



Comparative Runoff Modeling of Mediterranean Karst Using LSTM, HBV and Mapshed

Ozgun Akdegirmen^a, M. Erol Keskin^{a*}, Necati Agiralioglu^b

^a Department of Civil Engineering, Suleyman Demirel University, Isparta, Turkiye

^b Department of Civil Engineering, Antalya Bilim, University, Antalya, Turkiye

Received: 26 October 2025; **Received in revised form:** 07 December 2025; **Accepted:** 23 December 2025

ABSTRACT:

Despite the complex hydrology of karst systems, which typically necessitates specialized modeling approaches due to rapid flow dynamics and multiple flow regimes, hydrological models may yield adequate results in basins with limited data resolution. This study evaluates the performance of three hydrological models LSTM, Mapshed and HBV in the karstic Kucuk Aksu basin in Turkey's western Mediterranean coast. The research examined nine years (2013-2021) of daily meteorological and streamflow data to identify dominant hydrological elements and evaluate the accuracy of the model. The snow cover indices derived from satellite imagery analysis were used in the training of the LSTM model. The results indicate that LSTM model produced overall better performance metrics compare to two conceptual models. The results have been evaluated according to Nash-Sutcliffe Efficiency (NSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 performance indicators. During calibration, the NSE values for the LSTM, HBV, and Mapshed models were calculated as 0.82, 0.74, and 0.72, respectively, whilst during validation, the NSE values were calculated as 0.90, 0.75, and 0.50, in the same order. The study demonstrates that, despite karstic geology, complex Mediterranean climate patterns and low temporal resolution of the available data, LSTM and HBV models can provide reliable river flow forecast.

KEYWORDS:

LSTM, HBV, Mapshed, Karst, Mediterranean, Hydrology.

1. Introduction

Karst systems, formed by the deformation of soluble rocks, such as limestone, due to the effects of rainfall, can lead to complex hydrological consequences (Ford and Williams, 2007). Karst system aquifers differ from other aquifers through structures that facilitate rapid underground water movement, such as the advancement of water through fractured rock formations and karst conduits (Worthington et al., 2016). This rapid movement of water within the aquifer reduces the recharge to the aquifer and the outflow from the spring cycle to a matter of hours (Hartmann et al., 2014). These unique characteristics of karst

systems, while significantly complicating hydrological modeling, have prompted numerous studies to recommend the integration of rapid flow dynamics, multiple flow regimes, and complex recharge mechanisms into conceptual models to overcome the problem (Chen and Goldscheider, 2014). In addition to the karst structure, regions that exhibit a mountainous Mediterranean climate with high seasonal precipitation variability add another layer of uncertainty in terms of hydrological modeling. Despite the uncertainties introduced by the karstic structures and seasonal precipitation patterns, hydrological models may yield adequate results in catchments characterized by limited time series temporal data resolution. For this reason, a modeling study was conducted to

identify the dominant hydrological element in the Kucuk Aksu basin, located on the Western Mediterranean coast of Turkey, which represents the aforementioned conditions, and to examine the precision of hydrological models. In this study, three models were implemented and evaluated against streamflow observation data and meteorological time series: the Mapshed (GWLF-E) model, based on the SCS Curve Number method; the Hydrologiska Byrans Vattenbalansavdelning (HBV) model, a semi-distributed hydrological model; and the Long Short-Term Memory (LSTM), a machine learning approach. Since the mid-19th century, hydrological models have evolved rapidly. From the first application of the Rational Method, numerous models have been developed to establish the relationship between rainfall and runoff (Todini, 2007). However, hydrological models tend to produce more accurate results when applied to the conditions of the region in which they were developed (Beck et al., 2016). However, through calibration, it is possible to obtain results that are suitable for the conditions of the intended region of application. Mapshed, grounded in the Soil Conservation Service (SCS) CN method, is one of the most widely used approaches worldwide to determine rainfall-runoff relationships and flood volumes. A study conducted in a rural watershed in Louisiana, USA, reported that the model R^2 for discharge was 0.7 after calibration (Pokhrel and Paudel, 2019). In a 2015 study using the Mapshed model to examine the impact of land use data on watershed discharge in the Kerala region of India, the R^2 index for subbasins was found to range between 0.5 and 0.87 (Thottungal, 2015). In another study conducted in 2021 at Indian Mill Creek, Michigan, USA, to simulate water budget, land, and riverbed erosion, the Mapshed model was revealed to overestimate discharge by a factor of 2.8 to 11 (Myers et al., 2021). In a 2024 study using the Mapshed model to examine the impact of decreased rainfall on increased eutrophication in the Iberian Peninsula, the model's performance for watershed discharge yielded a Nash-Sutcliffe Efficiency (NSE) index of 0.61 (Catalan et al., 2024).

In contrast, the HBV model, a semi distributed hydrological model, was developed in 1976 and finalized in 2012 to simulate 2 catchment runoffs. Due to its capability to simulate snowmelt, the model has yielded successful results in Northern climates. In a study conducted in 2021 on the Ussuri River in Russia, it was observed that different versions of the HBV model produced similar runoff estimations, with R^2 values for sub-watersheds ranging up to 0.95 (Lupakov et al., 2021). In another study conducted in the same area, seasonal variations of the parameters of the HBV model and the performance of the model were analysed, and it was observed that the values of R^2 varied between 0.54 and 0.97 in different years

depending on the parameters that changed (Lupakov et al., 2022). On the other hand, a study that evaluated the performance of the HBV model in the Upper Benue River basin of northern Cameroon, the NSE values ranged between 0.68 and 0.89 (Nonki et al., 2021). In studies on HBV flow forecasting in the Russian Kama River, it was observed that the NSE values ranged between 0.81 and 0.91 in 8 of the 9 sub-watersheds (Simonov et al., 2021). In a study comparing different hydrological models using data from flow observation stations at six hydroelectric power plants in the Tocantins River basin in Brazil, the NSE values for the HBV model were found to be between 0.6 and 0.9 (Avila et al., 2023). In a study conducted by the General Directorate of Water Management in the south-east of the Mediterranean basin in Turkey using the HBV model, it was observed that 8 out of 10 basins had NSE values of 0.31 and below. However, the fact that calibration values were not disclosed in the study on a sub-basin basis made it difficult to evaluate the results (Yaslioglu et al., 2018).

On the other hand, the LSTM model is one of the machine learning approaches that has begun to gain importance in recent years and whose applicability has increased with the enhancement of computational power. Compared to other machine learning methods, the LSTM model stands out with the success it demonstrates in time series forecasting and hydrological applications. In a study conducted in the Velika Morava catchment in Serbia, a four-layer LSTM model was employed for river flow forecasting, and R^2 and NSE values were calculated as 0.85 and 0.77, respectively, during the test period (Leščešen et al., 2025). In a study conducted in the Posina River basin in the Italian Alps, the NSE values of HEC-HMS, ARIMAX, and LSTM models for river flow forecasting were calculated as 0.82, 0.67, and 0.93, respectively (Perazzolo et al., 2025). In a study conducted in four different sub-basins of Lake Urmia located in northwestern Iran, a comparison was made between LSTM and RNN models. In the conducted study, whilst both models yielded highly successful results, it was observed that the LSTM model exhibited NSE values ranging from 0.48 to 0.98 during the validation phase on a sub-basin basis, whereas the RNN model demonstrated values ranging from 0.73 to 0.97 (Nakhaei et al., 2024). In a study conducted in the Humber River basin located in western Newfoundland, a comparison was undertaken between LSTM and SWAT models. In this study, it was observed that the SWAT model overestimated peak flows. When the results were compared, it was observed that the SWAT and LSTM models achieved NSE values of 0.77 and 0.95, respectively (Islam et al., 2025). In a study conducted in the Shahrchay River in northwest Iran using an ensemble gene expression programming approach, NSE values of 0.96 and

0.90 were obtained during model validation for 1-day and 2-day ahead forecasts, respectively (Rahmani-Rezaei et al., 2020). In a study conducted in the Kentucky River basin in the United States examining LSTM model performance in comparison to conventional machine learning approaches, the LSTM model achieved R values of 0.99 in both training and testing across multiple scenarios (Rahimzad et al., 2021). Furthermore, in a study on the calibration of multigene genetic programming to simulate low flows in the Moselle River, it was observed that advanced data-driven approaches showed efficacy in capturing low-flow conditions, highlighting the importance of model selection for specific hydrological contexts (Mehr and Demirel, 2016).

This study aims to measure the performance of LSTM machine learning approach alongside Mapshed and HBV conceptual hydrological models in the Kucuk Aksu basin, where the seasonally fluctuating precipitation pattern of the Mediterranean climate converges with karst geology, and to provide guidance for hydrological studies to be conducted in the region. Whilst previous research has typically advocated for specialised modelling approaches in karst systems due to their complex flow dynamics, this investigation demonstrates that both conceptual and machine learning models can yield satisfactory results when appropriately calibrated in data scarce environments. A distinctive methodological contribution is the integration of satellite derived snow cover indices from Sentinel-2 and Landsat-8 imagery into the LSTM training process, thereby addressing the spatial disconnect between low-elevation meteorological stations and high-altitude snow accumulation zones. Furthermore, the study employs Bayesian optimisation across 2,000 model iterations to identify optimal hyperparameters for the LSTM architecture, ensuring robust performance evaluation. By systematically comparing these diverse modelling paradigms across a nine-year period (2013-2021) in the Kucuk Aksu basin, this research provides valuable insights into model selection criteria for Mediterranean karst watersheds and contributes to the broader discourse on hydrological model applicability under complex environmental conditions and data limitations.

2. Study area

The study was carried out in the sub-watersheds of the Antalya Basin, located in the south-west of Turkey, specifically in the Kucuk Aksu sub-watershed, located on the northeastern border of the Aksu Basin. The basin is surrounded by the Taurus Mountains in the north, with an elevation of 2,447 meters, and covers a total area of 241.84 km² (Fig. 1).

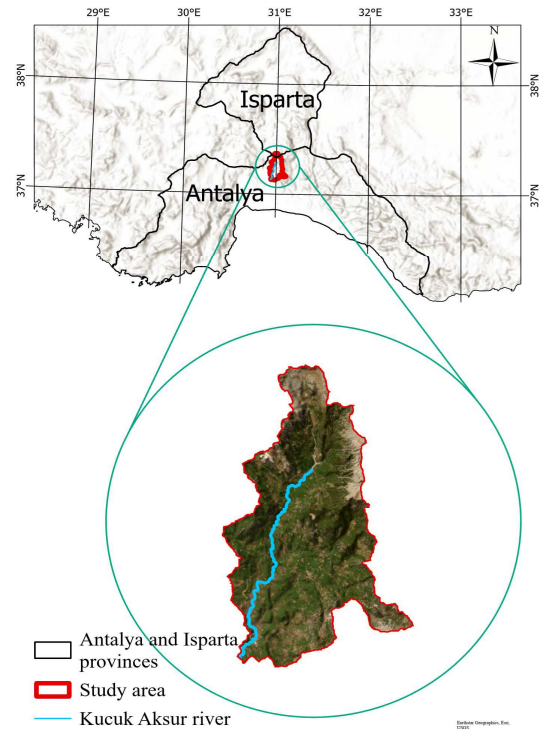


Fig. 1. The Kucuk Aksu Basin is one of the sub-basins of the Antalya Basin facing the Mediterranean, located between the provincial boundaries of Isparta and Antalya

Due to the moderate to high slope topography of the watershed and also to the various points of accumulation of flows within the area, the area could also be considered a high-risk flood zone (Baykal et al., 2023). At the same time, the region has a karstic geology, causing the water from the melting snow to rise among the limestone and form natural springs (Baykal et al., 2022). The basin was developed within the Tauride fold and thrust belt due to the long-term convergence of the African and Eurasian plates, recording a complex Neogene history of sedimentation and deformation. Its more than 1 km thick marine sedimentary infill, spanning from the Middle Miocene to the Pliocene, is characterized by distinct lithostratigraphic units and is significantly deformed by major thrust faults and folds. The geological record reflects a dynamic interplay between compressional and extensional tectonic phases, influenced by regional plate interactions and subduction processes beneath the Isparta angle (Wasoo and Koc, 2021). In the downstream part of the basin, the elevation is 55 meters; the highest areas of the basin are located in its northern and eastern parts. The Kucuk Aksu River, which flows through the centre of the basin, is 25 km long, and its upstream sections are mostly dry during the arid periods of the year. Nevertheless, during the rainy seasons, field investigations and satellite imagery showing water erosion patterns were used to identify points of lateral inflow to the

river. The corresponding drainage networks at these points were mapped using ArcGIS (Fig. 2).

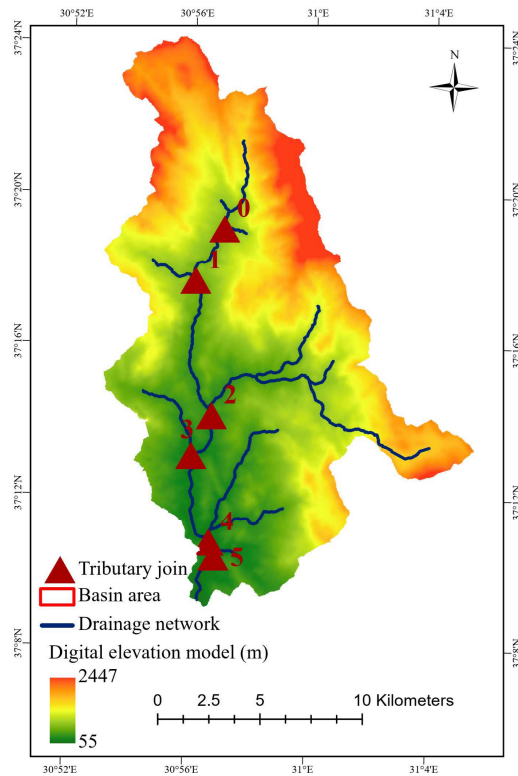


Fig. 2. Kucuk Aksu Basin digital elevation model, drainage network and tributary join locations

The system is dominated by the Kucuk Aksu River, which traverses the basin in a generally north to south orientation. Five tributaries join the main channel along its course, contributing to a dendritic drainage network that converges before progression toward the coastline. The combination of steep slopes, variable precipitation patterns, and high-relief terrain shapes both the flow regime of the river and the vegetation distribution (Fig. 3). The basin is covered with pine forests, reflecting the typical Mediterranean vegetation of its elevation (Kavgaci et al., 2021).

Field investigations and CORINE land use data consistently indicate that agricultural activities in the basin are minimal (Fig. 3). Therefore, the amount of water allocated for river irrigation is negligible, which aids hydrological models in establishing a natural precipitation-runoff relationship. Additional categories include sclerophyllous vegetation, natural grasslands, transitional woodland scrubs, and patches of agricultural land interspersed with pastures, particularly in the lower altitudinal zones and along valley floors. Several mineral extraction and construction sites, denoting localized anthropogenic influences (Fig. 3). As can be seen in Fig. 4, the months with the highest intensity of rainfall within the basin are January and December. On the other hand, July and August stand out as dry seasons.

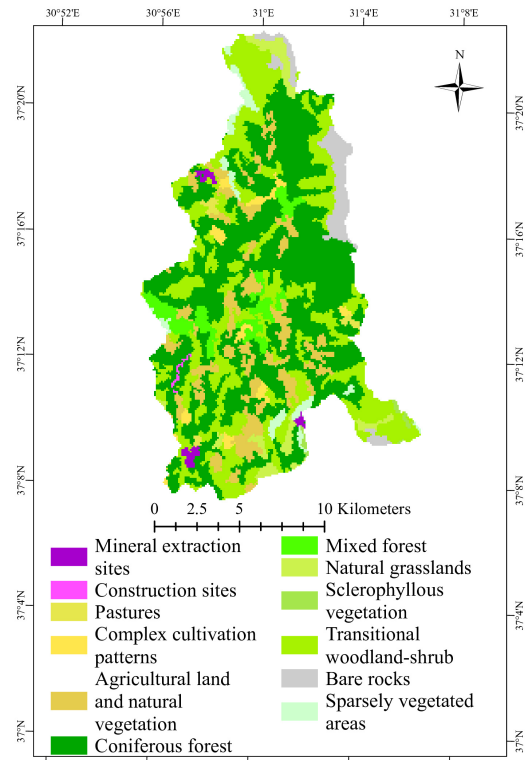


Fig. 3. Kucuk Aksu Basin digital elevation model, drainage network and tributary join locations (EEA, 2020)

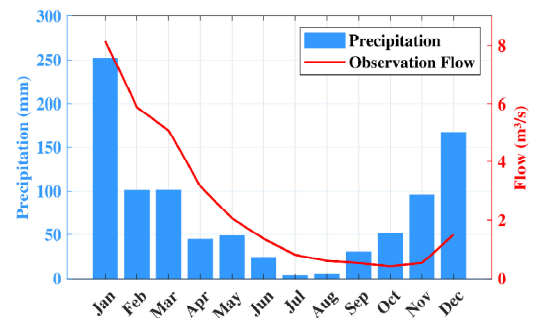


Fig. 4. Monthly average precipitation and flow observation height

According to meteorological data, it was observed that the average annual precipitation of 9 years was 1035 mm, which is one of the characteristics of the climate of the area (Aksu, 2023). Despite the fact that air temperature did not fall below 4 °C in the region during the analysis period according to meteorological data, Sentinel 2 and Landsat 8 satellite data from 2015-2021 were used to detect snow cover in the basin using the Google Earth Engine (ESA, 2015) (U.S. Geological Survey, 2013-2021) (Google Earth Engine Team, 2023). The examination revealed that snow cover in the mountains and foothills in the eastern part of the basin began in September, reached its maximum in January, and gradually decreased from February,

ending in May. As it will be discussed more in depth, the phenomena stand out as the most challenging basin characteristic for hydrological models due to under representation by the meteorological stations.

3. Methods

In the applications of the models, the input data have been aggregated to ensure that the input data used in the models were consistent to facilitate a more accurate analysis of the results. Due to the fact that the Mapshed model uses average data from two closest meteorological stations (Evans and Corradini, 2016), the same time series data were also used in the HBV and LSTM models. All meteorological data were obtained from the Turkish State Meteorological Service (MGM), and the elevations of the stations were 75 and 129 m. The study selected a 9-year period from 2013 to 2021, during which the data was uninterrupted, as the analysis period. The models were run for six predetermined segmentation points that corresponded to six sub-basin areas, contribute to flow hydrograph at the end of the basin outlet (Fig. 2). On the other hand, development of the LSTM model a more holistic approach has been followed due to the single point of flow measurement. Precipitation (mm), binary snow indicator driven from satellite imagery analysis and historical measured flow data (2013-2019) were used to train the model for 7 years of data and 2 years 2020, 2021 used for model validation. Two-layer LSTM model structure was implemented with the MATLAB deep learning toolbox. After sensitivity analysis most influential hyperparameters optimised with experiment manager interface with two thousand runs, and the best performing model on validation was used for comparison of the output forecast.

3.1. Mapshed model

Mapshed model is one of the modeling tools for the estimation of the nutrient, sediment loads and surface runoff for the watershed. The model has the ability to simulate water quality variations due to different types of land use and hydrology over long time periods. Daily weather parameters and time series input data, such as precipitation, temperature, and land use information, are utilised to model surface runoff, evapotranspiration, and base flow. The model uses the SCS-Curve number method and the Hamond method to establish precipitation and runoff relations and to calculate PET (potential evapotranspiration), respectively. The model is capable of simulating the transport of pollutants from point sources and non-point sources within the watershed (Fig. 5).

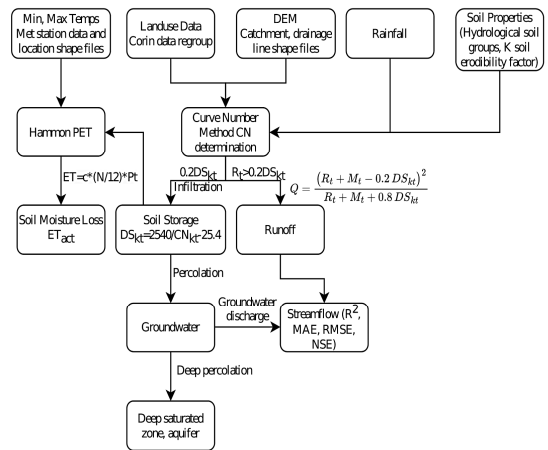


Fig. 5. Flow chart of Mapshed modeling analysis

Non-point source distributions, also called distributed pollution model, tracks nitrogen, phosphorus and sediment loads from areas such as urban, agricultural, and forest land. Transportation of the load is the recognition and incorporation of the transport of nutrients and sediments from land to water bodies through overland flow and subsurface paths. The model also enables the application of best management practices (BMPs) with the aim of reducing the pollutant loads with different application scenarios, empowering the model to investigate a possible scenario prior to field applications. The runoff from the watershed is calculated based on the curve number (CN), as shown in Eq. (1) (Evans and Corradini, 2016).

$$Q_{kt} = \frac{(R_t + M_t - 0.2DS_{kt})^2}{R_t + M_t + 0.8DS_{kt}} \quad (1)$$

Eq. 1 calculates the flow rate Q_{kt} in a specific time t . While R_t signifies rainfall, M_t stands for snow melt and DS_{kt} stands for soil storage for a specific time t . Eq. (1) is a representation of the CN method runoff equation with a combination of snowmelt. On the other hand, Eq. (2) calculates soil storage depending on CN_{kt} , which is an event-based CN parameter in a specific time t .

$$DS_{kt} = \frac{2540}{CN_{kt}} - 25.4 \quad (2)$$

The model applies the variable CN_{kt} according to the antecedent moisture conditions. Therefore, soil moisture changes due to consecutive precipitation events; correspondingly, the runoff value is modified throughout the course of the simulation (Haith et al., 2001).

As it can be seen from Table (1) initial CN values range between 30 and 100. However, for the areas, particularly covered with bare rocks, values were increased more than the models' initial estimations.

Table 1. Mapshed parameters

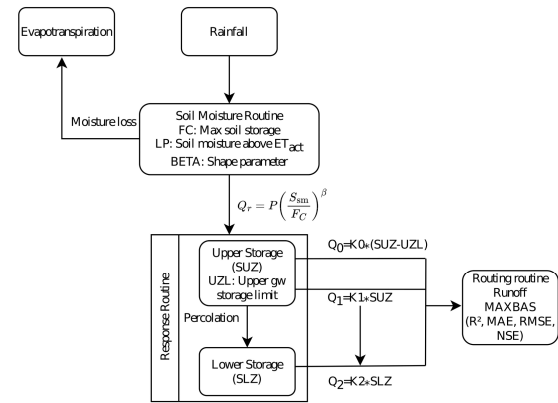
Parameter	Description	Value range	Value used
CN (Curve Number) *	Empirical value depending on land use, vegetation coverage, and topography	30–100	spatial variety
K (Soil erodibility factor)	Soil's potential for erosion	0.02–0.69	spatial variety (0.55–0.66)
ET Cover coefficient	ET adjustment coefficient according to environmental factors	0.3–1	Seasonal (0.55–1)
Groundwater seepage coefficient	Fraction of infiltrated water to deep percolation	0–1	0.02
Groundwater recession coefficient	Groundwater recession due to lateral flow	0–0.2	0.02

*Hydrological soil group "A" has been used for all the sub-basins inside of the watershed (Adanir, 2018)

On the other hand, the K soil erodibility factor was calculated using the parameter relationships presented in Tran's 2011 study (Tran et al., 2011) used, along with the soil texture parameters derived from MODIS data processed by machine learning (Hengl et al., 2017). Furthermore, the hydrological soil groups (Adanir, 2018) and the R rainfall erosivity parameters were determined on the basis of previous field studies (Dogan, 1987). K parameter for the watershed was found to be between 0.55 and 0.66, which is an indication of the high tendency of the soil to erosion. ET adjustment coefficient was reduced to 0.55 from the initial value of 1 due to the establishment of previous studies on the overestimation of the Hamond method (Cruff and Thompson, 1967).

3.2. HBV model

HBV model is a conceptual hydrological model developed in Sweden for the simulation of runoff in river basins (Nonki et al., 2021). The model focuses particularly on streamlining the modeling of certain critical hydrological processes such as snowmelt, changes in soil moisture, and the flow of groundwater within a basin. The strength of the model, especially data rich areas, lies in dividing the basin into smaller subbasins based on elevation and vegetation zones, making it applicable to different hydrological conditions within the basin. It has the capability of simulating runoff based on limited input data such as precipitation, temperature, and potential evapotranspiration. The runoff dynamics of the HBV model can be summarized in Fig. 6.

**Fig. 6.** Flow chart of HBV modeling analysis

The HBV model is widely used in flood prediction, water budget studies, flood risk analysis, and evaluation of the effects of climate change on water resources.

$$Q_{runoff} = Q_0 + Q_1 + Q_2 \quad (3)$$

The runoff in the HBV model can be investigated in three segments as Q_0 , Q_1 , and Q_2 respectively representing the runoff from the upper zone above the threshold (mm / time step), the runoff below the threshold from the upper zone (mm / time step) and the runoff from the lower soil storage (mm/time step).

$$Q_0 = K_0 \cdot (SUZ - UZL) \quad (4)$$

Where K_0 is the flow coefficient for the upper groundwater zone above the threshold value (1/time step), SUZ is the storage in the upper zone (mm), and UZL is the threshold value (mm).

$$Q_1 = K_1 \cdot SUZ \quad (5)$$

K_1 represents the flow coefficient or near surface flow, for the remaining upper groundwater zone below threshold value (1/time step) (Nonki, 2021).

$$Q_2 = K_2 \cdot SLZ \quad (6)$$

SLZ is the storage in the lower zone (mm), and K_2 is the recession coefficient for the lower groundwater zone (1/time step) (Lupakov, 2022).

$$\frac{Q_r}{P} = \left(\frac{S_{sm}}{FC}\right)^\beta \quad (7)$$

When, Q_r is the recharge into the soil via infiltration (mm), P is the total precipitation (mm), S_{sm} is the momentary soil moisture storage (mm), FC is the maximum soil storage (mm), and β is a constant controlling the non-linear relationship, also referred to as the shape constant. The range of values of the HBV model parameters and the values of the parameters that have been used after calibration are given in Table (2).

Table 2. HBV parameters

Parameter	Description	Value range	Value used
FC	Maximum soil moisture storage in mm	0–∞	150
LP	Evaporation reduction factor	0–1	1
β	Shape parameter of soil moisture storage and gw recharge	0–∞	1
PERC	Percolation in mm/day	0–∞	3
UZL	Groundwater recession due to lateral flow	0–∞	100
K_0	Quick flow coefficient	0–∞	0.85
K_1	Interflow coefficient	0–∞	0.07
K_2	Baseflow coefficient	0–∞	0.05
MAXBAS	Runoff transformation parameter	1–100	1

The FC, LP, and β parameters can be classified as soil moisture parameters under soil routine. On the other hand, PERC, UZL, K_0 , K_1 and K_2 can be classified as runoff parameters under response function. MAXBAS is the parameter that defines the runoff transformation. The increase in MAXBAS levels corresponds to a delay in peak flow in time and a reduction in peak flow values. Considering the basin scale and input time series data, this value has been capped at 1.

3.3. LSTM model

Although LSTM models were introduced in 1997 for the first time for hydrology applications, starting from Fangs' work on the extension of soil moisture data in 2017 (Hochreiter and Schmidhuber, 1997)(Fang et al., 2017). From that point onwards, the model's application to rainfall–runoff modeling rapidly gained prevalence owing to its ability to capture long-term temporal dependencies. For LSTM modeling, the 2-layer architecture was selected over single-layer or deeper networks to balance temporal feature extraction capability with overfitting risk given the limited training period (2013–2019). The two-layer configuration allows the first LSTM layer (sequence output mode) to extract low-level temporal patterns while the second layer (last output mode) consolidates these into a complete representation, an advantage over simpler GRU alternatives that may not fully capture the long-term dependencies in rainfall runoff processes. Although some researchers in the literature highlight the benefits of three-layer structures for model accuracy, the two-layer architecture was favoured due to its computational efficiency (Zhang et al., 2025). The architecture incorporates dropout (0.2) between LSTM layers and fully connected layers (20 units, then 1 unit) for non-linear transformation, providing regularization while maintaining computational efficiency

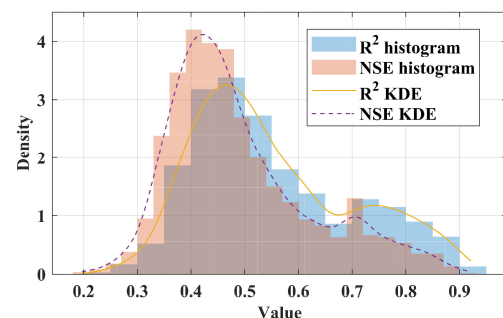
compared to more complex Transformer-based approaches. Following to determination of the model structure, Bayesian optimization was employed due to the hyperparameters of the model exerting a nonlinear influence on the model forecast.

The LSTM model is highly dependent on the optimization of its hyperparameters. Using MATLAB Experiment Manager with Bayesian optimization, two thousand model runs were carried out, and the set of hyperparameters that perform the best is presented in Table 3.

Table 3. LSTM hyperparameters

Parameter	Description	Value range	Value used
Look back	Look-back window for sequence creation	3–14	4
LSTM units	Number of hidden units in each LSTM layer	150–500	324
Learn rate drop factor	Learning rate drop factor	0.3–1.0	0.367211
Initial learn rate	Initial learning rate	1×10^{-4} – 1×10^{-3}	0.00058
L2 regularization	L2 regularization factor	1×10^{-5} – 1×10^{-2}	1.02×10^{-5}

As shown in Fig. 7, during the calibration the mode of the NSE values remained around 0.4, while the R^2 values were approximately 0.5.

**Fig. 7.** Validation R^2 and NSE Kernel Density Estimation (KDE) during calibration runs

When examining the KDE values of model parameter optimisation runs, it is observed that only the portion below 0.5 in terms of density among the total models achieves the desired NSE and R^2 values of 0.9 and above. The distribution of the performance indicators during the model runs highlights the importance of calibration for LSTM models.

4. Results

During the sensitivity analysis, each model parameter was varied within a range of +20% to -20%, and the effects on the overall model flow data were examined. The results indicated that, relative

to the baseline, the parameters of the HBV model FC, LP, β , K1 and MAXBAS influenced the model output of 2.74%, 1.23%, 1.5%, 0.14% and 0.54%, respectively. In contrast, the parameters PERC, UZL, K_0 , and K_2 were found to be ineffective. Similarly, in the Mapshed model, the parameters CN, ET, GWSC (groundwater seepage coefficient) and GWRC (groundwater recession coefficient) affected the model output by 0.01%, 0.44%, 2.35%, and 0.03%, respectively, while the K parameter was determined to have no effect on flow volume. Furthermore, calibrations for Mapshed and HBV were performed using a combined approach that incorporated sensitivity analysis and manual calibration, taking into account performance indicators for years 2013 and 2014. In the LSTM model, unlike the HBV and Mapshed models, sensitivity analysis was conducted on the following parameters whilst keeping other parameters constant: dropout rate, gradient threshold, initial learning rate, L2 regularisation, learning rate drop factor, learning rate drop period, lookback window, number of LSTM layers, LSTM units, and maximum epochs. These parameters were selected for optimisation within predetermined parameter ranges, taking computational resources into consideration. Having said that, during the parameter optimisation the number of LSTM layers has been limited to 2 due to the substantial negative impact on computational time and two-layer solution's proven track record (Li and Yuan, 2023). According to the results, the initial learning rate, L2 regularisation, learning rate drops factor, lookback window, and number of LSTM units' parameters were selected for optimisation, as they affected model forecasting by 37.4%, 23.1%, 10%, 35.9%, and 17.5%, respectively. These percentages represent the maximum differences observed between 3 to 5 observation points within the determined sample range. The analysis of models for the 9-year dataset is focused on 4 performance indicators (RMSE, MAE, R^2 and NSE). According to observation data, the runoff coefficient (river flow/precipitation) of the Kucuk Aksu Basin was determined as 32%, it was determined as 34%, 28% and 34% from the results of the Mapshed, HBV and LSTM models, respectively.

As can be seen in Fig. (9), the months with the highest rainfall within the basin are January and December. Despite the fact that air temperature did not fall below 0 °C in the region during the analysis period according to meteorological data, Sentinel 2 and Landsat 8 satellite data from 2015-2021 were utilised to detect snow cover in the basin (ESA, 2015) (U.S. Geological Survey, 2013-2021). The examination revealed that snow cover in the mountains and foothills in the eastern part of the basin began in November, reached its maximum in February, and gradually decreased from February, ending in May (Fig. 8).

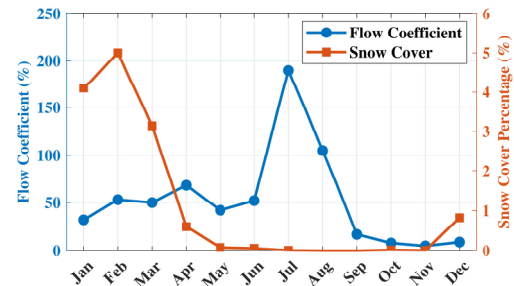


Fig. 8. Monthly average basin flow coefficient and snow cover in percentage

The seasonal variation of the percentage that snow cover occupies within the total basin area, obtained as a result of processing snow cover data acquired from Landsat 8 and Sentinel 2 satellites with a 30% cloud filter, is shown in (Fig. 8). With the analysis of the relationship between the basin runoff coefficient and snow cover, the effect of snow melt within the basin on the basin runoff coefficient can be observed from January through September (Fig. 8). On the other hand, the peak experienced in July following the snow cover reaching its maximum in February and approaching zero in May stems from these months being relatively dry in terms of precipitation compared to the rest of the year. The scarcity of precipitation has resulted in the proportionally high maintenance of snowmelt derived flow, which constitutes the river base flow, and consequently in a high runoff coefficient. In further articulation of the seasonal precipitation presented in Fig. (4) and Fig. (9), it is evident that July and August are the driest months and that the flow in the basin during these months is largely attributed to the effects of snowmelt.

The river flow response during the months of September, October, November, and December, in relation to increasing precipitation beginning in September, was found to be relatively lower compared to the rest of the year (Fig. 9). In addition, both physical models predicted higher flows compared to the river flow observed during these months (Fig. 9). However, during the dry months of July and August, both conceptual models forecast lower flows compared to the observed data.

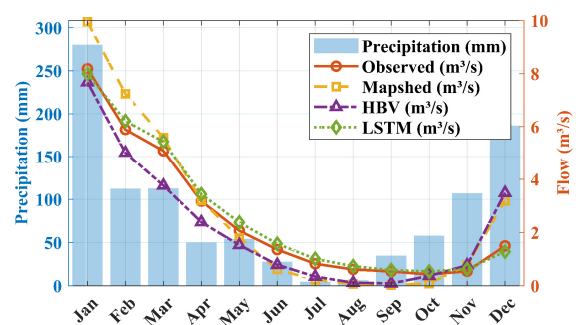


Fig. 9. 9 year monthly average precipitation and monthly average model forecasts

In contrast, the LSTM model produced monthly forecasts that, relative to the averages over nine years, were proportional to the observed data.

The model forecasts were subjected to linear regression analysis and performance indices such as R^2 , RMSE, NSE, and MAE were calculated to assess model performance.

Among the total data set, the HBV model showed the lowest performance in terms of NSE values with 0.48 in the year 2018 (Table 4). In contrast, NSE values for the Mapshed model of 0.26 and below have been observed in 2016, 2017, 2018, and 2021.

Table 4. HBV model performance indicators.

Year	MAE	RMSE	R^2	NSE
2013	0.99	1.4	0.775	0.714
2014	1.14	1.95	0.762	0.753
2015	1.88	2.77	0.75	0.649
2016	0.8	1.21	0.677	0.602
2017	0.64	1.15	0.804	0.774
2018	0.81	1.64	0.528	0.484
2019	1.67	2.43	0.915	0.89
2020	1.21	3.52	0.826	0.789
2021	1.17	4.42	0.673	0.592

Through further investigation of the performances of these years was carried out, it was observed that the Mapshed model forecast higher flow responses to precipitation in these years (Table 5). However, RMSE and MAE values were found to be similar in both models within the total analysis period. Furthermore, it was observed that the coefficients of determination values were greater than 0.5 in both models on a yearly basis and greater than 0.7 in the total analysis period.

Table 5. Mapshed model performance indicators

Year	MAE	RMSE	R^2	NSE
2013	0.83	1.2	0.857	0.789
2014	1.23	2.18	0.838	0.693
2015	1.32	1.96	0.876	0.825
2016	1.24	2.27	0.545	-0.396
2017	0.97	3.28	0.497	-0.837
2018	0.79	1.97	0.522	0.263
2019	1.53	2.37	0.952	0.896
2020	1.72	4.18	0.915	0.703
2021	1.51	7.49	0.733	-0.172

According to the results LSTM model demonstrate significantly better performance on average compared to the Mapshed and HBV model (Table 6). Thorough the 9 years analysis period

LSTM and R^2 and NSE values stayed consistently above 0.66 and 0.61 respectively. Although LSTM model performed poorly compared to the models' baseline on years 2016, 2017 and 2018 similar to Mapshed model.

Table 6. LSTM model performance indicators

Year	MAE	RMSE	R^2	NSE
2013	0.52	0.97	0.866	0.863
2014	0.95	1.57	0.848	0.84
2015	1.53	2.59	0.803	0.693
2016	0.72	1.19	0.798	0.619
2017	0.83	1.33	0.729	0.698
2018	0.9	1.43	0.66	0.611
2019	1.29	2.33	0.918	0.9
2020	1.26	2.64	0.906	0.882
2021	1.22	1.8	0.96	0.932

Comparative analysis of three hydrological forecasting models Mapshed, HBV, and LSTM through a nine-year evaluation period (2013-2021) reveals distinct performance characteristics that warrant careful consideration for operational implementation. The LSTM neural network consistently shows superior forecasting capability, achieving an overall coefficient of determination R^2 of 0.86 and NSE of 0.86 (Fig. 10).

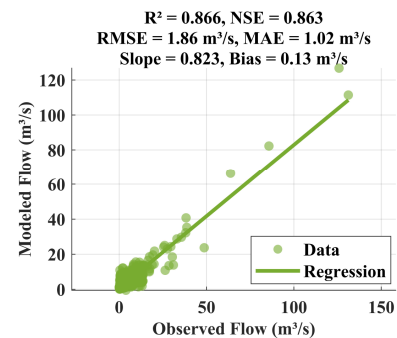


Fig. 10. LSTM linear regression analysis

Complementing these correlation-based metrics, the LSTM model exhibits the lowest error magnitudes with RMSE of 1.86 and MAE of 1.02, positioning it as the most robust performer across all evaluation criteria examined. It is also important to note that, although the LSTM model forecasts have a positive bias on average, during higher flows it has been observed that the model tends to forecast lower than the observed data.

The HBV model occupies an intermediate position within the performance hierarchy, achieving reasonable predictive capacity with an R^2 of 0.769 and an NSE of 0.749 (Fig. 11). However, the model demonstrates a notable negative bias of -0.32, suggesting a systematic tendency toward underestimation that requires consideration in operational contexts.

The associated RMSE 2.51 and MAE 1.15 values, while moderate in absolute terms, indicate substantially greater prediction uncertainty compared to the LSTM approach. The model's slope parameter of 0.660 further corroborates the systematic underestimation tendency, deviating considerably from the ideal relationship.

In contrast, the Mapshed model presents a more complex performance profile that raises significant concerns regarding operational reliability. Whilst achieving a superficially comparable R^2 value of 0.784 to the HBV model, the Mapshed approach demonstrates considerably poor NSE performance at 0.519, indicating substantial deviations from observed flow patterns in peak flow conditions that undermine its practical utility. The model's elevated RMSE of 3.48 and positive bias of 0.21 suggest systematic overestimation tendencies, particularly during extreme hydrological events (Fig. 12).

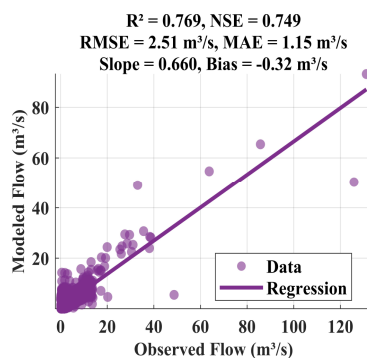


Fig. 11. HBV linear regression analysis

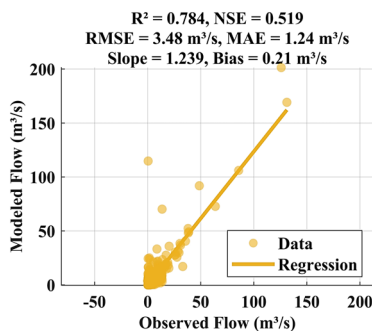


Fig. 12. HBV linear regression analysis

5. Conclusion

The LSTM model is observed to show superior and relatively consistent performance compared to the other two physical models in all performance indicators. Nevertheless, owing to the use of the final two years as the basis for testing to prevent biased behaviour of the model during hyperparameter optimisation, the LSTM model demonstrated relatively higher success in last two years compared to the years used for training. To avoid this situation and prioritise the general performance of the model, testing by sampling

from random years during the model calibration phase would constitute a reasonable approach. At this juncture, another applicable method could be climate parameter-based sample selection. Given that climate variables and precipitation patterns have tendency to change over time rather than subject to sudden changes. For this reason, target performance indicators selected from years in which different climate parameters are observed will ensure continuity in model performance over the total duration. On another hand NSE values of physical models indicate that the HBV model provides more accurate results compare to Mapshed model, despite the fact that the coefficient of determination values is similar for the total dataset. This suggests that the HBV model produces results closer to the observed data during peak discharges. However, it should be noted that both HBV and Mapshed models demonstrate lower performance during the dry period. In contrast, the LSTM model has exhibited acceptable results over 9 years, including during drought years.

Despite snow covered areas in the basin not exceeding 5% over the 9-year period, snow effects are observed within the basin. As can be seen from Fig. (8), snow cover begins to diminish from February onwards and falls below 1% in May; however, this effect is more prominently observed during the dry July and August months when precipitation is substantially reduced. Furthermore, with evaluation of Fig. (4) and Fig. (8) together, the absence of a response in river flow, despite increased precipitation in September, October, and November cannot be explained by snow effects. This is because snow cover is at a minimum during these months, indicating that precipitation generally occurs in the form of rainfall. However, this situation can be explained by the reduction in soil moisture following the dry July and August months and, consequently, the increased infiltration capacity, whereby precipitation occurring during these months increases soil moisture levels rather than contributing to river flow. As can also be observed from Fig. (9), the snow effects within the basin pose a challenge for the models. The primary reason for this is that whilst the elevation within the basin varies between 55 m and 2444 m, the meteorological stations are positioned at 75 m and 129 m. The inadequate representation of the basin by the meteorological stations diminishes the accuracy of the models in representing seasonal effects. A binary snow indicator has been created whereby months during which snowmelt occurs are designated as "1" for the LSTM model, and other months as "0". Furthermore, at this juncture, a superior approach could have been the incorporation of monthly average snow cover extents into the machine learning training. Nevertheless, no snow corrections have been implemented in the HBV and Mapshed models.

The use of daily total precipitation data and daily average flow data to evaluate the basin's hydrology complicates the examination of the region's karst hydrology and renders the observation of potential rapid responses. Together with that, no data pertaining to potential springs within the basin could be obtained. However, the similarity of the basin's runoff coefficient to those of other basins in the region may lead to interpretations that the karst structure is not influential in the basin's hydrology. On the other hand, both the uncertainty regarding the locations and discharge rates of springs and the low resolution of time series data impede the determination of the efficacy of the karst geology. In contrast, the low temporal resolution of the data has potentially enabled concept hydrological models to yield acceptable results comparable to those documented in the literature for other regions. Nevertheless, LSTM and HBV models produced acceptable results despite the complex climatic conditions and geology of the region.

Considering the semi-arid and highly seasonal rainfall variability of the Mediterranean climate, it is crucial to account for seasonal effects in the application of models. Furthermore, the hydrological soil groups, Group A, used in the Mapshed model were obtained from a previous study conducted in the region. Although the aforementioned study was extensive in terms of test locations and number of applications, seasonal effects were not considered during the experimental procedures. On the other hand, land use data have been regrouped in order to meet the input criteria of the Mapshed model so that some of the important land use and surface cover information might be lost in the procedure.

The intended use case of the models plays a significant role in evaluating their results. The better performance of the HBV model in terms of NSE values, especially during peak flow events, suggests that it could be favoured over the Mapshed model for tasks such as estimating hydraulic structure design load or flood analysis, where peak discharge is demanding. In addition, considering the runoff coefficients, both models produce results closer to the observed basin runoff coefficient, indicating that both models can be conveniently applied for basin water budget calculations. On the other hand, the LSTM model demonstrated that it constitutes a model that can be preferred across all application areas within the region by achieving more successful forecasts compared to SCS Curve Number and soil moisture accounting based conceptual models. Nevertheless, whilst the LSTM model does not possess high requirements in terms of data diversity, it requires continuous and long-term data to be used for model training, as 7 years of 9 years daily stream flow data utilised for model

training while last 2 years used for model validation in this study.

Declarations

- No funding has been used for this study.
- The authors report that there are no competing interests to declare.
- The data is available from the corresponding author upon request.

References

- Adanir S, "Asagi Aksu Cayi havzasindaki jeolojik birimlerin gecirimlilik ozellikleri ve SCS yuntemi ile yagis-akis iliskisinin incelenmesi", Master Thesis, Suleyman Demirel Universitesi, Turkey, 2018.
- Aksu HH, "Analysis of monthly precipitation at the basin scale in Turkiye", Turkish Journal of Agriculture-Food Science and Technology, 2023, 11, 1973-1985.
<https://doi.org/10.24925/turjaf.v11i10.1973-1985.6445>
- Avila L, Silveira R, Campos A, Rogiski N, Freitas C, Aver C, Fan F, "Seasonal streamflow forecast in the Tocantins River Basin, Brazil: An evaluation of ECMWF-SEAS5 with multiple conceptual hydrological models", Water, 2023, 15 (9), 1695. <https://doi.org/10.3390/w15091695>
- Baykal T, Sener E, Terzi U, "Application of analytical hierarchy process for flood risk analysis: A case study in Kucuk Aksu River basin (Antalya, Turkey)", Iranian Journal of Science and Technology, Transactions of Civil Engineering, 2023, 47 (4), 2449-2466.
<https://doi.org/10.1007/s40996-023-01055-4>
- Baykal T, Terzi U, Sener E, "Structural flood analysis based on the hydraulic model of the Kucuk Aksu River", Muhendislik Bilimleri ve Tasarim Dergisi, 2022, 10 (3), 1084-1096.
<https://doi.org/10.21923/jesd.1099665>
- Beck HE, Dijk AIJM, Roo A, Miralles DG, McVicar TR, Schellekens J, Bruijnzeel LA, "Global-scale regionalization of hydrologic model parameters", Water Resources Research, 2016, 52 (5), 3599-3622.
- Catalan J, Monteoliva AP, Vega JC, Domínguez A, Negro AI, Alonso R, Garcés BV, Batalla M, García-Gómez H, Leira M, Nuño C, Pahissa J, Peg M, Pla-Rabés S, Roblas N, Vargas JL, Toro M, "Reduced precipitation can induce ecosystem regime shifts in lakes by increasing internal nutrient recycling", Scientific Reports, 2024, 14 (1), 12408.
<https://doi.org/10.1038/s41598-024-62810-9>
- Chen Z, Goldscheider N, "Modeling spatially and temporally varied hydraulic behavior of a folded karst system with dominant conduit

- drainage at catchment scale, Hochifen-Gottesacker, Alps", *Journal of Hydrology*, 2014, 514, 41-52.
- Cruft R, Thompson TH, "A comparison of methods of estimating potential evapotranspiration from climatological data in arid and subhumid environments", Prepared by USGPO, US, 1967.
- Dogan O, "Turkiye yagislarinin erozyon potansiyelleri", Prepared by Tarim Orman ve Kuyisleri Bakanligi, Kuy Hizmetleri Genel Mudurlugu, Ankara, Turkey, 1987.
- EEA, "Corine Land Cover 2018, Version 2020", Copernicus Land Monitoring Service, 2020.
<https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac>
- ESA, "Sentinel-2 multispectral imaging data",
https://sentinel.esa.int/web/sentinel/mission/s/sentinel-2_2015
- Evans BM, Corradini KJ, "MapShed Version 1.5 Users Guide", 2016.
- Fang K, Shen C, Kifer D, Yang X, "Prolongation of SMAP to spatio-temporally seamless coverage of the continental US using a deep-learning neural network", *Geophysical Research Letters*, 2017.
- Ford D, Williams P, "Dissolution: Chemical and Kinetic Behaviour of the Karst Rocks", John Wiley and Sons, Ltd, Chichester, UK, 2007, 39-76.
- Google Earth Engine Team, "Google Earth Engine: A planetary-scale platform for Earth science data and analysis".
<https://doi.org/10.1016/j.rse.2017.06.031>
- Haith DA, Mandel R, Wu RS, "GWLF: Generalized Watershed Loading Functions, Version 2.0 User's Manual", 2001.
- Hartmann A, Goldscheider N, Wagener T, Lange J, Weiler M, "Karst water resources in a changing world: Review of hydrological modeling approaches", *Reviews of Geophysics*, 2014, 52 (3), 218-242.
- Hengl T, Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shanguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B, "SoilGrids250m: Global gridded soil information based on machine learning", *PLOS One*, 2017, 12 (2), 1-40.
- Hochreiter S, Schmidhuber J, "Long short-term memory", *Neural Computation*, 1997, 9, 1735-1780.
- Islam K, Daraio JA, Cheema M, Sabau G, Galagedara L, "Improved streamflow prediction accuracy in boreal climate watershed using a LSTM model: A comparative study", *PLOS Water*, 2025, 4 (4), 1-26.
<https://doi.org/10.1371/journal.pwat.0000359>
- Kavgaci A, Balpinar N, Uner HH, Arslan M, Bonari G, Chytry M, Carni A, "Classification of forest and shrubland vegetation in Mediterranean Turkey", *Applied Vegetation Science*, 2021, 24 (2), 12589.
<https://doi.org/10.1111/avsc.12589>
- Leščešen I, Tanhapour M, Pekárová P, Miklánek P, Bajtek Z, "Long short-term memory (LSTM) networks for accurate river flow forecasting: A case study on the Morava River basin (Serbia)", *Water*, 2025, 17 (6), 907.
<https://doi.org/10.3390/w17060907>
- Li J, Yuan X, "Daily streamflow forecasts based on cascade long short-term memory (LSTM) model over the Yangtze River basin", *Water*, 2023, 15 (6), 1019.
<https://doi.org/10.3390/w15061019>
- Lupakov SY, Bugaets AN, Shamov VV, "Application of different structures of HBV model to studying runoff formation processes: Case study of experimental catchments", *Water Resources*, 2021, 48, 512-520.
<https://doi.org/10.1134/S0097807821040126>
- Lupakov SY, Bugaets AN, Shamov VV, "Evaluation of the HBV hydrological model parameters using field observation data on experimental catchments (South of Primorsky Krai)", *Geography and Natural Resources*, 2022, 43, 287-294.
<https://doi.org/10.1134/S1875372822030088>
- Mehr AD, Demirel MC, "On the calibration of multigene genetic programming to simulate low flows in the Moselle River", *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 2016, 21 (2), 365-376.
- Myers DTL, Rediske RR, McNair JN, Allen ME, "Watershed and streambank erosion modeling in a coldwater stream using the GWLF-E model: application and evaluation", *Modeling Earth Systems and Environment*, 2021, 7, 1551-1564.
<https://doi.org/10.1007/s40808-020-00882-y>
- Nakhaei M, Zanjani H, Nakhaei P, Gheibi M, Moezzi R, Behzadian K, Campos LC, "Comparative evaluation of deep learning techniques in streamflow monthly prediction of the Zarrine River basin", *Water*, 2024, 16 (2), 208.
<https://doi.org/10.3390/w16020208>
- NASA/METI/AIST, "ASTER Global Digital Elevation Model (GDEM), Version 2", NASA EOSDIS Land Processes DAAC, US, 2011.
- Nonki RM, Lenouo A, Tshimanga RM, Donfack FC, Tchawoua C, "Performance assessment and uncertainty prediction of a daily time-step HBV-light rainfall-runoff model for the Upper Benue River Basin, Northern Cameroon", *Journal of Hydrology: Regional Studies*, 2021, 36, 100849.

- <https://doi.org/10.1016/j.ejrh.2021.100849>
- Perazzolo D, Lazzaro G, Fiume A, Fanton P, Grisan E, "Streamflow forecasting: A comparative analysis of ARIMAX, rolling forecasting LSTM neural network and physically based models in a pristine catchment", *Water*, 2025, 17 (15), 2341. <https://doi.org/10.3390/w17152341>
- Pokhrel BK, Paudel KP, "Assessing the efficiency of alternative best management practices to reduce nonpoint source pollution in a rural watershed located in Louisiana, USA", *Water (Switzerland)*, 2019, 11, 1714.
- Rahimzad M, Moghaddam Nia A, Zolfonoon H, Soltani J, Danandeh Mehr A, Kwon HH, "Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting", *Water Resources Management*, 2021, 35 (12), 4167-4187. <https://doi.org/10.1007/s11269-021-02937-w>
- Rahmani-Rezaeieh A, Mohammadi M, Danandeh Mehr A, "Ensemble gene expression programming: a new approach for evolution of parsimonious streamflow forecasting model", *Theoretical and Applied Climatology*, 2020, 139 (1), 549-564. <https://doi.org/10.1007/s00704-019-02982-x>
- Simonov YA, Semenova NK, Khristoforov AV, "Short-range streamflow forecasting of the Kama River based on the HBV model application", *Russian Meteorology and Hydrology*, 2021, 46, 388-395.
- Thottungal DBA, "Effect of landuse change on discharge using MapShed model in sub basins of bharathapuzha river basin", *International Conference on Hydraulics, Water Resources and River Engineering (Hydro-2015)*, 2015.
- Todini E, "Hydrological catchment modelling: past, present and future", *Hydrology and Earth System Sciences*, 2007, 11 (1), 468-482.
- Tran TA, Mitani Y, Ikemi H, Matsuki H, "Human impacts on erosion and deposition in Onga River basin, Kyushu, Japan", *Memoirs of the Faculty of Engineering, Kyushu University*, 2011, 71 (2), 47-65.
- U.S. Geological Survey, "Landsat 8 OLI/TIRS Collection 2 Level-1 Imagery: Mediterranean Region, Turkey", U.S. Geological Survey, Sioux Falls, South Dakota, US. <https://earthexplorer.usgs.gov/2013-2021>
- Wasoo M, Koc A, "Aksu Havzasi'nin (Antalya, Turkiye) Neojen stratigrafisi ve yapisal unsurlari", *Turkiye Jeoloji Bulteni*, 2021, 64 (1), 83-128. <https://doi.org/10.25288/tjb.682776>
- Worthington SRH, Davies GJ, Alexander EC, "Enhancement of bedrock permeability by weathering", *Earth-Science Reviews*, 2016, 160, 188-202.
- Yaslioglu R, Yaslioglu G, Uray G, Yaslioglu A, Eryildirim M, Yildizhan M, Durmaz Z, Kocbay GR, Bulut B, Demirtepe H, Dogan NB, Husemi Afsar M, Unen A, Tanir T, Dogulu N, Duzenli E, Ersoy NE, Tanrikulu V, "Dogu akdeniz havzasi kuraklik yonetim plani", Prepared by T.C. Orman ve Su Isleri Bakanligi Su Yonetimi Genel Mudurlugu, Turkey, 2018.
- Zhang P, Hu S, Xiao Y, Chen P, Sun D, "Application of stacked bidirectional LSTM neural networks in reservoir porosity prediction", *Scientific Reports*, 2025, 15, 39333. <https://doi.org/10.1038/s41598-025-23095-8>