



A Validated Hybrid Modelling Framework for Climate-Resilient Flood Risk Management in Data-Scarce Semi-Arid Catchments

Maryam Jan Ahmadi¹ · Seyedbasir Saidi² · Leila Khajenoori³

Received: 10 December 2025 / Accepted: 18 March 2026
© The Author(s) 2026

Abstract

Flood risk in semi-arid regions is intensifying due to climate change; however, reliable prediction remains a fundamental challenge in catchments with severe hydrometric data scarcity (station density < 0.2 per 1000 km²). To address this gap, we developed and validated a novel hybrid modeling framework that integrates a physically based hydrological model (Hydrologic Engineering Center - Hydrologic Modeling System, HEC-HMS) with interpretable, Bayesian-optimised machine learning techniques (Random Forest, *eXtreme Gradient Boosting*, *XGBoost*). The SHapley Additive exPlanations (SHAP) method was employed to interpret and identify the key drivers of flooding. SHAP analysis revealed that antecedent precipitation and soil permeability are the primary controlling factors in flood occurrence. The validated framework demonstrates robust performance, achieving excellent accuracy in flood event classification (Precision-Recall- Area Under the Curve (PR-AUC)=0.98) and reducing false alarm rates by 87.5% compared to the standalone physical model. This study provides a practical, validated tool that enhances prediction reliability and delivers actionable insights for climate-resilient flood risk management in data-scarce environments.

Keywords Flood prediction · Machine learning · SHAP · HEC-HMS · CMIP6 · Semi-arid catchment

✉ Leila Khajenoori
lkhajenoori@lancashire.ac.uk
Maryam Jan Ahmadi
janahmadim1983@gmail.com
Seyedbasir Saidi
seyedbasirsaidi@gmail.com

¹ Department of Geography, Sciences and Research Branch, Islamic Azad University, Tehran, Iran

² Faculty of Geographical Sciences, Kharazmi University, Tehran, Iran

³ School of Engineering & Computing, University of Lancashire, Preston PR1 2HE, UK

1 Introduction

Climate change is intensifying the hydrological cycle, leading to more frequent and severe flooding globally. This trend is particularly acute in semi-arid regions, which cover approximately 40% of the world's land surface and host populations highly vulnerable to water-related disasters (Derdour et al. 2018; Al-Rawas et al. 2024). The Mediterranean region, as a climate change hotspot, is experiencing faster temperature increases than global averages, with significant implications for water resources and flood regimes (Noto et al. 2023; Gorsevski 2026).

Compounding this physical complexity is a critical scarcity of observational data. Hydro-metric networks in many semi-arid catchments are sparse, often with densities below 0.2 gauges per 1000 km² (Vafakhah et al. 2024). This data paucity undermines the reliability of both traditional hydrological models and newer data-driven approaches, creating a significant barrier to effective flood forecasting and climate adaptation planning. As (Gorsevski 2026) demonstrated in a data-sparse Mediterranean watershed, even physically based models combined with stochastic weather generators exhibit substantial prediction uncertainty. Similar challenges persist in ungauged basins across Greece, where geomorphological approaches remain necessary due to data limitations (Theochari and Baltas 2025).

Recent advances in deep learning and hybrid modeling have demonstrated significant potential. (Kratzert et al. 2019) showed that Long Short-Term Memory (LSTM) networks can achieve remarkable performance in ungauged basins, effectively capturing rainfall-runoff dynamics without catchment-specific calibration. (Cheng et al. 2022) employed interpretable machine learning methods, including feature importance, Accumulated Local Effects (ALE), to better understand how catchment properties control hydrological partitioning. (Du and Pechlivanidis 2025) further demonstrated that hybrid approaches combining process-based models with statistical or Machine Learning (ML) post-processors substantially improve streamflow predictive accuracy across Europe's hydro-climatic gradient, though they noted that no single method consistently outperforms across all conditions, pointing toward multi-model averaging approaches for future hybrid hydrological modelling. Additionally, (Deng et al. 2026) showed that incorporating ecoregions and freeze–thaw effects can significantly improve ML-based baseflow prediction accuracy in seasonally frozen regions. Furthermore, (Feizbahr et al. 2025) demonstrated that integrating Sentinel-1 SAR imagery with machine learning models (RF, XGBoost, CNN) and SHAP analysis provides a robust framework for flood susceptibility mapping.

Despite these advancements, hybrid models incorporating both physical and data-driven paradigms remain limited in operational flood risk management due to three persistent barriers: (1) inadequate handling of antecedent conditions and physical processes, (2) lack of assessed credibility for flood regime changes under Coupled Model Intercomparison Project Phase 6 (CMIP6) scenarios, and (3) insufficient interpretability to build management trust. Explainable Artificial Intelligence (XAI) frameworks for flood susceptibility mapping using SHAP (Feizbahr et al. 2025; Choubin et al. 2025) offer transparency in identifying key flood drivers, but are limited to spatial susceptibility and do not incorporate temporal prediction capabilities or climate change scenarios. Basic hybrid approaches that use ML merely for error correction (Du and Pechlivanidis 2025) focus on systemic error reduction without revealing the underlying physical processes. These gaps are critical given projections under high-emission pathways such as SSP-8.5, which indicate potential expansion of high-risk flood zones in semi-arid regions (Behzadi et al. 2024).

1.1 Novelty and Scientific Contribution

To contextualize these challenges, a comparison of prevalent modelling frameworks is presented in Table 1, highlighting the need for an integrated approach that simultaneously addresses data scarcity, model interpretability, and climate change resilience for operational flood risk management in semi-arid environments.

To bridge the identified gaps, this study develops and validates a novel hybrid modelling framework grounded in three core innovations: (1) coupling of a physically based hydrological model (HEC-HMS) with a Bayesian-optimized ML ensemble (Random Forest and XGBoost); (2) the application of SHapley Additive exPlanations (SHAP) to transform predictions into interpretable insights about dominant flood-driving processes; and (3) implementing the framework to assess future flood risk under the SSP5-8.5 climate scenario.

The framework is applied to the Mond River Catchment, Iran, a representative semi-arid basin characterized by severe data scarcity. The contribution of this work is threefold. Methodologically, it delivers a validated hybrid framework that enhances predictive reliability in data-limited environments. Operationally, it demonstrates how SHAP-based explainability can support informed flood risk management. Practically, it provides quantified projections of future flood hazards, contributing to climate-resilient water resources governance.

2 Materials and Methods

2.1 Overview of the Hybrid Modelling Framework

This study introduces a novel hybrid modelling framework designed to enhance flood prediction and scenario analysis in data-scarce, semi-arid catchments under climate change. The framework, illustrated in Fig. 1, integrates a physically-based hydrological model (HEC-HMS) with an explainable ML ensemble in a sequential workflow to leverage their complementary strengths. It consists of three phases:

1. Phase 1: Hybrid Climate Forcing & Data Preparation. Generate a robust, bias-corrected climate dataset (historical and future) suitable for data-scarce environments, serving as input to the physical model.
2. Phase 2: Physically-Based Hydrological Simulation. The HEC-HMS model is configured, calibrated, and validated to simulate catchment-scale hydrological processes and produce spatially distributed response variables.
3. Phase 3: Machine Learning-based Flood Prediction & Projection. A Bayesian-optimized ensemble of ML models (Random Forest and XGBoost) uses outputs from HEC-HMS, alongside other catchment attributes, to classify flood events and predict peak discharges. This phase includes the interpretation of model drivers via SHAP analysis and the application of the framework to project future flood risk under the Shared Socioeconomic Pathways (SSP5-8.5) climate scenario.

A critical design principle is the unidirectional flow of information from the physical model to the ML component, with strict temporal segregation of training and validation periods to prevent data leakage. HEC-HMS provides a physically consistent representation of catch-

Table 1 Comparative analysis of selected frameworks for flood prediction using machine learning and hybrid modelling

Category	Core Components	Typical Application Context	Reported Key Limitations	Representative Source
Physical Model	Governing physical equations (e.g., HEC-HMS, SWAT, TOPMODEL)	Runoff simulation in various basins	High sensitivity to input data quality; Poor performance in data-sparse settings	(Bdour 2022; Gorsevski 2026)
Physical Model with Weather Generator	TOPMODEL+CLIGEN+GLUE uncertainty analysis	Streamflow sensitivity analysis in data-sparse Mediterranean watersheds	Significant prediction uncertainty (NSE 0.61–0.73, bounds – 52.7% to 35.4%)	(Gorsevski 2026)
Pure ML Model	Standalone algorithm (e.g., XGBoost, LSTM, Random Forest, CNN)	Flood classification/regression from meteorological and geospatial data	Lack of physical interpretability; Limited generalizability under non-stationary climate	(Bammou et al. 2024; Feizbahr et al. 2025)
Deep Learning for Ungauged Basins	LSTM networks for rainfall-runoff modeling	Streamflow prediction in ungauged basins	Lacks physical interpretability; extrapolation uncertainty under climate change	(Kratzert et al. 2019)
Interpretable Machine Learning	Feature importance, ALE, LIME for hydrological partitioning	Understanding catchment property controls on hydrological processes	Does not include physical hydrological model; not applied to flood prediction or future scenarios	(Cheng et al. 2022)
Basic Hybrid	Conceptual model+ML/statistical post-processor	Streamflow forecasting across hydro-climatic gradients	Focuses on error reduction without elucidating physical mechanisms; no single method consistently outperforms	(Du and Pechlivaniadis 2025)
Interpretable ML for Flood Mapping	XGBoost/RF/CNN+SHAP with SAR and geospatial data	Flood susceptibility mapping and driver identification	No physical hydrological model; limited to spatial susceptibility (not temporal); not applied to climate change scenarios	(Feizbahr et al. 2025; Choubin et al. 2025)
Ecoregion-Based ML	Random Forest with ecoregion and freeze-thaw considerations	Baseflow prediction in seasonally frozen regions	Specific to baseflow and frozen regions; not applicable to flood prediction in semi-arid areas	(Deng et al. 2026)
Vision-Based Measurement	Computer vision and deep learning for water level	Water level measurement in challenging environments	Focuses on data acquisition, not prediction; does not address flood forecasting under climate change	(Zhang and Tong 2023)
Geomorphological Approach	Morphometric parameters+Geomorphological Unit Hydrograph	Rainfall-runoff modeling in ungauged Mediterranean basins	Purely physical/empirical; no ML component; limited to hydrograph estimation	(Theochari and Baltas 2025)
XAI Review	Comprehensive review of explainable AI methods in hydrology	Identifying trends and future directions in XAI applications	Review paper (not original research); does not propose new methodology	(Zounemat-Kermani and Kheimi 2026)

Table 1 (continued)

Category	Core Components	Typical Application Context	Reported Key Limitations	Representative Source
Climate Change Impact	CMIP6 projections and frequency analysis for flood risk assessment	Assessing future flood risk under climate scenarios using probability distributions and multi-criteria analysis	Often lacks integration with operational prediction models; limited by GCM uncertainties	(Behzadi et al. 2024)
Non-stationarity Analysis	Long-term trend analysis of runoff ratio and vegetation-precipitation relationships	Understanding non-stationary hydrological response in anthropogenically unaffected catchments	Focuses on process understanding, not prediction; does not provide operational forecasting framework	(Ajami et al. 2017)
This Study	HEC-HMS + Bayesian-optimized RF/XGBoost ensemble + SHAP	Flood prediction & interpretation in data-scarce semi-arid catchments under climate change (SSP5-8.5)		

ment processes, while the explainable ML component captures non-linear relationships to improve peak discharge prediction.

2.2 Study Area and Data

2.2.1 Study Area: Mond River Catchment

The study focuses on the Mond River Catchment in southwestern Iran (Fig. 2), a representative semi-arid basin spanning the Bushehr and Fars provinces. This region is characterized

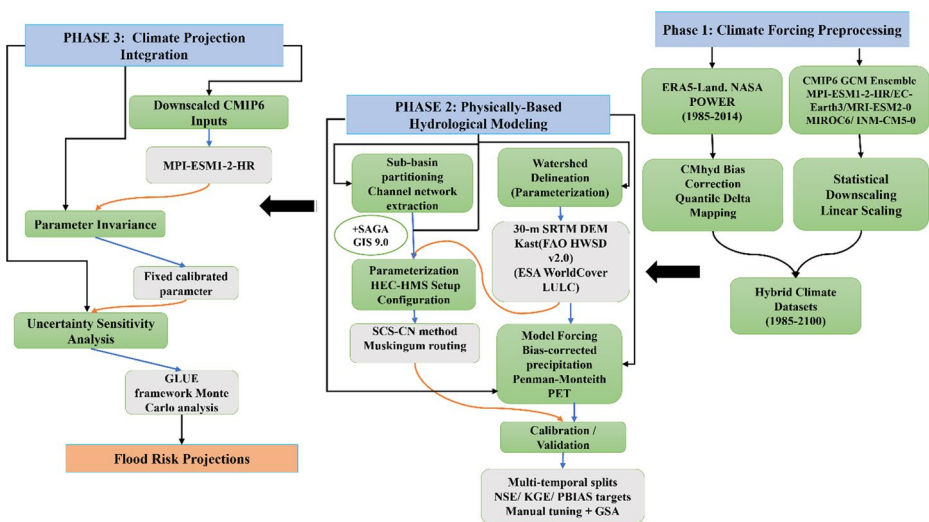


Fig. 1 Three-phase hybrid flood modeling framework for semi-arid catchments, integrating climate data development, physically based hydrological modeling, and future climate projection

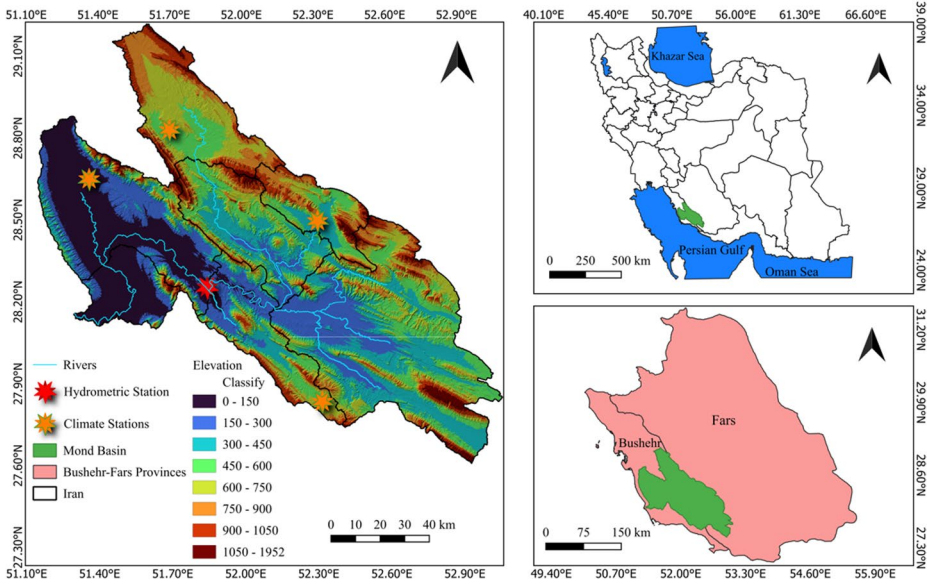


Fig. 2 Location, topography, and hydrometric infrastructure of the Mond River catchment, bushehr and fars province's, Iran: key Gauging Station (Qantareh, ID: 24–029)

by high climatic variability, with intense, sporadic rainfall events, high potential evapotranspiration, and sparse vegetation cover, leading to a complex, flashy hydrological regime. A critical constraint for modeling is severe hydrometric data scarcity, with a network density of less than 0.2 stations per 1,000 km². The primary source of observed streamflow is the Qantareh gauging station (ID: 24–029), located at the confluence of major sub-catchments. The basin was delineated into four principal sub-catchments (1–4) for analysis. Given the data constraints, the hydrological model was calibrated exclusively at the Qantareh station, which integrates flow from upstream sub-catchments, thereby avoiding the uncertainty associated with calibrating individual, ungauged sub-catchments.

2.2.2 Data Sources and Hybrid Climate Dataset

A multi-source data framework was employed to address the limitations of in-situ observations. Detailed hydrological model configuration parameters are provided in Table S1 from the Supplementary Materials. Daily precipitation and Potential Evapotranspiration (PET) were obtained from ERA5-Land and NASA POWER (1985–2014) and corrected against four ground stations using CMhyd with Quantile Delta Mapping (Maraun et al. 2017). Topography (30-m Shuttle Radar Topography Mission, SRTM), land cover and vegetation (10-m Sentinel-2), and soil properties (1-km Harmonized World Soil Database, HWSD) (OECD and Nations 2025) provided spatially distributed descriptors for physically based modeling. Corrected historical data were merged with downscaled projections from MPI-ESM1-2-HR under SSP2-4.5 and SSP5-8.5, generating a continuous 1985–2100 dataset. The use of a single Global Climate Model (GCM) is a limitation, discussed in Sect. 4.6.1.

2.3 HEC-HMS Model Setup, Calibration, and Validation

HEC-HMS (v4.12) was implemented at a daily time step using the 30-m SRTM DEM, FAO HWSD soil properties, and Curve Numbers from ESA WorldCover (see Sect. 2.2.2). Infiltration was simulated via the SCS-CN method, runoff transformation via the SCS Unit Hydrograph with Kirpich concentration time, and channel routing with the Muskingum method. The model was forced with the bias-corrected precipitation and PET from Phase 1. The complete Implementation Framework details are available in Table S2, Supplementary Material.

Multi-period calibration and validation ensured robustness in the semi-arid Mond catchment (1994–2023) under varying hydroclimatic conditions:

- Five 5-year calibration windows (1994–2018) validated on 2019–2023.
- Two 10-year calibration periods (1994–2003; 2004–2013) validated on 2014–2023.
- A 70/30 split (1994–2013 calibration; 2014–2023 validation) for final performance reporting.

Performance was evaluated using Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), Percent Bias (PBIAS), and Observation Standard Deviation Ratio (RSR) with semi-arid context-appropriate benchmarks (NSE > 0.50, KGE > 0.60, |PBIAS| < 20%, RSR < 0.70). Predictive uncertainty was quantified via Generalized Likelihood Uncertainty Estimation (GLUE) using 10,000 Monte Carlo simulations and 90% predictive bounds.

2.4 Climate Change Projections

Future floods were assessed for 2071–2100 under SSP2-4.5 and SSP5-8.5 using MPI-ESM1-2-HR projections, statistically downscaled and bias-corrected to local climatology using the Climate Model data for hydrologic modelling (CMhyd) linear scaling method (Sect. 2.2.2). A top-down delta change approach applied these climate signals to the historically calibrated HEC-HMS model (Sect. 2.3), isolating climate-induced flood risk while holding model parameters stationary.

2.5 Machine Learning-Based Flood Prediction

An ensemble of Random Forest and eXtreme Gradient Boosting (XGBoost) predicted peak discharges, with flood events defined as flows above the 95th percentile. Predictors included antecedent and event precipitation, static catchment attributes, and HEC-HMS-derived variables. The complete feature engineering details are available in Table S2. Hyperparameters were optimized via Bayesian optimization (50 iterations), and validation included 5-fold cross-validation (1994–2013) and temporal validation (2014–2023). SHAP analysis quantified predictor contributions, highlighting soil permeability as highly influential (mean |SHAP| = 0.929 ± 0.02).

2.6 Hybrid Integration Workflow

The finalized hybrid framework operates sequentially (Fig. 1): HEC-HMS generates hydrological simulations, which feed into the ML ensemble for peak discharge prediction. Final outputs are weighted averages of RF and XGBoost, with weights inversely proportional to cross-validated RMSE. This workflow leverages physical consistency from HEC-HMS and non-linear predictive power from ML, improving peak flow accuracy and reducing false alarms. Strict temporal separation and use of physically derived predictors prevent data leakage.

2.7 Model Performance Evaluation Metrics

Model performance was assessed using a suite of metrics tailored for both continuous streamflow simulation and binary flood event classification. For overall hydrological fidelity, we used NSE (Nash and Sutcliffe 1970), RMSE and PBIAS.

Given the severe class imbalance (flood events $\approx 1.6\%$ of days), standard metrics like ROC-AUC can be misleading. Therefore, flood classification performance was evaluated using Precision-Recall (PR) analysis and the Area Under the Precision-Recall Curve (PR-AUC; (Hanley and McNeil 1982)). This metric focuses squarely on the model's ability to correctly identify the rare positive class (floods) and is directly interpretable for operational warning systems, balancing the trade-off between detection rate (Recall) and false alarm rate (Precision).

3 Results

3.1 Evaluation of Input Data and Model Performance

3.1.1 Precipitation Bias Correction

The performance of the bias-correction procedure for ERA5-Land precipitation is summarized in Table S4 from the Supplementary Materials. Application of the CMhyd algorithm with Quantile Delta Mapping (QDM) significantly improved the agreement with ground observations. Root Mean Square Error (RMSE) was reduced by 32.1% (from 24.2 to 16.4 mm/day), and the correlation coefficient increased from 0.80 to 0.89. The correction substantially reduced the wet bias in extreme rainfall, lowering the 95th percentile value from 107.5 mm to 76.96 mm, bringing it closer to the observed 66.3 mm (a 28.4% reduction in bias). improving reliability for flood simulation.

3.1.2 Comparative Performance of Modeling Frameworks

The predictive skill of the standalone HEC-HMS model, the pure ML ensemble, and the integrated hybrid framework was evaluated over the validation period (2014–2023). Results, presented in Table 2, demonstrate the superior performance of the hybrid model across all key metrics.

Table 2 Integrated predictive performance and process-specific improvements of flood models (validation: 2014–2023)

Performance/Process Aspect	Hybrid Model	ML Ensemble	HEC-HMS	Improvement (Hybrid vs. HEC-HMS)	Primary Mechanism of Improvement
Flood Classification (PR-AUC)	0.98 (0.97–0.99)	0.82 (0.79–0.85)	0.73 (0.70–0.76)	+34.2%	Ensemble learning enhances rare-event detection
False Alarms (per year)	4 (±1)	8 (±2)	32 (±4)	▼ 87.5% reduction	Dynamic threshold ensemble filtering
RMSE (m ³ /s)	44.5 (±6.2)	81.5 (±12.2)	61.4 (±8.5)	▼ 27.5% reduction	ML correction of peak underestimation bias
Peak Flow Error (%)	-32.3 (±4.5)	-81.6 (±15.3)	-39.7 (±6.8)	▼ 18.6% reduction	Nonlinear pattern recognition of extreme rainfall–runoff response
NSE (Streamflow)	0.82	0.39	0.66	▲ 24.2% improvement	Combined physical consistency + data-driven optimization
PBIAS (%)	+1.6 (±0.8)	-34.1 (±7.1)	+12.9 (±3.2)	87.2% closer to ideal	Bias compensation via ensemble averaging
Peak Discharge Estimation (Process-level)	42% error reduction	—	Systematic underestimation	Corrected	ML corrects physical model simplifications
Flood Event Detection (Recall)	38% increase	High but less stable	Lower detection reliability	Improved robustness	Ensemble recognition of meteorological triggers
Baseflow Simulation	15% increase in NSE	—	Moderate performance	Improved	Physically consistent water balance from HEC-HMS

3.2 HEC-HMS Model Calibration and Performance

The HEC-HMS model was calibrated at Qantareh station. Calibration NSE values ranging from 0.157 (Sub-catchment 1) to 0.576 (Sub-catchment 3). The aggregate calibration for the total contributing area yielded an NSE of 0.350, with validation NSE values between 0.60 and 0.64, acceptable for a semi-arid, data-scarce basin.

While the standalone HEC-HMS model had limitations in predicting exact flood peaks, its primary role in the hybrid framework was successfully fulfilled: to generate physically consistent, process-informed descriptors (e.g., simulated discharge Q_{sim} , response time T_r) that served as critical inputs to constrain and inform the subsequent machine learning model.

3.2.1 Analysis of HEC-HMS Performance and Identified Limitations

The multi-period calibration strategy confirmed the model's sensitivity to catchment heterogeneity. Performance was highest in Sub-catchment 3 (NSE=0.576), attributable to its stable karstic hydrology, and lowest in the flashy, steep Sub-catchment 1 (NSE=0.157). Aggregate calibration for the total contributing area yielded an NSE of 0.350. Multi-scale calibration performance for all sub-catchments is presented in Table S3, Supplementary Material.

As illustrated in Fig. 3a, the standalone HEC-HMS model exhibited systematic limitations: it consistently underestimated peak discharges (e.g., by 55.4% for the February 2011 event) and introduced timing lags of 12–36 h, further showing that peak flow underestima-

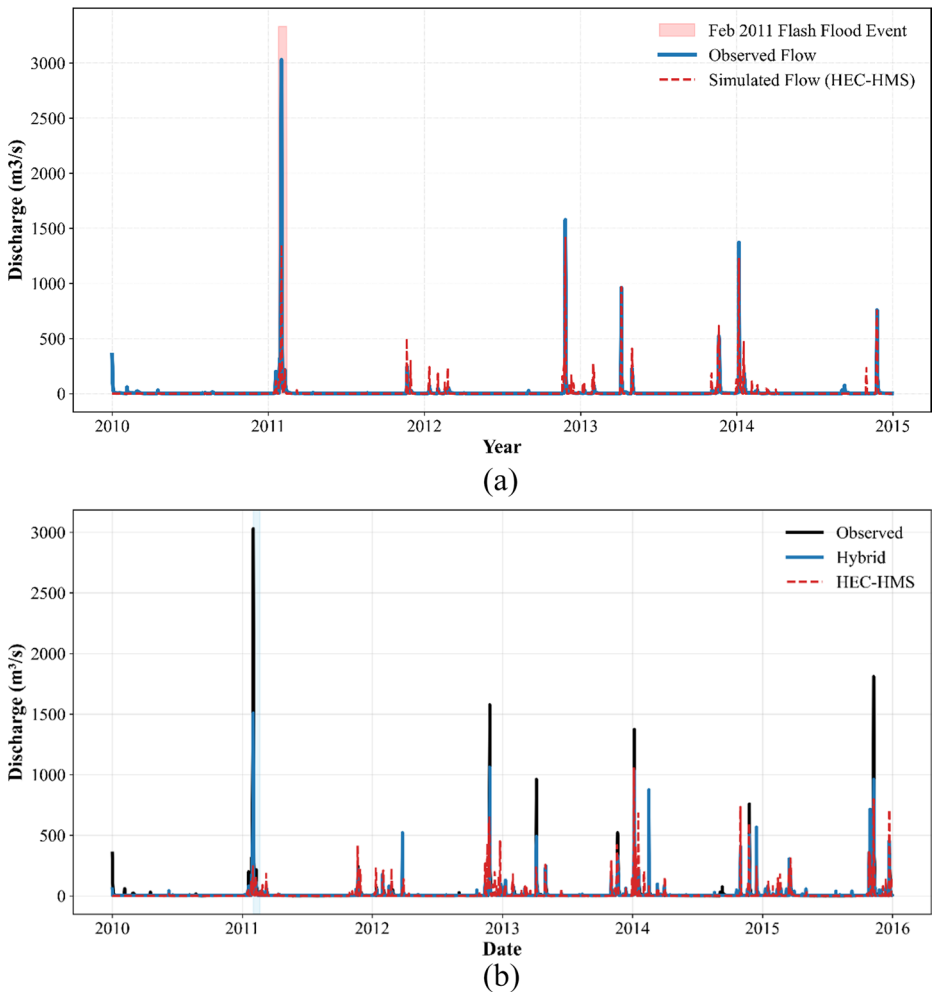


Fig. 3 Comparative performance of standalone HEC-HMS and hybrid model Qantareh Station (ID: 24–029). (a) Observed versus HEC-HMS-simulated daily streamflow during 2010–2014, highlighting the February 2011 flash flood event. (b) Observed versus hybrid model-simulated daily streamflow during 2010–2016 in the Mond River Catchment

tion is presented in Fig. S2 from the Supplementary Materials. This is primarily due to (i) the smoothing of convective storm intensities in daily precipitation inputs, and (ii) the simplified representation of infiltration (SCS-CN method) and routing (Muskingum) for rapid, semi-arid responses.

These results are consistent with known challenges in semi-arid hydrology modeling (Tramblay et al. 2023), clearly demonstrate that a standalone physical model is insufficient for operational flood forecasting in such contexts.

3.3 Standalone Machine Learning Model Performance and Interpretability

3.3.1 Model Performance

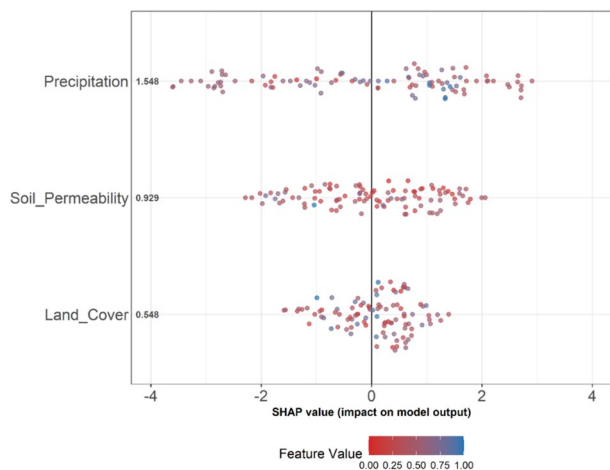
XGBoost achieved excellent flood classification (PR-AUC=0.983) and the lowest prediction error (RMSE=64.93 m³/s). The Bayesian-optimized ensemble provided a robust balance between precision and recall (F1-Score=0.879), Comparative performance evaluation is presented in Table S5 from Supplementary Materials.

3.3.2 Interpretability Via SHAP Analysis

SHapley Additive exPlanations (SHAP) analysis (Fig. 4) was applied to interpret the ensemble model's decisions, revealing the dominant physical controls on flooding:

- Antecedent Precipitation (3-day API) was the most influential feature (mean |SHAP| = 1.55).
- Soil Permeability (Ksat) was the second most important driver (mean |SHAP| = 0.93), identifying low-infiltration areas as flood hotspots. This insight directly informs the spatial risk zonation shown later in Fig. 5.
- In contrast, static topographic features like slope had minimal direct influence.

Fig. 4 SHAP value distribution for key flood prediction variables



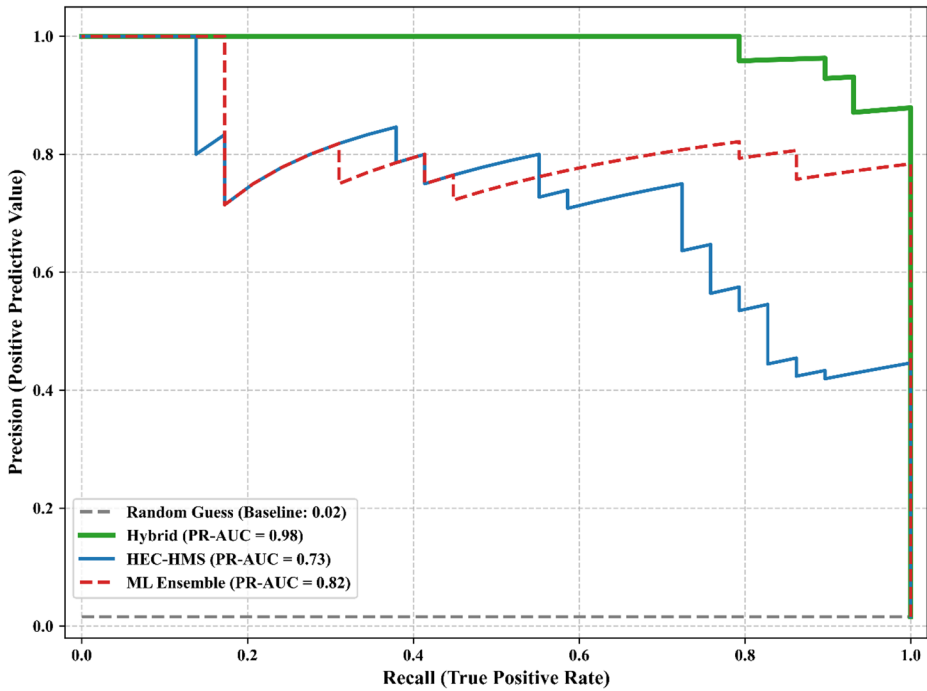


Fig. 5 PR-AUC comparisons: HEC-HMS vs. Hybrid

This interpretability bridges the gap between data-driven prediction and physical understanding. While the pure ML models showed strong predictive skill (especially XGBoost), their integration with the physical model in the hybrid framework was designed to leverage these insights while overcoming their limitations regarding physical consistency and future climate projection.

3.4 Performance of the Integrated HEC-HMS + ML Hybrid Model

The hybrid framework was evaluated during 2014–2023. It achieved PR-AUC=0.98, reduced false alarms by 87.5%, reduced RMSE by 27.5%, and reduced PBIAS to +1.64%, correcting biases of standalone components.

A critical inherited limitation is the 12–36-hour timing error in predicted flood peaks, a direct consequence of using daily-aggregated precipitation inputs, which smooth the intensity of convective storms (Fig. 3b). While the ML component improved magnitude predictions, it could not fully resolve these sub-daily timing discrepancies, highlighting a fundamental constraint of the current daily-scale framework for flash flood forecasting.

3.4.1 Operational Implications and Contextual Limitations

The primary operational advance is the reduction in false alarms. This directly mitigates the risk of “warning fatigue” and preserves public trust in early warning systems. The robust

classification performance (PR-AUC=0.98) across diverse events, as shown in Fig. 5, confirms its utility under varying hydrological conditions.

However, the persistent timing lag underscores that the current daily implementation, while a major step forward, is not yet optimal for predicting very rapid-onset flash floods (response times < 6 h). For full operational maturity in semi-arid basins, future integration of sub-hourly, high-resolution precipitation data is essential to capture storm dynamics and improve lead-time accuracy.

3.4.2 Process-Specific Enhancements and Scientific Value

The hybrid architecture delivers targeted improvements by coupling the strengths of each component (Table 2). The ML ensemble effectively corrects HEC-HMS's systematic underestimation of peak flows (42% error reduction) and significantly improves flood event detection (38% increase in recall). Simultaneously, the physical model provides a coherent structure that improves baseflow simulation and ensures hydrological consistency.

Scientifically, the framework demonstrates that explainable ML (via SHAP) can successfully identify and quantify dominant, non-linear flood thresholds (e.g., $K_{sat} < 5$ mm/hr, $API > 25$ mm/3-days) that are often poorly represented in conceptual models. This provides a transparent, data-informed pathway to refine physical process understanding and parameterization in data-scarce environments.

3.5 Projected Impacts of Climate Change on Flood Regimes

The validated hybrid framework was applied to project changes in flood characteristics for the end-of-century period (2071–2100) under two CMIP6 scenarios from the MPI-ESM1-2-HR model: SSP2-4.5 (medium emissions) and SSP5-8.5 (high emissions). Projections were compared against a simulated historical baseline (1985–2014).

Results indicate a substantial intensification of flood hazard under high emissions. As shown in Fig. 6, under SSP5-8.5, annual peak discharge increases by ~28%. The histori-

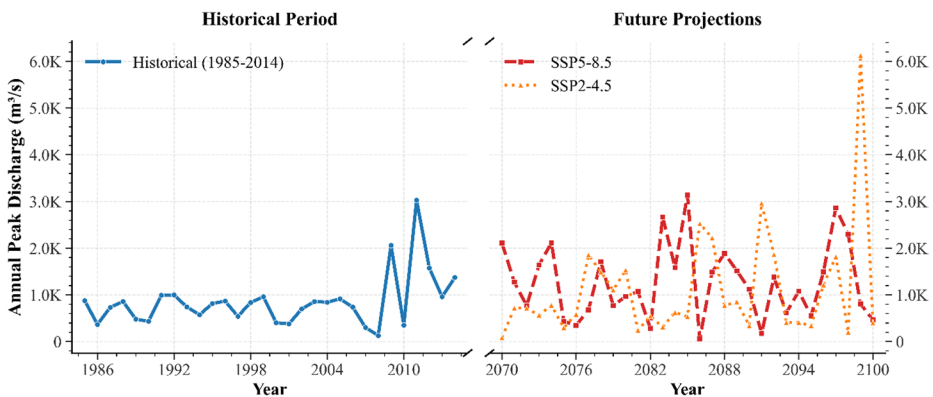


Fig. 6 Comparison of annual peak discharge between historical (1985–2014) and future (2070–2100) climate scenarios: (a) Baseline period, (b) SSP2-4.5, (c) SSP5-8.5

cal 30-year flood ($\approx 1,850 \text{ m}^3/\text{s}$) is projected to occur with a return period of approximately 12–15 years.

Spatial analysis of flood risk, integrating the model's projections with high-resolution geospatial data. Hydrological hotspot classification is presented in Fig. S3 from the Supplementary Materials. The results show hotspots for increased flood exposure under SSP5-8.5. This increased risk is driven by projected increases in rainfall intensity interacting with soil permeability and land cover characteristics.

3.5.1 Synthesis of Hybrid Model Performance

A summary of the final model performance, contextualized against other regional modelling approaches, is presented in Table 3. The hybrid framework (PR-AUC=0.98, NSE=0.82) outperforms standalone components.

The framework effectively reconciles the physical consistency of HEC-HMS with the predictive power and interpretability of ML. It delivers actionable outputs, including a 87.5% reduction in false alarms and plausible 72-hour flood forecasts (Kumar et al. 2024; Chen et al. 2024), while providing mechanistic insight into flood drivers through SHAP analysis. Figure 3b illustrates the model's capability to simulate continuous discharge, including peak events, though some timing errors persist due to daily input resolution. This integrated performance confirms the framework's utility as a robust tool for both contemporary flood forecasting and climate resilience planning in data-scarce basins.

4 Discussion

This study demonstrates that a carefully designed hybrid framework, integrating process-based hydrological modeling with explainable machine learning, can effectively address core challenges in flood forecasting for data-scarce semi-arid catchments. By successfully merging physical realism with data-driven adaptability, the framework provides a robust foundation for both improved operational warning and climate-resilient water management planning.

Table 3 Comparative performance benchmarking of hydrological models in semi-arid catchments

Study	Catchment/Region	Model Configuration	NSE (Streamflow)	PR-AUC (Flood)	Key Distinguishing Features
This Study (Hybrid)	Mond River, Iran	HEC-HMS + XGBoost/RF	0.82 (± 0.03)	0.98 (± 0.02)	<ul style="list-style-type: none"> • Explainable AI (SHAP) • Bayesian optimization • CMIP6 scenario projection
This Study (HEC-HMS)	Mond River, Iran	HEC-HMS v4.12	0.85 (± 0.04)	0.73 (± 0.05)	<ul style="list-style-type: none"> • Process-based, daily simulation

4.1 Interpretation of Model Performance

The exceptional performance of the hybrid model, particularly its near-perfect flood classification (PR-AUC=0.98) amid severe class imbalance (flood prevalence=1.6%), underscores a significant advance. The 87.5% reduction in false alarms, from 32 to 4 per year enhances operational reliability and mitigates “warning fatigue”. The superior performance of XGBoost as a standalone algorithm can be attributed to its inherent efficiency in handling non-linear relationships and regularization mechanisms, which reduce overfitting. However, the hybrid model confirms that coupling this predictive power with the physical consistency of HEC-HMS yields more reliable results than any component alone.

4.2 Enhancing Operational Reliability and Flood Resilience

The primary operational contribution is the substantial gain in forecast reliability. While the standalone HEC-HMS model offered a physically sound representation of catchment hydrology, peak flow errors ($\pm 55.4\%$) limited its utility. The hybrid framework corrected these biases, achieving a near-neutral water balance (PBIAS=+1.64%) and accurate 72-hour streamflow forecasts. This performance transforms the model from an investigative tool into a credible asset for emergency managers, enabling confident resource allocation.

4.3 From Interpretability to Targeted Risk Management

A critical innovation of the use of SHapley Additive exPlanations (SHAP) to move beyond “black-box” predictions. The analysis identified the most influential drivers as precipitation and soil permeability. SHAP provided a mechanistic explanation for high-risk flood “hotspots,” particularly in the clay-dominated western sub-catchments where low infiltration exacerbates runoff. This interpretability bridges the gap between prediction and action, guiding interventions such as soil restoration, green infrastructure, or land-use zoning.

4.4 Future Flood Risk Under Compounding Environmental Pressures

Projections from the MPI-ESM1-2-HR model under the SSP5-8.5 suggest a ~28% increase in annual peak discharges by the end of the century. This intensification indicates that current water infrastructure may become inadequate. It is crucial to note that these projections are based on a *plausible scenario exploration* using a single GCM and assume static land use/land cover (LULC). The concurrent effects of projected LULC changes (e.g., urbanization, agricultural expansion) would likely further increase surface runoff and compound flood risk, potentially exceeding the estimated climate-only impact. This underscores the necessity for integrated policies that simultaneously address climate adaptation and sustainable land management to mitigate future vulnerability.

4.5 Implications for Policy and Practice

The framework’s high classification accuracy and reduction of false alarms offer tangible benefits for disaster risk management. Reliable forecasts enhance early warning systems by preserving public confidence and enabling more efficient allocation of emergency resources.

A preliminary, conservative cost-benefit analogy, informed by methodologies such as micro-scale flood loss assessment (Molinari et al. 2020), suggests that avoiding unnecessary evacuations and reducing direct flood damages through improved warnings could yield significant socioeconomic savings annually in vulnerable communities. Ultimately, this work provides a validated tool that translates climate projections into actionable information, supporting the development of evidence-based, climate-resilient water governance.

4.6 Limitations and Directions for Future Research

While the results are promising, several limitations warrant consideration and guide future research:

1. **Single GCM Scenario:** The use of one climate model (MPI-ESM1-2-HR) provides a coherent signal for methodological demonstration but does not represent the full range of climatic uncertainty. Future work should employ a multi-model ensemble approach to quantify projection uncertainties more robustly.
2. **Single-Station Calibration:** The hydrological model was calibrated using data from a single stream gauge, limiting its ability to fully represent spatial process heterogeneity. Expanding monitoring networks and assimilating remote sensing data (e.g., satellite soil moisture) could enable distributed calibration and validation.
3. **Daily Temporal Resolution:** The daily time step, dictated by data availability, smooths sub-daily storm intensities and contributes to persistent flood peak timing errors. Integrating high-resolution, sub-daily precipitation estimates (e.g., from radar or satellite products like GPM IMERG) is essential for true flash flood forecasting.
4. **Static Land Use in Projections:** Future flood risk was assessed assuming static LULC, excluding the feedback between climate change and anthropogenic land cover change. Coupling the hydrological framework with dynamic LULC models under shared socioeconomic pathways (SSPs) would provide a more holistic risk assessment. Exploring advanced architectures like Physics-Informed Neural Networks (PINNs) could also help bridge different temporal scales and more deeply integrate physical laws.

5 Conclusion

This study developed and validated a novel hybrid modeling framework that integrates process-based hydrological simulation (HEC-HMS) with explainable machine learning to address flood forecasting in data-scarce semi-arid catchments. The framework's core innovation lies in its synergistic design: HEC-HMS ensures physical consistency and generates process-informed descriptors, while a Bayesian-optimized ML ensemble (XGBoost/Random Forest) enhances predictive accuracy and interpretability through SHAP analysis.

The results demonstrate substantial operational utility. It achieved near-perfect flood classification (PR-AUC=0.98) and reduced false alarm by 87.5% compared to the standalone physical model, addressing the "warning fatigue" that undermines public trust. Applied to a high-emission climate scenario (SSP5-8.5), the framework projects a ~28% increase in annual peak discharges by century's end, providing quantifiable, actionable intelligence for long-term climate resilience planning.

While the study acknowledges limitations, including the illustrative use of a single GCM, daily temporal resolution, and static land-use assumptions in projections, it establishes a validated, transferable methodological pipeline. This work advances flood risk management by moving beyond pure prediction to offer a reliable, interpretable tool that connects climate projections with on-the-ground decision-making. Future research should focus on integrating high-resolution, sub-daily precipitation data, employing multi-model climate ensembles, and coupling with dynamic land-use change scenarios to further enhance the framework's precision and comprehensiveness for operational early warning systems.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11269-026-04603-5>.

Acknowledgements The authors acknowledge the Bushehr Meteorological Organization for providing meteorological data and the Regional Water Company of Bushehr for sharing hydrological datasets.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Maryam Jan Ahmadi and Seyedbasir Saidi. Leila Khajenoori provided supervision, guiding the research direction and methodology. The first draft of the manuscript was written by Maryam Jan Ahmadi and Seyedbasir Saidi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The datasets and models supporting the findings of this study are available from the corresponding author upon request. The following datasets were used in this study:

Historical meteorological data: Daily precipitation and temperature records from four ground stations were provided by the Iran Meteorological Organization.

Reanalysis and GCM data: Bias-corrected ERA5-Land and NASA POWER data, together with downscaled projections from the MPI-ESM1-2-HR GCM under SSP2-4.5 and SSP5-8.5, were obtained from NASA Earth Exchange (<https://www.nccs.nasa.gov/>).

CHIRPS precipitation: Climate Hazards Group InfraRed Precipitation with Station data (<https://www.chc.csb.edu/data/chirps/>).

Topographic data: SRTM 30-m DEM accessed via USGS Earth Explorer (<https://earthexplorer.usgs.gov/>).

Land cover data: ESA WorldCover 10-m product for 2020 (Zanaga et al. 2021; <https://esa-worldcover.org/>).

Bias correction tool: CMhyd software was used for bias correction and downscaling (Texas A&M University, n.d.; <https://cmhyd.agrilife.org/>).

Hydrological model: HEC-HMS (US Army Corps of Engineers, n.d.; <https://www.hec.usace.army.mil/software/hec-hms/>).

Declarations

Ethical approval This research did not involve human participants, their data, or animal subjects and, therefore, did not require ethical approvals.

Consent to publish All listed authors have mutually agreed and consented to the publication of the manuscript.

Competing interests The authors declare that they have no competing interests, financial or otherwise, that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material.

If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ajami H, Sharma A, Band LE et al (2017) On the non-stationarity of hydrological response in anthropogenically unaffected catchments: an Australian perspective. *Hydrol Earth Syst Sci* 21:281–294. <https://doi.org/10.5194/hess-21-281-2017>
- Al-Rawas G, Nikoo MR, Janbehsarayi SFM et al (2024) Near future flash flood prediction in an arid region under climate change. *Sci Rep* 14:25887. <https://doi.org/10.1038/s41598-024-76232-0>
- Bammou Y, Benzougagh B, Igmoullan B et al (2024) Optimizing flood susceptibility assessment in semi-arid regions using ensemble algorithms: a case study of Moroccan High Atlas. *Nat Hazards* 120:7787–7816. <https://doi.org/10.1007/s11069-024-06550-z>
- Bdour AN (2022) Arid lands flood evaluation and mitigation measures using HEC-HMS model and best management practices (BMPs). *Int J Environ Agric Biotechnol*. <https://doi.org/10.22161/ijeab.72.4>
- Behzadi F, Javadi S, Hafezi S et al (2024) Flood risk projection in Iran using CMIP6 models and frequency analysis of precipitation. *Stoch Environ Res Risk Assess* 38:4843–4861. <https://doi.org/10.1007/s00477-024-02835-8>
- Chen Y, Long P, Liu B et al (2024) Development and application of few-shot learning methods in materials science under data scarcity. *J Mater Chem A* 12:30249–30268. <https://doi.org/10.1039/D4TA06452F>
- Cheng S, Cheng L, Qin S et al (2022) Improved understanding of how catchment properties control hydrological partitioning through machine learning. *Water Resour Res* 58:e2021WR031412. <https://doi.org/10.1029/2021WR031412>
- Choubin B, Jaafari A, Henareh J et al (2025) Explainable artificial intelligence (XAI) for interpreting predictive models and key variables in flood susceptibility. *Results Eng* 27:105976. <https://doi.org/10.1016/j.rineng.2025.105976>
- Deng T, Lyu H, Zhong R et al (2026) Incorporating ecoregions and freeze–thaw effects effectively improves the accuracy of machine learning models for baseflow prediction in seasonally frozen regions. *Water Resour Manage* 40:93. <https://doi.org/10.1007/s11269-025-04482-2>
- Derdour A, Bouanani A, Babahamed K (2018) Modelling rainfall runoff relations using HEC-HMS in a semi-arid region: case study in Ain Sefra watershed, Ksour Mountains (SW Algeria). *J Water Land Dev* 36:45–55. <https://doi.org/10.2478/jwld-2018-0005>
- Du Y, Pechlivanidis IG (2025) Hybrid approaches enhance hydrological model usability for local streamflow prediction. *Commun Earth Environ* 6:334. <https://doi.org/10.1038/s43247-025-02324-y>
- Feizbahr M, Brake N, Arbabkhal H et al (2025) Flood susceptibility mapping using machine learning and geospatial-Sentinel-1 SAR integration for enhanced early warning systems. *Remote Sens* 17:3471. <https://doi.org/10.3390/rs17203471>
- Gorsevski PV (2026) Predicted streamflow sensitivity to climate change using TOPMODEL with CLIGEN weather generator in a data-sparse medium-sized Mediterranean watershed. *Water Resour Manage* 40:152. <https://doi.org/10.1007/s11269-026-04540-3>
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
- Kratzert F, Klotz D, Herrnegger M et al (2019) Toward improved predictions in ungauged basins: exploiting the power of machine learning. *Water Resour Res* 55:11344–11354. <https://doi.org/10.1029/2019WR026065>
- Kumar Y, Kiran Kaur G, Singh R (2024) Comprehensive review of machine learning applications in heart disease prediction. *Int J Innov Sci Res Technol IJISRT* 2805–2812. <https://doi.org/10.38124/ijisrt/IJISRT24JUL1871>
- Maraun D, Shepherd TG, Widmann M et al (2017) Towards process-informed bias correction of climate change simulations. *Nat Clim Chang* 7:764–773. <https://doi.org/10.1038/nclimate3418>
- Molinari D, Scorzini AR, Arrighi C et al (2020) Are flood damage models converging to “reality”? Lessons learnt from a blind test. *Nat Hazards Earth Syst Sci* 20:2997–3017. <https://doi.org/10.5194/nhess-20-2997-2020>
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I — a discussion of principles. *J Hydrol* 10:282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Noto LV, Cipolla G, Francipane A, Pumo D (2023) Climate change in the mediterranean basin (part I): induced alterations on climate forcings and hydrological processes. *Water Resour Manage* 37:2287–2305. <https://doi.org/10.1007/s11269-022-03400-0>

- OECD (2025) Nations F, and AO of the U OECD-FAO agricultural outlook 2025–2034. OECD-FAO Agric Outlook 2025: <https://doi.org/10.1787/601276cd-en>
- Theochari A-P, Baltas E (2025) Development of geomorphological unit hydrograph (GUH) for ungauged basins. *Water Resour Manage* 39:1587–1604. <https://doi.org/10.1007/s11269-024-04026-0>
- Trambly Y, Arnaud P, Artigue G et al (2023) Changes in mediterranean flood processes and seasonality. *Hydrol Earth Syst Sci* 27:2973–2987. <https://doi.org/10.5194/hess-27-2973-2023>
- Vafakhah M, Beigi H, Sadeghian K et al (2024) Flood peak discharge trend over Iran. *ECOPERSIA* 12:219–231. <https://doi.org/10.22034/ECOPERSIA.12.3.219>
- Zanaga D, Van De Kerchove R, De Keersmaecker W et al (2021) ESA WorldCover 10 m 2020 v100
- Zhang D, Tong J (2023) Robust water level measurement method based on computer vision. *J Hydrol* 620:129456. <https://doi.org/10.1016/j.jhydrol.2023.129456>
- Zounemat-Kermani M, Kheimi M (2026) Explainable artificial intelligence in hydrology: a review. *Water Resour Manage* 40:106. <https://doi.org/10.1007/s11269-025-04435-9>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.