"W	297.16	532	4.70	0.12	12.69
A170	Upper Aragón	Aragón-Yesa	42°36'35.0"N	0°59'55.3"W	1473.31	497	36.05	4.30	81.19
A188	Omecillo	Omecillo-Berguenda	42°47'01.2"N	3°02'40.9"W	341.93	483	2.77	0.19	7.21
A326	Irati	Urrobi-Espinal	42°58'03.9"N	1°21'14.7"W	48.21	860	1.44	0.19	3.31

480

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix B

Table B1: Performance metrics per catchment of the proposed regional LSTM model ("Dist."), which uses spatially distributed raw meteorological station data as inputs, and the benchmark model ("Avg."), which uses catchment-averaged meteorological forcings. Values for the proposed model are reported as mean ± standard deviation across five evaluated seeds.

	NSE		$nRMSE_{T0.5}$		$nRMSE_{T2}$		$nRMSE_{T10}$		Peak timing (days)	
	Dist.	Avg.	Dist.	Avg.	Dist.	Avg.	Dist.	Avg.	Dist.	Avg.
A003	**0.830** ± 0.015	0.788	**0.433** ± 0.032	0.583	**0.219** ± 0.016	0.503	**0.344** ± 0.034	0.674	-0.108 ± 0.076	**-0.054**
A004	0.713 ± 0.004	**0.826**	0.468 ± 0.015	**0.390**	0.537 ± 0.003	**0.515**	0.734 ± 0.002	**0.649**	**0.085** ± 0.040	0.628
A005	**0.639** ± 0.011	0.333	**0.512** ± 0.013	0.788	**0.580** ± 0.008	0.704	-	-	**-0.104** ± 0.32	-0.156
A020	**0.609** ± 0.033	0.526	**0.317** ± 0.012	0.566	**0.464** ± 0.009	0.646	-	-	**0.263** ± 0.029	0.333
A021	**0.494** ± 0.050	0.406	**0.374** ± 0.030	0.548	**0.568** ± 0.110	0.823	-	-	0.308 ± 0.150	0.800
A023	**0.501** ± 0.061	0.491	**0.248** ± 0.019	0.474	**0.698** ± 0.020	0.770	-	-	**-0.167** ± 0.120	0.556
A040	**0.746** ± 0.019	0.735	**0.265** ± 0.018	0.413	**0.468** ± 0.094	0.840	-	-	-0.352 ± 0.141	**0.055**
A061	**0.550** ± 0.013	0.460	**0.357** ± 0.009	0.387	**0.508** ± 0.007	0.751	-	-	**-0.128** ± 0.089	0.270
A062	**0.716** ± 0.013	0.714	0.402 ± 0.035	**0.396**	**0.307** ± 0.084	0.736	-	-	**-0.218** ± 0.074	0.250
A063	**0.722** ± 0.007	0.693	**0.553** ± 0.023	0.571	**0.400** ± 0.030	0.685	**0.514** ± 0.013	0.852	**0.094** ± 0.072	0.152
A064	**0.699** ± 0.009	0.633	**0.906** ± 0.079	1.076	**0.401** ± 0.035	0.619	**0.560** ± 0.015	0.832	**-0.087** ± 0.128	0.196
A065	**0.651** ± 0.039	0.570	**0.506** ± 0.007	0.714	**0.512** ± 0.012	0.579		-	**0.000** ± 0.101	-0.357
A067	**0.797** ± 0.031	0.792	**0.678** ± 0.036	0.732	**0.376** ± 0.015	0.759		-	**0.257** ± 0.064	1.063
A068	0.658 ± 0.003	**0.773**	**0.552** ± 0.043	0.735	**0.670** ± 0.004	0.785		-	**0.340** ± 0.107	0.796
A069	0.645 ± 0.018	**0.816**	**0.803** ± 0.046	1.056	**0.600** ± 0.006	0.711	**0.769** ± 0.003	0.791	**0.318** ± 0.098	0.634
A070	0.738 ± 0.017	**0.788**	0.479 ± 0.015	**0.404**	**0.378** ± 0.145	0.607	**0.244** ± 0.260	0.618	**-0.333** ± 0.057	0.512
A071	**0.795** ± 0.009	0.768	0.571 ± 0.008	**0.508**	**0.266** ± 0.018	0.584	**0.375** ± 0.045	0.714	**-0.073** ± 0.103	0.244
A074	**0.767** ± 0.012	0.721	**0.515** ± 0.036	0.805	**0.318** ± 0.014	0.527	**0.449** ± 0.008	0.715	0.276 ± 0.105	**0.200**
A075	0.704 ± 0.016	**0.761**	**0.674** ± 0.072	0.841	**0.255** ± 0.024	0.445	**0.129** ± 0.026	0.664	**0.063** ± 0.071	0.595
A079	**0.704** ± 0.024	0.614	0.950 ± 0.092	**0.593**	**0.351** ± 0.040	0.703	**0.594** ± 0.023	0.905	**-0.059** ± 0.118	0.578
A080	**0.325** ± 0.014	0.285	0.641 ± 0.029	**0.484**	**0.719** ± 0.034	0.823	**0.732** ± 0.057	0.897	**0.236** ± 0.202	0.603
A091	0.576 ± 0.039	**0.618**	**0.331** ± 0.010	0.453	**0.527** ± 0.005	0.821		-	-0.513 ± 0.022	**0.243**
A092	0.570 ± 0.023	**0.668**	0.484 ± 0.017	**0.434**	**0.491** ± 0.010	0.619	**0.692** ± 0.003	0.806	**0.542** ± 0.133	0.667
A111	**0.611** ± 0.020	0.595	**0.277** ± 0.016	0.429	**0.667** ± 0.046	0.821		-	**-0.023** ± 0.100	0.256
A137	**0.731** ± 0.013	0.666	**0.209** ± 0.008	0.312	**0.402** ± 0.054	0.767		-	-0.591 ± 0.037	**-0.273**
A159	**0.758** ± 0.033	0.720	**0.707** ± 0.069	0.876	**0.253** ± 0.034	0.597	**0.466** ± 0.033	0.863	**0.009** ± 0.151	0.711
A165	0.659 ± 0.035	**0.723**	**0.444** ± 0.012	0.470	**0.354** ± 0.022	0.588	**0.514** ± 0.031	0.802	**0.156** ± 0.096	0.298
A166	0.682 ± 0.016	**0.730**	0.504 ± 0.016	**0.427**	**0.220** ± 0.024	0.457	**0.383** ± 0.057	0.760	**0.252** ± 0.075	0.531
A170	**0.438** ± 0.009	0.342	**0.417** ± 0.008	0.438	**0.692** ± 0.012	0.826		-	**-0.152** ± 0.144	-0.211

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

A188	0.727 ± 0.014	**0.732**	**0.896** ± 0.122	0.944	**0.317** ± 0.072	0.478		-	**-0.056** ± 0.068	0.556
A326	**0.718** ± 0.024	0.555	0.970 ± 0.101	**0.766**	**0.173** ± 0.055	0.654	**0.353** ± 0.017	0.904	**0.183** ± 0.121	0.843

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Acosta, C. M., Herath, H. M. V. V., Lim, J. Y., Saha, A., Rasnayaka, S., and Marshall, L.: DUALFloodGNN: Physics-informed Graph Neural Network for Operational Flood Modeling, https://doi.org/10.48550/arXiv.2512.23964, 11 May 2026., 2026.

Acuña Espinoza, E., Loritz, R., Kratzert, F., Klotz, D., Gauch, M., Álvarez Chaves, M., and Ehret, U.: Analyzing the generalization capabilities of a hybrid hydrological model for extrapolation to extreme events, Hydrology and Earth System Sciences, 29, 1277–1294, https://doi.org/10.5194/hess-29-1277-2025, 2025.

Agencia Estatal de Meteorología (AEMET). (2026). Meteorological station observations [Dataset]. https://opendata.aemet.es/

Anderson, S. and Radić, V.: Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling, Hydrology and Earth System Sciences, 26, 795–825, https://doi.org/10.5194/hess-26-795-2022, 2022.

Batalla, R. J., Gómez, C. M., and Kondolf, G. M.: Reservoir-induced hydrological changes in the Ebro River basin (NE Spain), Journal of Hydrology, 290, 117–136, https://doi.org/10.1016/j.jhydrol.2003.12.002, 2004.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18–36, https://doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K.: Rainfall-runoff modelling: the primer, 2nd ed., Wiley-Blackwell, Chichester, West Sussex Hoboken, NJ, 1 p., https://doi.org/10.1002/9781119951001, 2012.

Bourdin, D. R., Fleming, S. W., and Stull, R. B.: Streamflow Modelling: A Primer on Applications, Approaches and Challenges, Atmosphere-Ocean, 50, 507–536, https://doi.org/10.1080/07055900.2012.734276, 2012.

Centre for Public Works Studies and Experimentation (CEDEX). (2026). Spanish National Streamflow Gauging Network [Dataset]. https://ceh.cedex.es/anuarioaforos/default.asp.

Confederación Hidrográfica del Ebro (CHE). (2026). SAIH Ebro Streamflow Monitoring Network [Dataset]. https://www.saihebro.com/tiempo-real/mapa-aforos-HG-toda-la-cuenca.

Feng, T., Zhou, Z., Tarun, J., and Nair, V. N.: Comparing Baseline Shapley and Integrated Gradients for Local Explanation: Some Additional Insights, https://doi.org/10.48550/arXiv.2208.06096, 12 August 2022., 2022.

Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L. M., Gupta, H. V., and Nearing, G. S.: Deep learning rainfall–runoff predictions of extreme events, Hydrol. Earth Syst. Sci., 26, 3377–3392, https://doi.org/10.5194/hess-26-3377-2022, 2022.

French, M. N., Krajewski, W. F., and Cuykendall, R. R.: Rainfall forecasting in space and time using a neural network, Journal of Hydrology, 137, 1–31, https://doi.org/10.1016/0022-1694(92)90046-X, 1992.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Cohen, D., and Gilon, O.: How to deal w___ missing input data, Hydrol. Earth Syst. Sci., 29, 6221–6235, https://doi.org/10.5194/hess-29-6221-2025, 2025.

González-Planet, I. C. and Juez, C.: Hydrological insights from a comparative evaluation of LSTM and MC-LSTM networks in the Ebro River basin (Spain), Journal of Hydrology: Regional Studies, 61, 102719, https://doi.org/10.1016/j.ejrh.2025.102719, 2025.

Grouillet, B., Ruelland, D., Vaittinada Ayar, P., and Vrac, M.: Sensitivity analysis of runoff modeling to statistical downscaling models in the western Mediterranean, Hydrology and Earth System Sciences, 20, 1031–1047, https://doi.org/10.5194/hess-20-1031-2016, 2016.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D.: A Survey of Methods for Explaining Black Box Models, ACM Comput. Surv., 51, 93:1-93:42, https://doi.org/10.1145/3236009, 2018.

Hochreiter, S.: The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions, Int. J. Unc. Fuzz. Knowl. Based Syst., 06, 107–116, https://doi.org/10.1142/S0218488598000094, 1998.

Hochreiter, S. and Schmidhuber, J.: Long Short-term Memory, Neural computation, 9, 1735–80, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.



Holland, J. H.: Genetic Algorithms, , Scientific American26766–73https://doi.org/https://www.jstor.org/stable/24939139, 1992.

Holzmann, H.: Status and perspectives of hydrological research in small basins in Europe, Cuadernos de Investigación Geográfica, 44, 601–614, https://doi.org/10.18172/cig.3406, 2018.

Hosseini, F., Prieto, C., and Álvarez, C.: An explainable AI approach for interpreting regionally optimized deep neural networks in hydrological prediction, Journal of Hydrology, 661, 133689, https://doi.org/10.1016/j.jhydrol.2025.133689, 2025.

Juez, C., Lacasta, A., Murillo, J., and García-Navarro, P.: An efficient GPU implementation for a faster simulation of unsteady bed-load transport, Journal of Hydraulic Research, 54, 275–288, https://doi.org/10.1080/00221686.2016.1143042, 2016.

Juez, C., Garijo, N., Nadal-Romero, E., and Vicente-Serrano, S. M.: Wavelet analysis of hydro-climatic time-series and vegetation trends of the Upper Aragón catchment (Central Spanish Pyrenees), Journal of Hydrology, 614, 128584, https://doi.org/10.1016/j.jhydrol.2022.128584, 2022.

Juez, C., Garijo, N., Vicente-Serrano, S. M., and Beguería, S.: Six Decades of Hindsight Into Yesa Reservoir (Central Spanish Pyrenees): River Flow Dwindles as Vegetation Cover Increases and Mediterranean Atmospheric Dynamics Take Control, Water Resources Research, 59, e2022WR033304, https://doi.org/10.1029/2022WR033304, 2023.

Kindermans, P.-J., Hooker, S., Adebayo, J., Alber, M., Schütt, K. T., Dähne, S., Erhan, D., and Kim, B.: The (Un)reliability of saliency methods, https://doi.org/10.48550/arXiv.1711.00867, 2 November 2017., 2017.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrology and Earth System Sciences, 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S., and Klambauer, G.: NeuralHydrology -- Interpreting LSTMs in Hydrology, vol. 11700, 347–362, https://doi.org/10.1007/978-3-030-28954-6_19, 2019a.

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, Hydrology and Earth System Sciences, 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019b.

Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling, Hydrology and Earth System Sciences, 25, 2685–2703, https://doi.org/10.5194/hess-25-2685-2021, 2021.

Kratzert, F., Gauch, M., Klotz, D., and Nearing, G.: HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin, Hydrology and Earth System Sciences, 28, 4187–4201, https://doi.org/10.5194/hess-28-4187-2024, 2024.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., and Dadson, S. J.: Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models, Hydrology and Earth System Sciences, 25, 5517–5534, https://doi.org/10.5194/hess-25-5517-2021, 2021.

Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, https://doi.org/10.48550/arXiv.1705.07874, 25 November 2017., 2017.

Ma, K., Feng, D., Lawson, K., Tsai, W.-P., Liang, C., Huang, X., Sharma, A., and Shen, C.: Transferring Hydrologic Data Across Continents – Leveraging Data-Rich Regions to Improve Hydrologic Prediction in Data-Sparse Regions, Water Resources Research, 57, e2020WR028600, https://doi.org/10.1029/2020WR028600, 2021.

Mariano, M., Abella, S., Araujo, R., Ibisate, A., and Ollero, A.: Nature-Human-River Relationships at the Ebro River and its Delta (Spain), in: River culture: life as a dance to the rhythm of the waters, edited by: Wantzen, K. M., UNESCO, 745–782, https://doi.org/10.54677/KGYR6965, 2023.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F.: The future of Earth observation in hydrology, Hydrology and Earth System Sciences, 21, 3879–3914, https://doi.org/10.5194/hess-21-3879-2017, 2017.

https://doi.org/10.5194/egusphere-2026-5583
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.



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

Patterson, N. K., Lane, B. A., Sandoval-Solis, S., Persad, G. G., and Ortiz-Partida, J. P.: Projected Effects of Temperature and Precipitation Variability Change on Streamflow Patterns Using a Functional Flows Approach, Earth's Future, 10, e2021EF002631, https://doi.org/10.1029/2021EF002631, 2022.

Sun, A. Y., Jiang, P., Yang, Z.-L., Xie, Y., and Chen, X.: A graph neural network (GNN) approach to basin-scale river network learning: the role of physics-based connectivity and data fusion, Hydrology and Earth System Sciences, 26, 5163–5184, https://doi.org/10.5194/hess-26-5163-2022, 2022.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic Attribution for Deep Networks, https://doi.org/10.48550/ARXIV.1703.01365, 2017., 2017.

Theissler, A., Spinnato, F., Schlegel, U., and Guidotti, R.: Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions, IEEE Access, 10, 100700–100724, https://doi.org/10.1109/ACCESS.2022.3207765, 2022.

Viviroli, D., Archer, D. R., Buytaert, W., Fowler, H. J., Greenwood, G. B., Hamlet, A. F., Huang, Y., Koboltschnig, G., Litaor, M. I., López-Moreno, J. I., Lorentz, S., Schädler, B., Schreier, H., Schwaiger, K., Vuille, M., and Woods, R.: Climate change and mountain water resources: overview and recommendations for research, management and policy, Hydrology and Earth System Sciences, 15, 471–504, https://doi.org/10.5194/hess-15-471-2011, 2011.

Wilbrand, K., Taormina, R., Ten Veldhuis, M.-C., Visser, M., Hrachowitz, M., Nuttall, J., and Dahm, R.: Predicting streamflow with LSTM networks using global datasets, Front. Water, 5, 1166124, https://doi.org/10.3389/frwa.2023.1166124, 2023.

Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., and Shen, C.: Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships, Journal of Hydrology, 603, 127043, https://doi.org/10.1016/j.jhydrol.2021.127043, 2021.

Yu, B., Zheng, Y., He, S., Xiong, R., and Wang, C.: Physics-encoded deep learning for integrated modeling of watershed hydrology and reservoir operations, Journal of Hydrology, 657, 133052, https://doi.org/10.1016/j.jhydrol.2025.133052, 2025.