Machine learning model for stage-discharge curve calculation

Jakub Langhammer*, Miroslav Šobr, Doudou Ba

Charles University, Faculty of Science, Department of Physical Geography and Geoecology, Czechia
* Corresponding author: jakub.langhammer@natur.cuni.cz

ABSTRACT
Stage-discharge relationships (rating curves) are fundamental in hydrology but remain challenging to establish in experimental catchments, where observations are sparse, irregular, and uncertain. Conventional regression models provide simple and interpretable solutions, yet often fail to capture nonlinearities in hydraulically complex environments. Purely data-driven machine learning (ML) models offer flexibility, but their performance deteriorates under data scarcity and they often produce physically implausible results. We present a hybrid physics-informed machine learning (PIML) framework that integrates a log-log regression baseline with residual corrections from Support Vector Regression (SVR) and Multilayer Perceptron (MLP) models. By embedding hydrological constraints such as monotonicity, non-negativity, and continuity, the framework ensures physically consistent rating curves while leveraging ML to capture nonlinear deviations. The approach was developed in four contrasting catchments and validated across 20 independent evaluation sites. Results show that both hybrid models outperform conventional regression, with the Hybrid MLP consistently providing the most accurate and generalizable predictions (median R^2 and NSE > 0.98) even when calibrated with as few as 8–15 discharge measurements. The framework is particularly effective in irregular or hydraulically complex basins, while differences to conventional regression are minimal in stable profiles. These findings demonstrate that PIML enables systematic, transferable, and reproducible rating curve development under sparse and uncertain data conditions. The framework offers a practical alternative to subjective or ad hoc methods, advancing discharge estimation in experimental hydrology and supporting applications in data-limited and hydraulically complex environments.

KEYWORDS
rating curve; discharge; water level; machine learning; physics informed model

Received: 9 September 2025
Accepted: 11 November 2025
Published online: 10 December 2025

Langhammer, J., Šobr, M., Ba, D. (2025): Machine learning model for stage-discharge curve calculation.
AUC Geographica 60(2), 296–315
https://doi.org/10.14712/23361980.2025.25

1. Introduction

Stage-discharge relationships, often referred to as rating curves, are a cornerstone of open channel hydrology. These curves provide a means to estimate river discharge based on measurements of water stage, forming the basis for a wide array of hydrological analyses, including flood forecasting, water resource management, and the design of hydraulic infrastructure. Discharge ratings for gauging stations are usually determined empirically by means of discharge measurements made in the field (Kennedy 1984). Common practice is to measure the discharge of the stream periodically, usually by current meter or Acoustic Doppler Current Profiler (ADCP) system in case of flood events, and to note the concurrent stage. Measured discharge is then plotted against a concurrent stage on graph paper to define the rating curve. At a new station many discharge measurements are needed to define the stage discharge relation throughout the entire range of stage (WMO 2020).

The long-standing development of stage-discharge estimation has led to a variety of methodological approaches, from empirical regression models to physically-based hydraulic simulations. Traditional techniques for constructing rating curves have predominantly relied on simple statistical relationships, such as linear and polynomial regression. These models are favored for their computational simplicity and ease of implementation.

The development of robust stage-discharge modeling techniques is particularly important in experimental catchments. Unlike conventional gauging stations, which are typically maintained by water agencies under standardized and long-term monitoring protocols, experimental profiles are often constrained by numerous limitations. Nevertheless, they represent a critical source of information for hydrological research and water management, particularly in basins where no other data are available.

Experimental catchments are typically characterized by small drainage areas, steep gradients, dynamic flow regimes, and irregular channel morphologies, but also by challenges in data collection. Measurements are often sparse due to the remote locations, and acquired irregularly during episodic field campaigns, in different years, using various instruments and by different surveyors. All these factors contribute to observational noise and increase the uncertainty of the resulting data.

Although such data are often far from optimal quality and consistency when compared to standard hydrometric records, they are typically the only available observations for the catchments. Their scientific and practical value, particularly in understanding hydrological processes in understudied environments, remains substantial. Therefore, there is a need to identify and develop modeling approaches that are both accurate and robust, capable of coping with irregularities, limited sample sizes, and inherent measurement uncertainties. Such methods must ensure physical plausibility while allowing generalization across diverse hydrological conditions, making them suitable for applications in data-scarce, experimental settings.

Conventional regression models, such as power-law or polynomial fits, often fail to capture nonlinear behavior in irregular channels, leading practitioners to rely on ad hoc corrections (e.g., piecewise regression, correction factors, localized hydraulic modeling (Di Baldassarre and Montanari 2009; Dobrovolski et al. 2022; Kiang et al. 2018).

This niche creates an opportunity for the application of machine learning models, which are capable of learning complex, nonlinear relationships directly from data and adapting to site-specific conditions without requiring explicit physical formulations (Bhasme et al. 2022; Nearing et al. 2021; W. Xu et al. 2024). While ML models offer strong predictive power and flexibility, they face substantial challenges:

(i) ML model performance under data-sparse conditions. A significant limitation of conventional ML models is their reliance on large volumes of high-quality training data, which are typically unavailable in stage-discharge monitoring, especially in experimental catchments. For example, montane basins or experimental watersheds in remote areas frequently lack the infrastructure for long-term gauging. To address this, recent studies have explored regional modeling and data synthesis techniques (Poulinakis et al. 2023).

(ii) Lack of integration of physical laws into ML models. A persistent issue with black-box ML models is their disregard for elementary hydrological principles, such as the monotonic increase of discharge with stage or conservation of mass. Traditional models impose these physically meaningful constraints explicitly, whereas many ML models often fail to enforce them due to their purely statistical nature. This drawback has driven the development of PIML modeling approaches, which embed hydrological knowledge into learning processes (Bhasme et al. 2022; W. Xu et al. 2024).

(iii) Poor transferability across catchments. ML models are typically site-specific, limiting their applicability across regions with differing hydrological, topographic, or climatic characteristics. Although some attempts have been made to model across regions using large-sample learning (Kratzert et al. 2019b), cross-basin generalization remains an unsolved challenge.

To address these challenges, we propose a PIML framework for stage-discharge curve estimation, with the following objectives:

(i) To develop a ML-based model that incorporates physical constraints relevant to hydrological processes,

(ii) To test and validate the model across four hydrologically and topographically diverse catchments,
(iii) To benchmark the performance of the proposed model against conventional regression-based approaches.

In recent years, the emergence of PIML models has begun to reshape hydrological modeling by embedding physical principles into data-driven frameworks (Feng et al. 2023; Kratzert et al. 2019b). These approaches address key criticisms of purely statistical models by constraining predictions to obey mass conservation, monotonicity, and other hydrological laws. Applications of PIML in hydrology have so far focused primarily on streamflow prediction in large-sample datasets and ungauged basins, demonstrating improved robustness and physical plausibility compared to conventional machine learning (Bhasme et al. 2022; Esmaeilzadeh and Amirzadeh 2024). However, their use with sparse data, such as for rating curve development in experimental catchments, remains unexplored.

To address the aforementioned challenges, we developed and tested a physics-informed machine learning (PIML) model for stage-discharge estimation. The proposed framework integrates the structure of conventional rating curve models, used here as a baseline, with machine learning techniques that enhance adaptability and robustness to uncertainty. By explicitly enforcing key hydrological constraints such as monotonicity and non-negativity, the model ensures physically reliable behavior even in irregular or data-scarce conditions. We propose that physics-informed machine learning, despite its conventional requirement for large datasets, can provide a more systematic, objective, and reproducible framework as an alternative approach for extracting maximum information from sparse but high-quality measurements.

To evaluate the model performance and generalizability, we applied the model across four hydrologically and geomorphologically distinct test catchments: a steep high-gradient alpine stream in the Tien Shan Mountains, a mid-latitude montane basin in Central Europe, a low-gradient small stream in rural landscape, and a larger lowland river. These diverse test cases were selected to reflect the broad range of environments in which conventional methods often fail, providing a validation of model performance. Consequently, the model was applied to a set of 20 independent assessment sites in the mid-latitude experimental catchments to verify the model‘s applicability to real-world conditions.

2. Materials and methods

2.1 Rating curve modeling

Rating curve development is a central task in hydrometry, providing the functional relationship between river stage and discharge. This relationship may be derived empirically, from hydraulic theory, or by combining both approaches (Herschy 2019; ISO 2021).

Reliable hydrometric measurements form the basis of rating curve construction, with erroneous or inconsistent observations removed during data screening to ensure internal consistency (WMO 2010). The stage-discharge relationship is typically represented using polynomial, exponential, or piecewise functions, and dividing the curve into height zones is often necessary to capture regime shifts across flow stages (Herschy 2019; Lane 1998).

Empirical methods rely on plotting stage-discharge pairs and fitting a smooth curve, but this procedure is inherently subjective and sensitive to sparse or noisy datasets. Power-law formulations remain the standard in operational hydrometry, with parameters estimated through nonlinear regression (Braca 2008; Hrafnkelsson et al. 2022). Hydraulic methods, in contrast, derive the relationship from physical principles, for example using Manning’s equation with cross-sectional geometry, slope, and roughness (ISO 2021). Statistical regression approaches, often based on log-log transformations, reduce subjectivity and provide diagnostic measures of model fit (WMO 2010).

In practice, rating curve modeling reflects a trade-off between empirical fitting and physical reasoning. Manual and regression-based approaches can reproduce observed patterns but often lack robustness under sparse or uncertain data conditions. Hydraulic approaches ensure physical consistency but depend on detailed channel information that is not always available. Traditional regression models may also struggle in irregular natural channels, where observed discharges deviate markedly from theoretical assumptions (Ali and Maghrebi 2023; Di Baldassarre and Montanari 2009).

Recent advances in machine learning (ML) provide alternative strategies for rating curve estimation. ML algorithms are effective at capturing nonlinear relationships and handling large, complex datasets (Liu et al. 2022). However, their application in stage-discharge modeling is constrained by the data-driven nature of these methods. Direct discharge measurements are often sparse and uncertain, especially in experimental catchments, which increases the risk of overfitting and can lead to physically implausible results (Ali and Maghrebi 2023; Feng et al. 2023; Roelofs et al. 2019; Ying 2019).

2.2 Proposed hybrid physics-informed ML model

2.2.1 Model principles

To address the limitations of both traditional regression-based rating curves and purely data-driven ML approaches, we developed a hybrid PIML framework. This approach integrates a theoretical baseline model with ML-based residual corrections, thereby preserving physical consistency while capturing

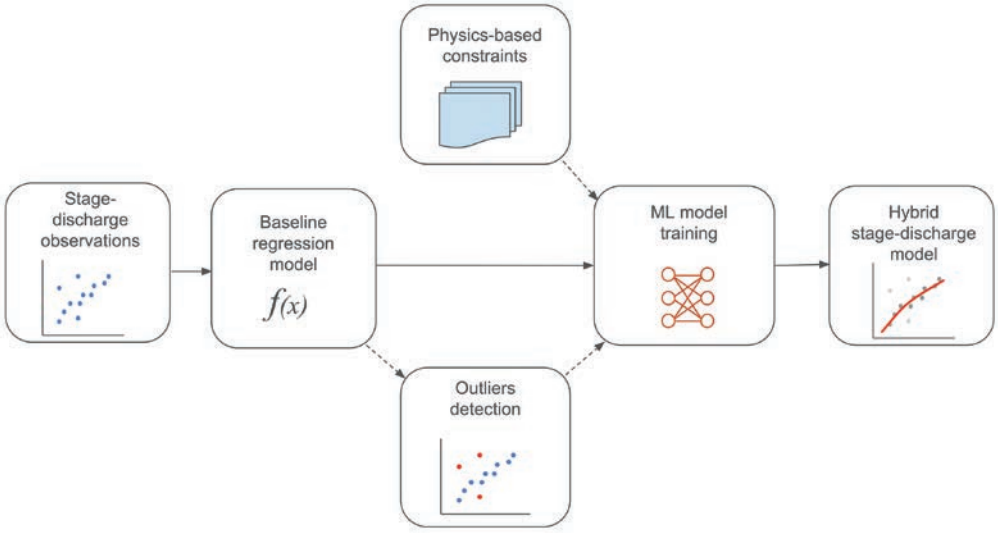

Fig 1 Principal steps in the design of the hybrid rating curve model.

nonlinear deviations (Feng et al. 2023; Raissi et al. 2019).

The framework consists of two components: (i) a log-log regression baseline that represents the primary stage-discharge relationship, and (ii) an ML residual model trained on deviations between observed and baseline predictions. By embedding physical constraints – non-negativity, monotonicity, continuity, and bounded parameter ranges – the approach ensures hydrological realism while allowing ML to refine predictions. Final discharges are obtained by combining baseline predictions with corrected residuals, resulting in smooth rating curves suitable for operational use (Fig. 1).

ML algorithm is then used for fitting the model to the real-world distribution, while being trained on the residuals between the baseline model and the observed values. The ML residual model is trained on the deviations between observed discharges and baseline predictions. To ensure physically consistent behavior, three constraints are enforced: (i) discharge remains non-negative, (ii) discharge increases monotonically with stage, and (iii) rating curves are continuous and site-specific, and (iv) basic shape of the stage-discharge curve is defined by a theoretical distribution. These preserve hydrological realism while allowing the ML model to capture nonlinear deviations. Final predictions are generated by combining the baseline log-log regression model outputs with the machine learning residual predictions.

2.2.2 Baseline model

For the selection of baseline regression models, we considered approaches commonly used in stage-discharge curve reconstruction, namely power law, polynomial regression, and log-log regression (Herschy 2019; WMO 2010). The log-log regression form has been shown to be comparatively more robust under sparse and noisy data conditions, owing to its stabilizing transformation of both variables (Kiang et al. 2018; Le Coz et al. 2014). Baseline model was thus established through a log-log regression model that captures the primary relationship between water levels and discharge measurements. This baseline model follows the form:

$$Q = exp\ exp\ (b) \times (H - H_0)^a,$$

where Q is discharge, H stage, H_0 a reference stage, and a, b, fitted parameters.

Parameters are optimized using non-linear least squares with bounds $a \in [0.1, 5.0]$ and the log coefficient bounded $b \in [-10, 10]$, and H_0 set to 90% of the minimum observed stage. These constraints prevent unrealistic scaling while ensuring positive effective depths. The baseline captures fundamental hydraulic behavior and provides the physical structure for ML refinement.

This baseline model captures the primary hydraulic behavior and serves as the physical constraint for subsequent machine learning corrections, ensuring that any data-driven improvements operate within a physically consistent framework rather than attempting to learn fundamental hydraulic principles solely from observations.

To extend the single-segment log-log baseline, a piecewise power-law model was implemented to represent gradual changes in the slope of the stage-discharge relationship.

The model divides the relationship into several monotonic segments, each described by its own power-law parameters, with continuity enforced at the breakpoints. Model parameters and breakpoint locations were estimated by nonlinear least-squares fitting in log-transformed space. A configuration of three connected segments provided adequate flexibility

while avoiding overfitting, and the model served as both an independent benchmark and a reference baseline for the hybrid machine-learning models.

2.2.3 Physics constraints

Physical realism of the model is maintained by enforcing key constraints.

First, monotonicity is enforced to guarantee that the stage-discharge relationship is non-decreasing. This is achieved via sorting and validation of the input data, followed by a post-processing correction with a tolerance of 0.01 m^3/s, leading to curve smoothness. Non-negativity is ensured by removing non-positive inputs and constraining predictions to a minimum discharge of 0.001 m^3/s. Continuity is achieved by interpolating 200 equally spaced points across the observed stage range, preventing discontinuities in curve generation.

The energy consistency conservation principle is preserved by embedding the log-log power-law formulation, which reflects energy conservation in open-channel flow and is consistent with Manning's equation and open channel hydraulic principles.

2.2.4 Residual modeling with machine learning

Eight ML algorithms were initially tested for residual modeling, representing a range of different methodological families used in hydrological applications (Kratzert 2019a; Mosavi et al. 2018; T. Xu and Liang 2021). Specifically, the Artificial Neural Networks (ANN), Multi-Layer Perceptron (MLP), Support Vector Regression (SVR), Random Forest (RF), K-Nearest Neighbors (KNN), Gaussian Process (GP), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost) were tested.

Performance was evaluated using R^2 and RMSE, complemented by qualitative assessment of physical plausibility, since some algorithms may produce non-monotonic curves under limited training data (Nearing et al. 2021; Shortridge et al. 2016).

The evaluation was based on observed stage-discharge data from primary test catchments, ensuring that algorithm performance was tested under data-poor conditions typical of experimental basins with variable physiographies. This approach enabled us to assess both statistical performance and physical plausibility, which guided the selection of algorithms for the PIML framework. As a result, two machine learning models, Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP) were selected for implementation in the hybrid model.

To address the systematic deviations from the baseline predictions, ML models were trained on the residuals between observed values and baseline predictions. To ensure physical consistency and non-negative discharge predictions, the residuals were log-transformed during the training process. Outlier removal was performed using the Interquartile Range (IQR) method to enhance model stability.

Two complementary machine learning approaches, selected from ML model testing, were implemented to provide robust evaluation of the physics-informed framework: Support Vector Regression (SVR) with Radial Basis Function kernel and Multi-Layer Perceptron (MLP) neural networks. This dual approach enables comprehensive assessment of different algorithmic families, allowing the data to determine optimal performance rather than assuming superiority of any particular method. SVR represents kernel-based learning with strong theoretical foundations for small datasets, while MLP represents neural network approaches with proven capability for complex nonlinear relationships.

2.2.5 Data preprocessing and model integration

Preprocessing removed missing or non-physical values, ensured correct ordering of stage-discharge pairs, and identified outliers using the interquartile range method. Residuals were log-transformed with offsets to handle negative values, then back-transformed for final predictions.

Hybrid predictions were obtained as

$$Q_{\mathrm{PIML}} = Q_{\mathrm{baseline}} + \mathrm{Residuals}_{ML},$$

ensuring that baseline hydraulic structure was preserved while allowing ML to capture systematic deviations.

Continuous rating curves were generated using 200 equally spaced stage values from the minimum observed stage to 110% of the maximum, with monotonicity corrections applied selectively to maintain smoothness. Post-processing includes gentle monotonicity enforcement that selectively corrects only when $Q(i) < Q(I - 1) - 0.01$, preserving natural curve smoothness while maintaining hydraulic principles. Final predictions are converted back through exponential transformation and offset removal, ensuring that the original residual scale is preserved while maintaining numerical stability throughout the machine learning process.

2.2.6 Model performance evaluation

Model performance was assessed using Coefficient of Determination (R^2), Nash-Sutcliffe Efficiency (NSE), and Index of Agreement (IoA). Comparisons between baseline and hybrid models quantified predictive improvement, while cross-validation across catchments demonstrated robustness under varying physiographic and data conditions.

The computational workflow processes each monitoring station individually through sequential data loading and validation, model fitting, prediction generation, and performance assessment. Error handling safeguarded against insufficient datasets (<10 points), non-convergence, and numerical instability, ensuring reliable model performance and consistent results across diverse sites.

Model evaluation was restricted to the observed stage range; no extrapolation beyond the highest measured stage was performed. The predictive uncertainty of the hybrid rating-curve models was quantified using a residual-based, non-parametric method adapted from Kiang et al. (2018). Residuals in log-discharge space were evaluated along the stage axis, and for each stage, the k nearest observations were used to derive 5th–95th percentile bounds of the residual distribution.

These local quantiles were smoothed and applied to the fitted curves to form 95% prediction intervals, common in hydrometry (ISO 2020), and providing a transparent, data-driven estimate of discharge uncertainty for models without native uncertainty propagation.

2.3 Study sites and data

The modeling framework was first developed and evaluated in four contrasting test catchments, representing diverse physiographic settings and hydrological regimes: a high mountain basin, a montane headwater catchment, a hilly agricultural stream, and a lowland river section (Fig. 2). To assess robustness and transferability, the approach was subsequently applied to 20 additional evaluation catchments, independent from model development, located in the

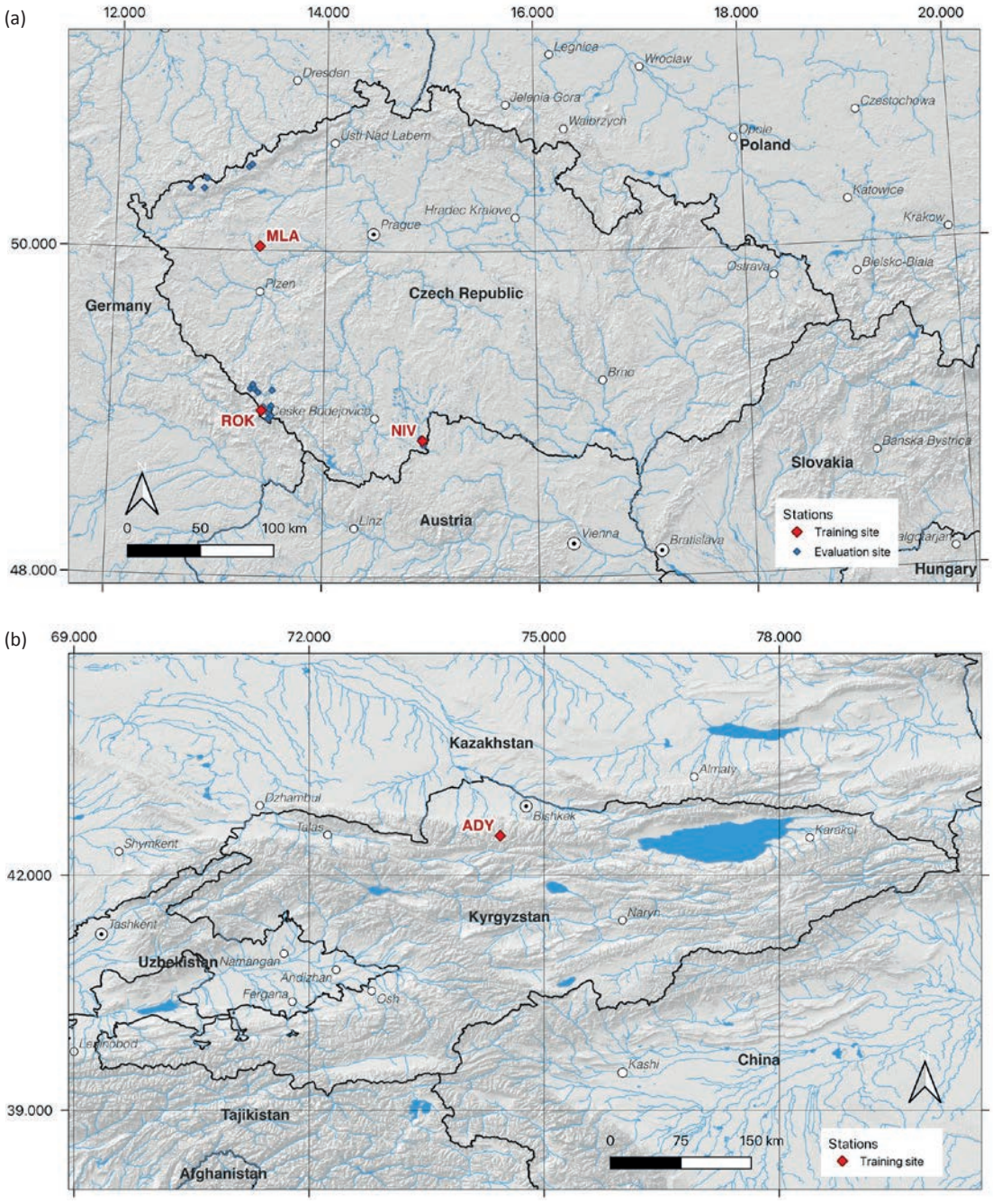

Fig. 2 Location of study sites. (a) Stations in the Czech Republic and (b) Kyrgyzstan.

Šumava, Krušné hory, and Krkonoše mountain ranges as well as in rural midland and lowland basins of Czechia.

The primary test sites encompass substantial variability in natural conditions and hydrological dynamics (Tab. 1). The Adygine catchment (ADY), situated in the Tien Shan Mountains, Kyrgyzstan, represents a high-alpine setting with a gauging station at the outlet of Lake Adygine (Fig. 3a). Its glacial regime produces pronounced daily flow fluctuations. The Rokytka catchment (ROK), located in the Šumava Mountains, Czechia, is characterized by a rain-snow runoff regime with peak flows in spring. The gauging station is positioned downstream of a tunnel through the dam of a former reservoir (Fig. 3b). The Lužnice River site (LUZ), in the Třeboň basin lowlands of Czechia, differs from the others by scale: the gauging station is situated on the middle river reach within a wide floodplain, with a deep channel and very low velocities even at high flows (Fig. 3c). The Mladotický Creek site (MLA), in the hilly agricultural landscape of Western Bohemia, Czechia, is characterized by a rain-snow runoff regime and an intensively farmed basin. The station is installed on a stable concrete bridge (Fig. 3d).

Stage-discharge rating curves at these sites were constructed from direct hydrometric measurements. Data were collected using a SonTek RiverSurveyor ADCP, a SonTek FlowTracker velocimeter, and Ott C2 and C31 propeller meters, following ISO 748 guidelines for velocity measurements. The resulting discharge data have an estimated uncertainty of 2–5%. All gauging stations are located at stable river sections, with repeated measurements confirming rating curve stability over time.

Between 8 and 32 discharge measurements were available per station (Tab. 1). The relatively small sample sizes reflect typical conditions in experimental catchments, while also highlighting the challenges of reliably reconstructing rating curves from limited datasets.

2.4 Hybrid model implementation

Initial testing of baseline regression models indicated that the log-log formulation provided the most suitable foundation for the hybrid approach. Outliers were identified using the interquartile range method and excluded prior to model fitting. Eight machine learning algorithms were evaluated for residual modeling, of which Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) demonstrated the best performance. Both were trained on the residuals between observed discharges and baseline predictions, thereby capturing systematic deviations from the theoretical relationship.

The SVR implementation employed a radial basis function kernel with regularization parameter C = 10, gamma set to “scale” for automatic variance-based scaling, and epsilon = 0.5 to control tolerance around the support vectors. These settings were chosen to balance generalization and curve smoothness, a requirement for hydraulic applications. The MLP architecture comprised two hidden layers with 50 and 25 neurons, respectively, and used hyperbolic tangent activation functions to produce smooth, bounded outputs. Optimization was performed with the Limited-memory BFGS solver, supported by L2 regularization ($\alpha = 0.01$) to reduce overfitting. The

Fig. 3 Variability of physical properties of gauging stations. (a) Outlet from the Adygine lake.Water drains away via a stable bedrock ridge (riegel), (b) Gauging station of the Rokytka River at the time of maximum water level during the flood situation on December 1, 2015, (c) Lužnice gauging station site at the time of maximum measured water level, (d) Gauging station of the Mladotický Creek during period of increased water level.

Tab. 1 Basic characteristics of stations in experimental catchments used in the study. Data: Charles University.

Station code	Stream	Station name / location	Elevation [m a.s.l.]	Catchment area [sq. km]	Station type*	Number of observations
ADY	Adygine	Adygine	3400	2.8	T	8
ANT	Antýgl	Antýgl	930	1.6	E	9
BRE	Březnický p.	Březník	1140	3.4	E	14
BYS	Bystřice	Abertamy	855	11.1	E	15
CER	Černý	Nová Hůrka	910	1.5	E	13
CIK	Cikánský p.	*catchment outlet*	1055	2.2	E	13
CHH	Chomutovka	Hora	806	8.7	E	17
CHT	Chomutovka	Tišina	650	21.9	E	11
JAV	Javoří p.	*catchment outlet*	1035	14.2	E	20
LOS	Losenice	Rejštejn	570	53.9	E	13
MLA	Mladotický p.	Přehořov	415	34.2	T	10
MOD	Modravský p.	Modrava	991	42.1	E	14
NIV	Lužnice	Niva	450	935.0	T	15
POP	Lužnice	Popelnice	720	1.8	E	8
PTA	Ptačí p.	Ptačí nádrž	1130	5.5	E	30
RKL	Roklanský p.	Modrava	990	47.6	E	32
ROK	Rokytka	Rokytecká nádrž	1090	3.9	T	15
SCH	Lužnice	Suchdol	454	955.0	E	15
SLA	Slatinný p.	*catchment outlet*	850	27.7	E	9
SLP	Slatinný p.	Nové Hamry	755	17.8	E	11
VES	Lužnice	Nová Ves	475	917.0	E	17
ZLA	Zlatý p.	Zlatý kopec	770	5.9	E	12

* Station types: T – Primary test site, E – Independent evaluation site.

maximum number of iterations was set to 1000, and a fixed random state (42) ensured reproducibility.

Model performance was evaluated using multiple criteria, including Mean Squared Error (MSE), Nash-Sutcliffe Efficiency (NSE), coefficient of determination (R^2), and Index of Agreement (IoA), computed for both the baseline and hybrid models.

The hybrid model implementation was developed in Python, relying primarily on the Scikit-learn library for ML modeling. Statistical procedures and baseline regression fitting were carried out with SciPy, including non-linear least squares optimization and continuous curve generation through interpolation routines. Additional preprocessing and smoothing were supported by the scipy.ndimage and scipy.signal modules. Performance evaluation was conducted with HydroEval, while visualization was performed using Matplotlib and Seaborn.

3. Results

3.1 Baseline regression models and ML algorithms

3.1.1 Baseline regression models

Based on theoretical assumptions, three regression models were compared to test the suitability for rating curve determination: log-log, polynomial, and power law (Fig. 4).

All three models exhibited a very high level of fit, with coefficients of determination (R^2) exceeding 0.995. However, despite these strong statistical metrics, some models produced physically implausible behavior. In particular, the polynomial regression curve displayed unrealistic bending in the low-flow region (Fig. 4c), which may yield ambiguous discharge estimates. This highlighted a key limitation: performance metrics alone may mask physical inconsistencies and should therefore not serve as the sole criterion for model selection (McMillan and Westerberg 2015).

The log-log model demonstrated greater robustness, largely due to the logarithmic transformation of both variables. This transformation stabilizes the regression across the full range of observed flows, enhancing reliability even in small, irregularly shaped mountain streams where low-flow conditions are often subject to high measurement uncertainty due to channel morphology. Nevertheless, log-log models also have limitations in steep or morphologically complex channels, where step-pool sequences, backwater effects, or local controls may disrupt the theoretical monotonic relationship (Herschy 2019; Kiang et al. 2018).

3.1.2 Machine learning models

The ML algorithms tested for potential integration into the PIML model revealed two key findings. First, most models, including Artificial Neural Network (ANN), Multi-Layer Perceptron (MLP), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors (KNN), Gaussian Process (GP), and XGBoost (XGB), achieved high values of statistical performance metrics. However, as the ML models are trained on a limited set of observations, most of them resulted in physically unrealistic rating curves. The high fit achieved by the ML models, with most algorithms reaching R^2 values above 0.96 and low RMSE values, was largely the result of mechanistic fitting to the data rather than capturing the underlying principles governing the stage-discharge relationship. This led to overfitting and unrealistic model behavior, including sharp discontinuities, non-monotonic segments, and excessive sensitivity to sparse data points, all of which are inconsistent with physically plausible rating curves (Mosavi et al. 2018).

This behavior is apparent in the example from the ROK profile, which contains observations spanning the full range of stage values, with a high density of measurements in the mid-range, a distribution typical for experimental basins (Fig. 5). GB, XGB, and GP yielded near-perfect metrics; however, the exceptionally high fits with R^2 values approaching 1.0 are more indicative of overfitting than of true predictive skill (Nearing et al. 2021; Shortridge et al. 2016). RF (R^2 = 0.940) produced the characteristic stepwise predictions typical of ensemble tree-based approaches. KNN, with a similar stepwise fit, was the weakest performer (R^2 = 0.598), failing to reproduce even the basic functional form of the stage-discharge relationship. ANN achieved a balanced performance (R^2 = 0.989) with a realistic curve shape. Based on the combination of performance metrics and smooth, physically relevant curves, MLP (R^2 = 0.997) and SVR (R^2 = 0.995) were among the best-performing models. Importantly, the smooth fits achieved by ANN, MLP, and SVR despite the low number of observations were

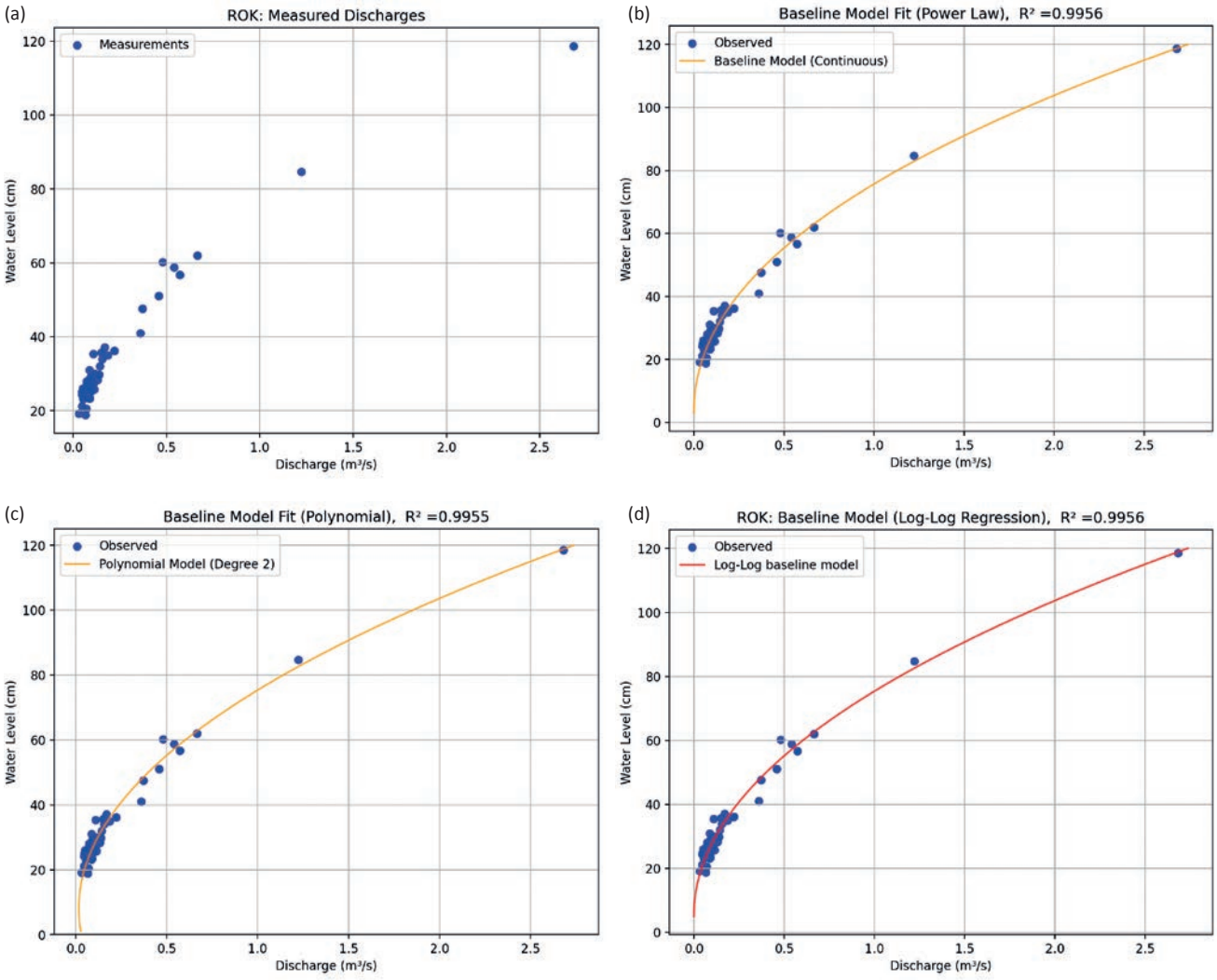

Fig. 4 Stage-discharge reconstruction using conventional regression models for ROK. a) Measured discharges, b) Power Law fit, c) Polynomial fit, d) Log-log fit.

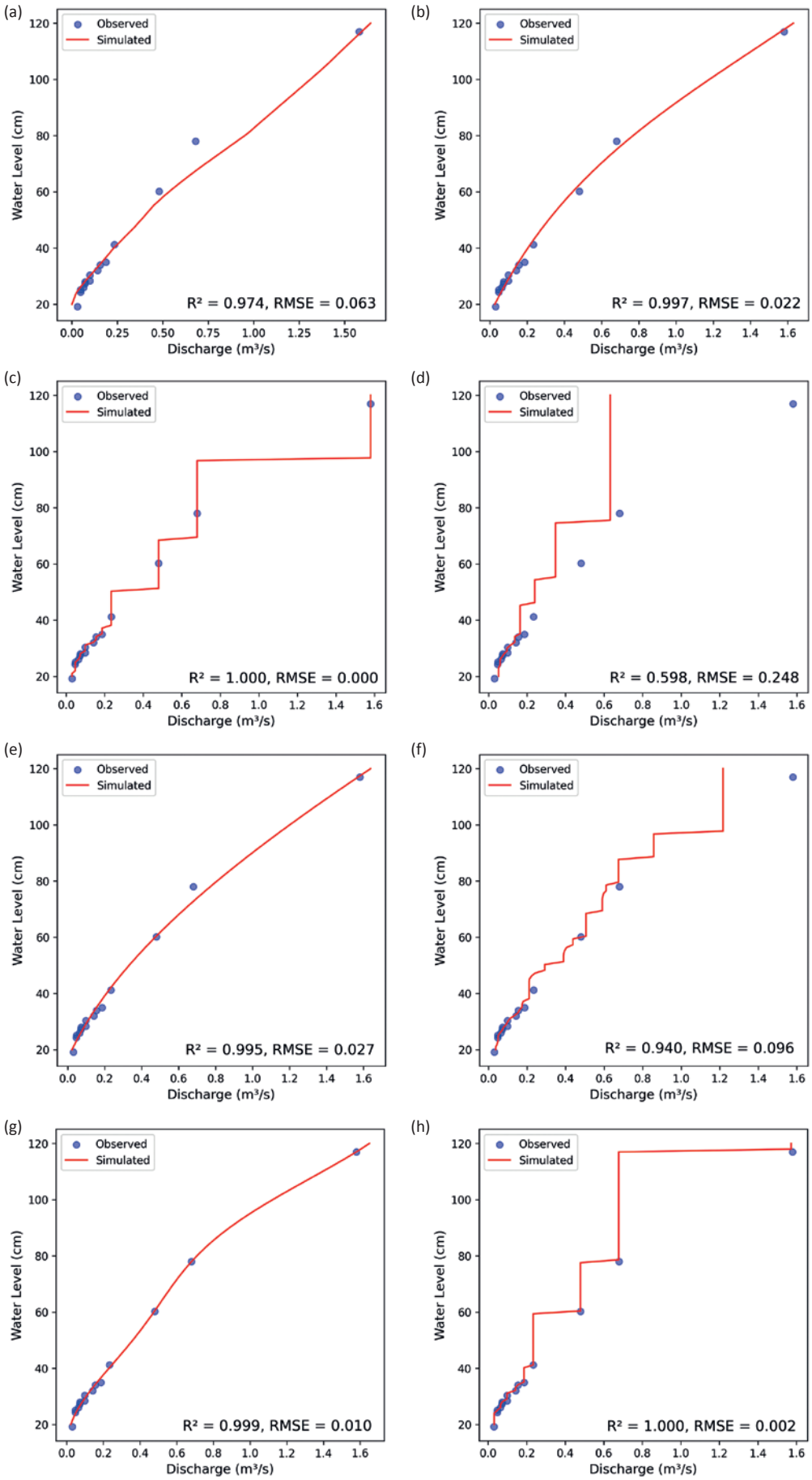

Fig. 5 Stage-discharge reconstruction using ML models for ROK basin. a) Artificial Neural Network, b) Multi-Layer Perceptron, c) Gradient Boosting, d) KNN, e) Support Vector Regression, f) Random Forest, g) Gaussian Process, h) XGBoost.

possible only after careful optimization of model settings and hyperparameters.

MLP and SVR models, representing neural network and kernel-based approaches, were therefore selected for integration into a hybrid modeling framework as representative robust ML methods, combining empirical flexibility with physical plausibility.

3.2 Hybrid model performance in test catchments

The hybrid models demonstrated close agreement with observed stage-discharge relationships, particularly at lower flows where the conventional log-log transformation often underperforms (Fig. 6). Hybrid MLP model was the most consistent performer, producing rating curves closely aligned with observations and maintaining physical plausibility (Fig. 6). Hybrid SVR model achieved similarly strong results but was slightly less stable in extrapolation ranges. Both substantially improved on baseline regression, particularly at low and high flows.

In the ADY basin (n = 8, Fig. 6a), hybrid models outperformed the baseline at higher discharges, with MLP achieving the closest fit (R^2 = 0.983 vs. baseline R^2 = 0.976). In ROK (n = 14) and MLA (n = 10, Fig. 6b–c), all models performed nearly identically, reflecting hydraulically stable conditions (R^2 > 0.996). The NIV basin (n = 12, Fig. 6d) showed the largest improvement from hybridization: the baseline underestimated medium and high flows (R^2 = 0.897), while both hybrid models markedly reduced this deviation, with MLP performing best (R^2 = 0.996).

Model testing in contrasting environments showed that the hybrid framework improved the ability to capture nonlinearities in the rating curves, particularly in the transition from low to high flow regimes. Both SVR and MLP improved predictive accuracy, with MLP consistently achieving the highest R^2 values. The effect of the hybrid models was minimal in hydraulically simple channels (ROK, MLA), but pronounced in complex environments represented by a highly dynamic high mountain stream (ADY) or a large lowland basin (NIV).

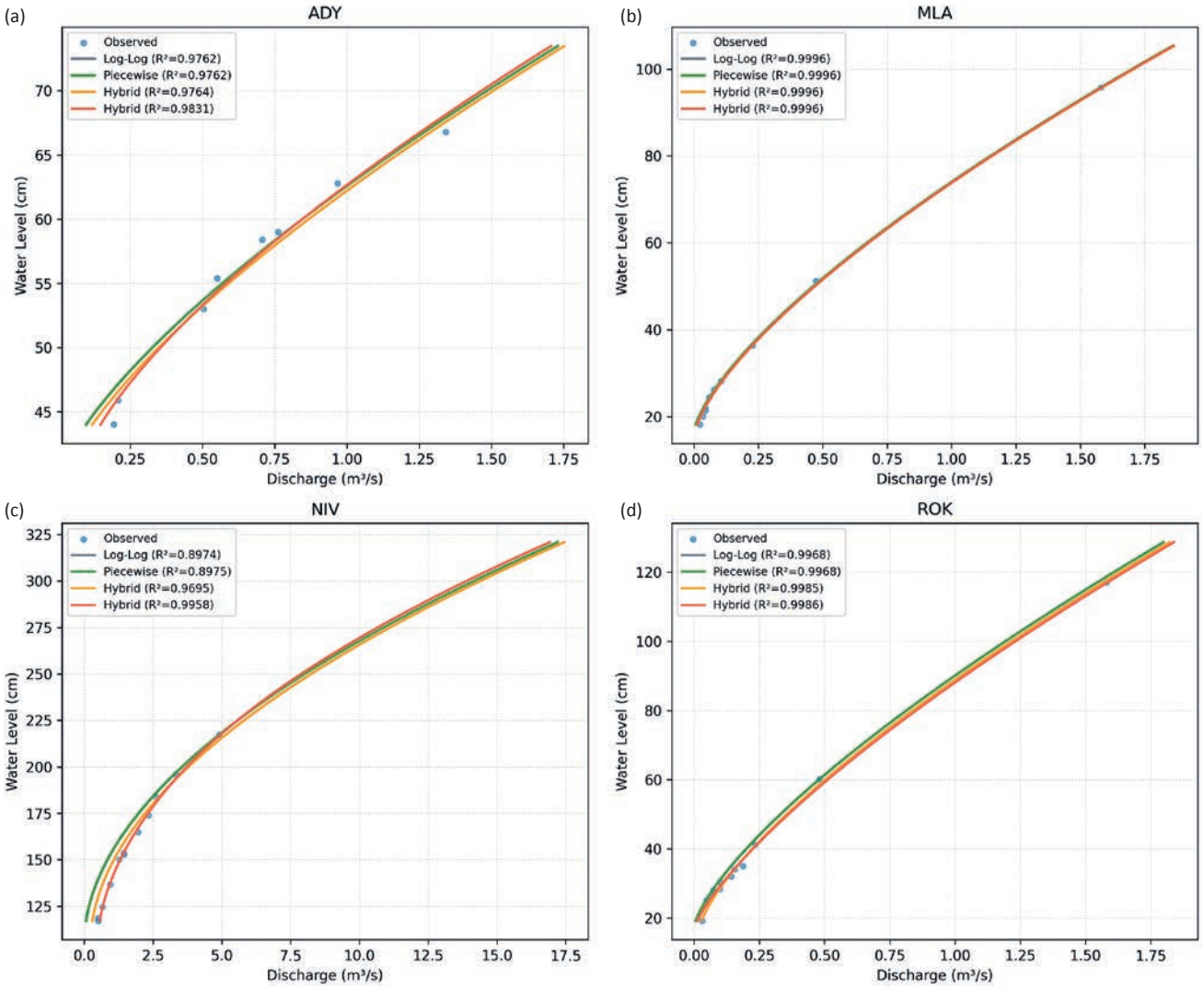

Fig. 6 Observed stage-discharge relationships and model predictions from the baseline log-log regression, hybrid SVR, and hybrid MLP models across the four experimental basins: (a) ADY (n = 8), (b) ROK (n = 14), (c) MLA (n = 10), and (d) NIV (n = 12).

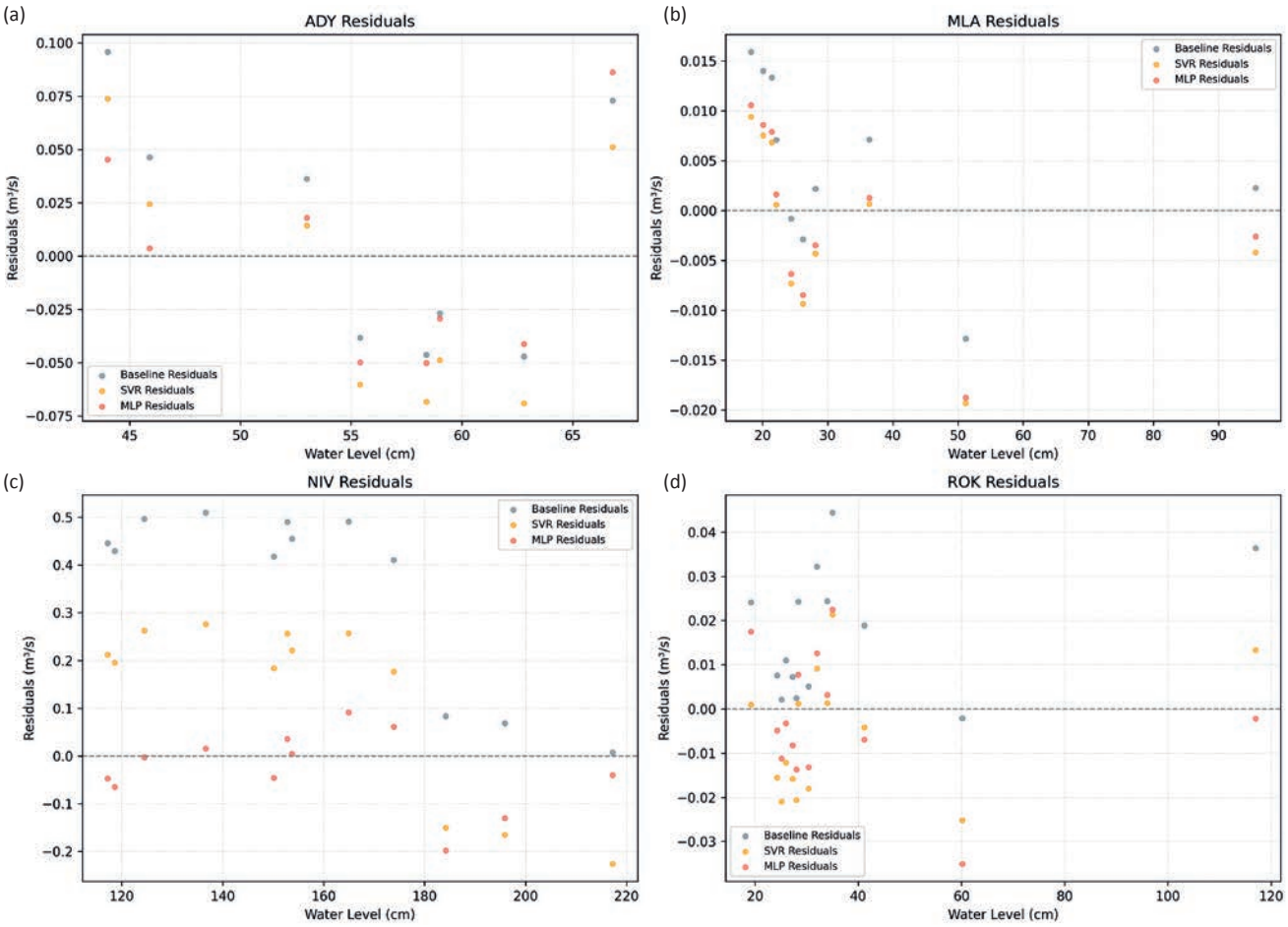

Fig. 7 Residuals from the baseline, SVR and MLP models (a) ROK, (b) JAV, (c) CIK, (d) PTA.

Model performance metrics (R^2, NSE, IoA; Table 2) confirm the effect of hybrid models on the improvements of the fit. MLP achieved the highest predictive skill (mean R^2 = 0.9853, mean IoA = 0.9986), followed closely by SVR (mean R^2 = 0.9825, mean IoA = 0.9968), while the baseline log-log model showed lower accuracy (mean R^2 = 0.9564, mean IoA = 0.9913). Basin-specific patterns are consistent: hybrid models were nearly indistinguishable in MLA and ROK but produced the largest gains in NIV, and moderately improved performance in ADY. These results demonstrate that hybrid models are particularly effective in basins where the stage-discharge relationship departs from simple log-log behavior, while in hydraulically stable basins the differences between models are less pronounced.

Residual plots (Fig. 7) reveal patterns consistent with these findings. In ADY, the baseline model exhibited systematic positive residuals at low flows and negative residuals at high flows, which were mitigated by both hybrid models. ROK and MLA showed compact residual distributions with minor differences among models, while NIV exhibited the greatest residual spread; hybrid models, particularly MLP, substantially reduced residual magnitudes at high flows.

The residual plots (Fig. 7) indicate some degree of heteroscedasticity across all basins, with residual variance appearing to vary with water level. This pattern is most evident in the NIV basin, where residual scatter increases at higher water levels, though the hybrid models help mitigate this effect.

Tab. 2 Model performance metrics by basin and model type, calculated using R2, NSE, and IoA metrics.

Basin	ADY			ROK			MLA			NIV		
Model	log-log	SVR	MLP	log-log	SVR	MLP	log-log	SVR	MLP	log-log	SVR	MLP
R^2	0.9762	0.9764	0.9831	0.9968	0.9985	0.9986	0.9996	0.9996	0.9997	0.8974	0.9695	0.9958
NSE	0.9762	0.9764	0.9831	0.9968	0.9985	0.9986	0.9996	0.9996	0.9997	0.8974	0.9695	0.9958
IoA	0.9943	0.9943	0.9957	0.9992	0.9996	0.9997	0.9999	0.9999	0.9999	0.9777	0.9933	0.9990

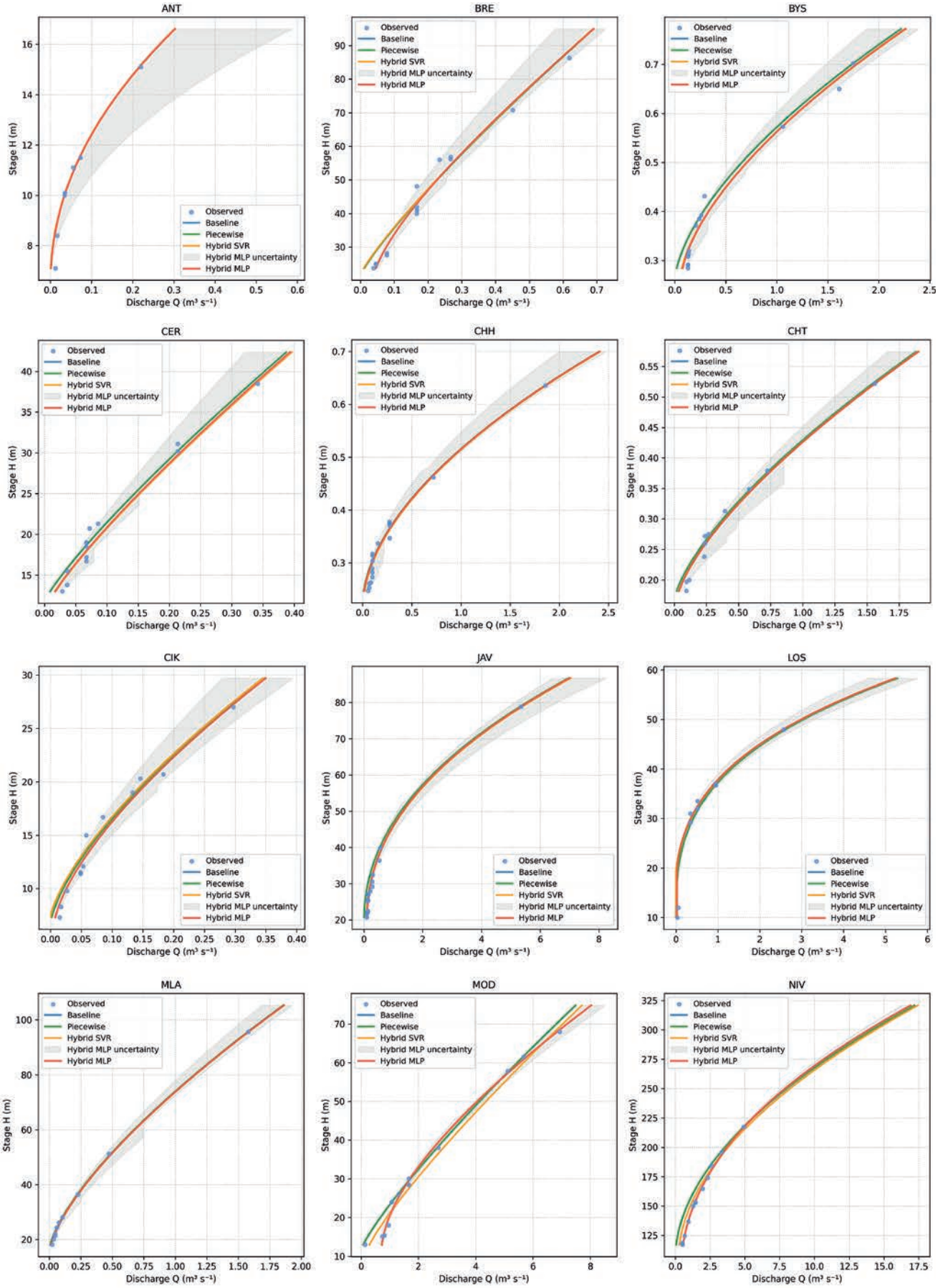

Fig. 8 Observed stage-discharge data and rating curves derived from the log-log baseline, Hybrid SVR, and Hybrid MLP models for 20 independent evaluation catchments.

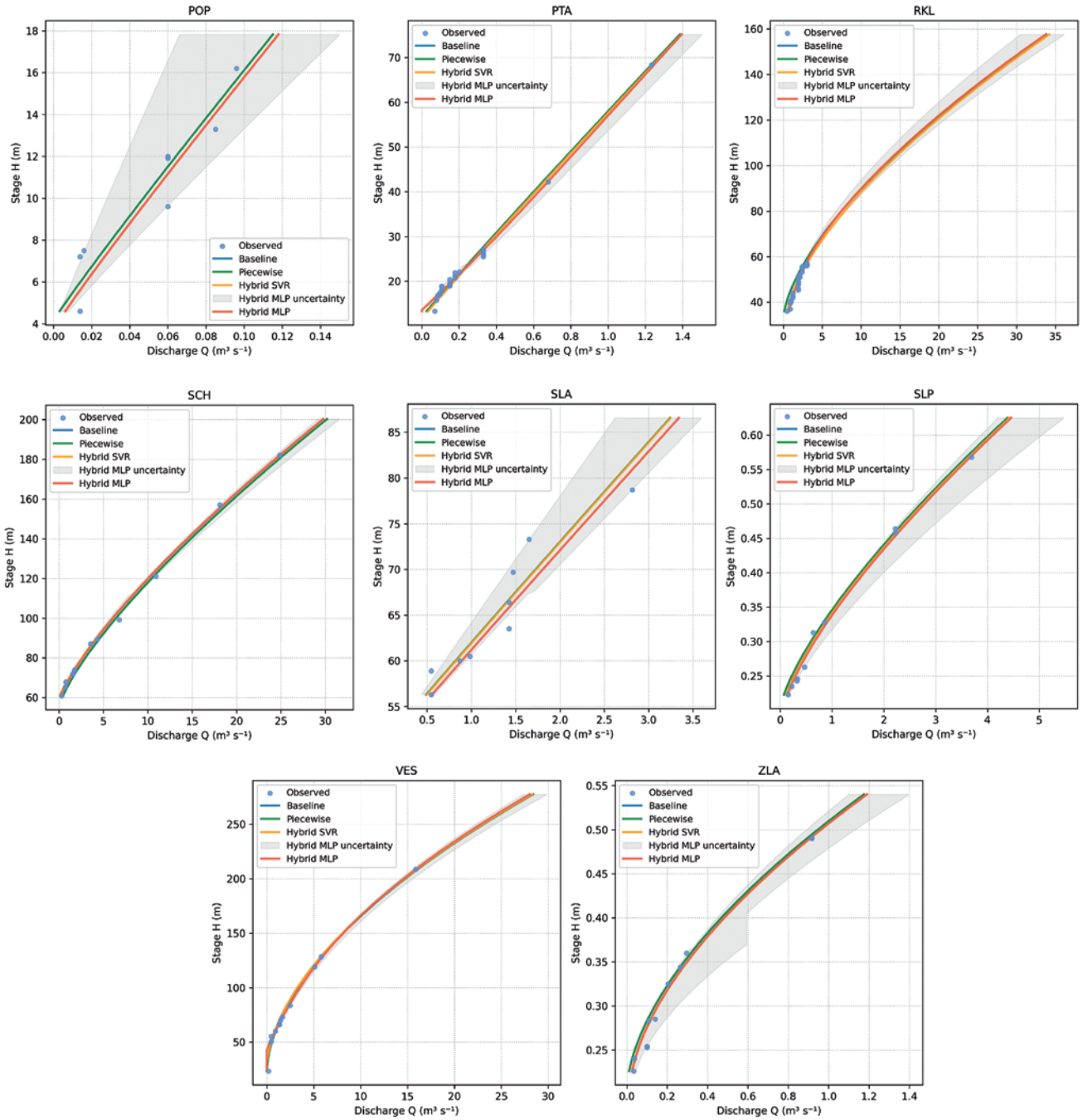

Fig. 8 Observed stage-discharge data and rating curves derived from the log-log baseline, Hybrid SVR, and Hybrid MLP models for 20 independent evaluation catchments.

3.3 Hybrid model performance in independent catchments

The proposed approach was evaluated on 20 independent catchments to assess its applicability and reliability (Fig. 8). Rating curves derived from the log-log baseline, hybrid SVR, and hybrid MLP models were compared with observed stage-discharge measurements.

Across most sites, the Hybrid MLP closely matched observations, capturing both slope and curvature, even under sparse measurement conditions. The Hybrid SVR generally performed well but showed larger deviations at high or low discharges, while the log-log baseline reproduced overall trends but tended to underfit nonlinear relationships in sites with wide discharge ranges. These results demonstrate the capacity of the Hybrid MLP to generalize the stage-discharge relationship across diverse physiographic and hydrological settings.

In addition to model performance, the predictive uncertainty of the hybrid MLP curves was assessed using residual-based quantile envelopes (Fig. 8). The shaded envelopes indicate 95% prediction intervals, typically narrowing in the central stage-discharge

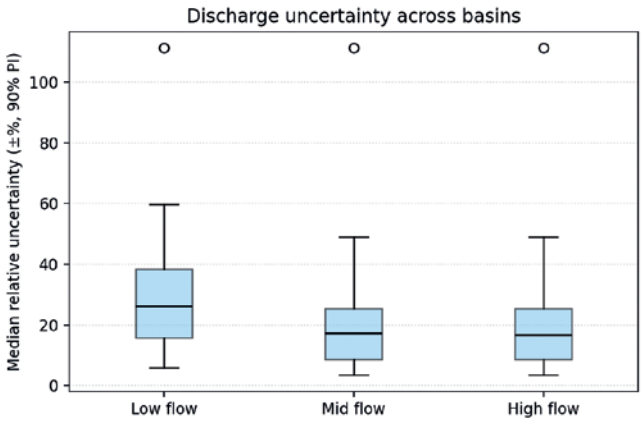

Fig. 9 Distribution of relative discharge uncertainty (±%, 90% prediction interval) across all catchments, grouped by flow terciles.

domain and widening toward the high-flow range. The hybrid MLP reproduces both slope and curvature of the stage-discharge relationship across diverse hydrological conditions, while maintaining narrow uncertainty envelopes for most of the observed range. The widening of prediction intervals toward high stages reflects the limited availability of extreme-flow measurements and the increased extrapolation uncertainty typical of rating-curve applications.

The corresponding summary of relative uncertainty across flow categories (Fig. 9) shows that median uncertainty decreases from low- to high-flow conditions, with median values typically below ±25% for mid- and high-flow regimes, indicating consistent predictive stability of the hybrid MLP model across varying flow conditions. The systematically higher uncertainty at low flows highlights the influence of measurement noise and channel-bed variability, emphasizing the importance of representative low-stage gauging when calibrating machine-learning rating curves.

Distribution of model performance metrics by station (Fig. 10) indicate high goodness of fit for all models, with median R^2 and NSE above 0.99. Nevertheless, the Hybrid MLP showed the narrowest variability and the highest consistency, while Hybrid SVR exhibited wider variation, and the log-log baseline displayed the lowest median performance and broader spread, particularly at low-performance tails (Fig. 10).

All models achieved high median values ($R^2 \approx 0.99$, NSE ≈ 0.98, IoA ≈ 0.99), their variability revealed differences in robustness (Tab. 3). The log-log baseline

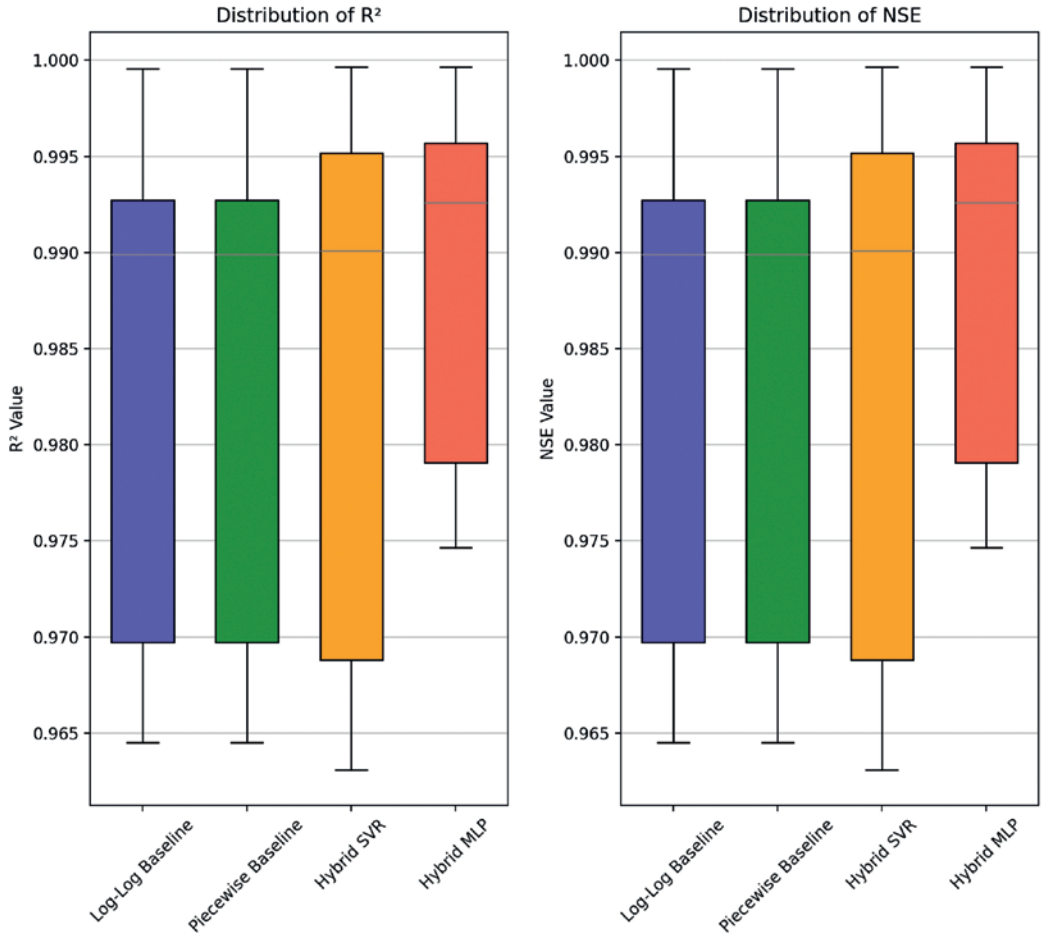

Fig. 10 Distribution of (a) R^2 and (b) NSE model performance values for the log-log and piecewise baseline, Hybrid SVR, and Hybrid MLP models across 20 independent evaluation catchments.

Tab. 3 Performance metrics for the baseline, hybrid MLP, and hybrid SVR models in 20 independent assessment catchments.

Metrics	R2			NSE			IoA			Standard deviation		
	median	max	min	median	max	min	median	max	min	R2	NSE	IoA
Hybrid MLP	0.9926	0.9997	0.8812	0.9926	0.9997	0.8812	0.9982	0.9999	0.9690	0.0343	0.0343	0.009
Model Hybrid SVR	0.9901	0.9996	0.8929	0.9901	0.9996	0.8929	0.9975	0.9999	0.9718	0.0360	0.0360	0.009
log-log Baseline	0.9899	0.9996	0.4669	0.9899	0.9996	0.4669	0.9975	0.9999	0.8929	0.1182	0.1182	0.024
Piecewise Baseline	0.9899	0.9996	0.4673	0.9899	0.9996	0.4673	0.9975	0.9999	0.8930	0.1182	0.1182	0.024

showed the largest spread (std R^2 = 0.1182) and lowest minimum performance (min R^2 = 0.467), indicating underfitting in catchments with nonlinear behavior. This comparison also revealed virtually identical results between the single-segment and three-segment log-log models (mean ΔNSE < 0.001, ΔR^2 < 0.0001), indicating that the additional breakpoints did not improve the fit within the observed range (Tab. 3). Both hybrid models reduced performance variability (std R^2 ≈ 0.034–0.036), with the MLP achieving the highest median values and narrowest spread, highlighting its generalizability. Maximum metrics were similar across all models, showing accurate fits in well-behaved catchments, while differences emerged primarily in complex or extreme stage-discharge conditions.

Analysis of best-performing models per station (Fig. 11), based on R^2 showed that the Hybrid MLP achieved the highest performance in 14 out of 20 catchments. The Hybrid SVR model achieved the highest performance in four catchments, while the baseline log-log and piecewise models achieved the highest performance in one catchment each.

4. Discussion

4.1 Effect of sample size on model performance

Our results show that hybrid PIML models perform reliably even with sparse calibration data, challenging the assumption that large datasets are essential. The Hybrid MLP, in particular, consistently outperformed both SVR and the log-log baseline across most stations. This highlights the framework's potential for experimental catchments where only 8–15 discharge measurements are typically available.

Specifically, hybrid MLP performed best at 15 out of 20 stations, while hybrid SVR performed best at 4 stations, and the log-log baseline model performed best at only 1 station.

These results demonstrate that ML-based approaches, particularly the Hybrid MLP, can provide accurate rating curve estimates and high NSE and R^2 values even with sparse calibration data (Fig. 12). The Hybrid MLP showed consistently strong performance regardless of the number of calibration points, indicating robustness under data-scarce conditions.

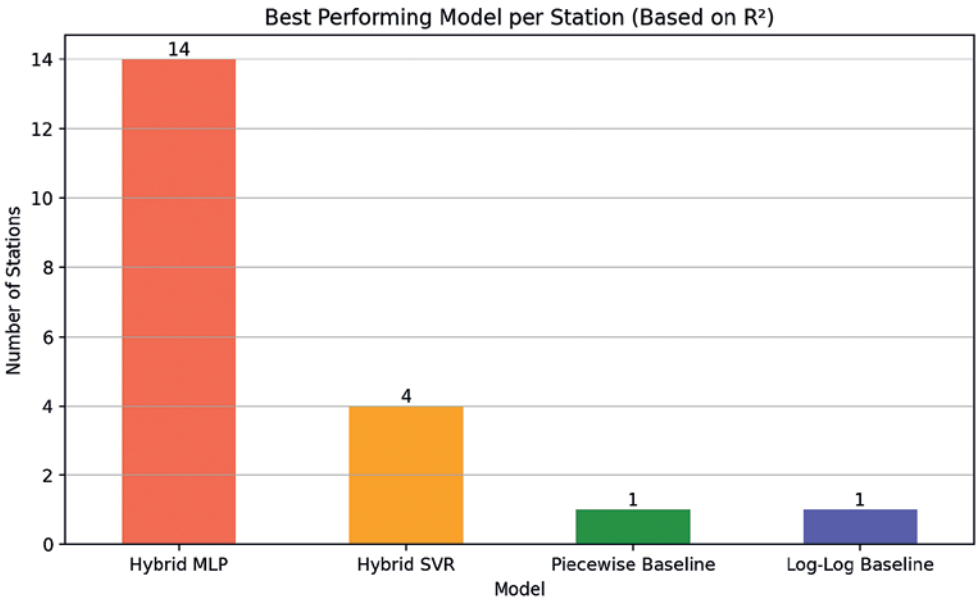

Fig. 11 Number of stations where each model achieved the highest R^2.

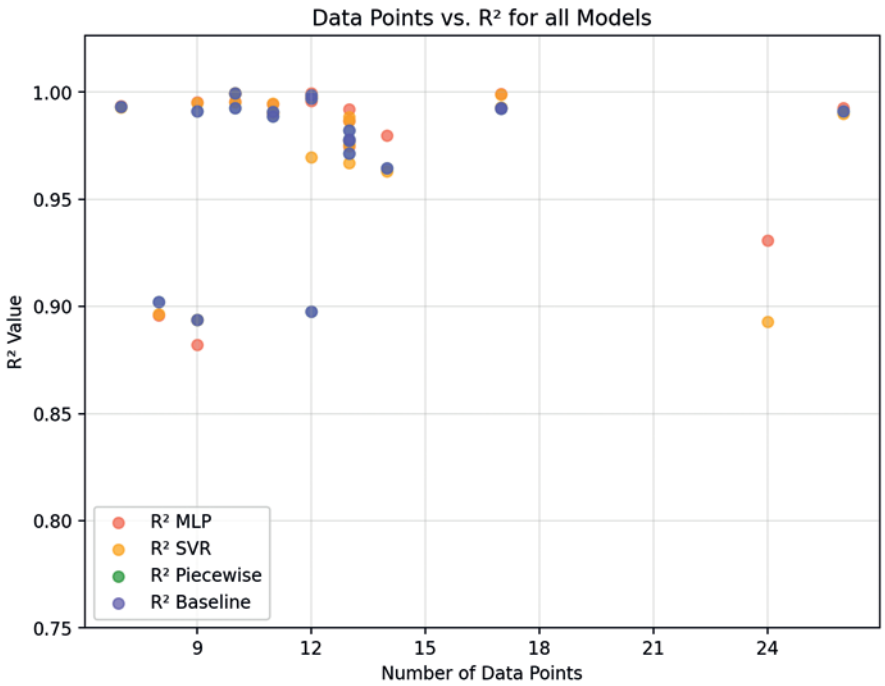

Fig. 12 Model performance metrics according to the number of points, used for rating curve modeling by the model type.

In contrast, the Hybrid SVR and the log-log baseline exhibited greater variability, especially when fewer than 12 observations were available.

While the number of data points influences performance, the accuracy of the underlying hydrometric observations exerts a stronger control. Poor-quality data characterized by noise or systematic errors cannot be compensated for by algorithmic sophistication.

The MLP outperformed SVR despite the latter's theoretical advantages for small datasets. This highlights the importance of empirical evaluation in physics-informed modeling. While SVR's margin-based learning and kernel methods are often recommended for limited data scenarios, the smooth and continuous relationships required by hydrological rating curves, appear better suited to the MLP architecture. This emphasizes that algorithm selection in physics-informed ML should prioritize alignment with the physical characteristics of the system rather than general assumptions about small-data performance.

4.2 Applications of the hybrid model

Hybrid machine learning (ML) models show particular promise for rating curve development in situations where conventional approaches fail or require extensive manual postprocessing. ML component offers robustness and capability to capture complex, nonlinear relationships that traditional methods may struggle to represent (Liang et al. 2023). This can be beneficial in cases where stage-discharge relationships vary due to differences in channel morphology, bed roughness, or flow conditions (Feng et al. 2023).

Our results demonstrate that systematic, physics-informed methods can extract meaningful relationships even from datasets that would be considered inadequate for traditional ML applications. While hybrid approaches cannot compensate for deficiencies in observation quality or coverage, they contribute in several important ways:

(i) Objective methodology: By replacing subjective manual curve fitting with a standardized procedure, hybrid models provide reproducible results that reduce practitioner bias.
(ii) Maximum information extraction: Combining physical constraints with flexible learning algorithms allows greater use of limited data than purely empirical or purely physical models.
(iii) Transferability: The consistent methodology supports comparative hydrology by enabling application across diverse catchments.

These characteristics make hybrid models attractive for several hydrometric contexts. Hybrid models, particularly those using nonlinear algorithms such as multi-layer perceptrons (MLPs) or support vector regression (SVR), can better capture complex stage-discharge relationships (Mosavi et al. 2018). PIML models are likewise useful in rivers with high flow variability, turbulence, or backwater effects, where strong nonlinearities reduce the performance of traditional rating curves (Di Baldassarre and Montanari 2009). Hybrid models can also integrate data from different instruments with varying levels of precision that would otherwise degrade predictive accuracy.

Applications extend to long-term records, where channel morphology may evolve over time. Finally,

hybrid models hold potential for reconstructing historical records from sparse or fragmented observations. By combining physically plausible baseline functions with ML corrections, they provide more robust reconstructions than either method alone (Belitz and Stackelberg 2021).

4.3 Limitations and future directions

Despite its advantages, this study remains a proof-of-concept rather than a definitive framework for physics-informed ML in rating curve modeling. A key limitation is the black-box nature of the ML component: unlike physically based or empirical models, its internal decision processes are opaque, which reduces interpretability and complicates transferability in operational contexts (Beven 2019).

Model performance also depends strongly on data quality and quantity. Sparse or noisy observations, typical of experimental catchments, increase the risk of overfitting and poor generalization (Huntingford et al. 2019). The small datasets in this study (8–30 measurements per site) highlight these constraints, limiting statistical inference about algorithm superiority. Nonetheless, within such conditions, the hybrid framework offers a systematic and reproducible alternative to subjective manual fitting while embedding essential physical constraints.To mitigate these challenges, recent research advocates physics-informed ML (Raissi et al. 2019), embedding physical laws such as conservation principles or monotonicity directly into model structures.

Future research should advance physics-informed ML by developing uncertainty quantification methods tailored to small-sample hydrology, testing performance on controlled synthetic datasets, and defining evidence-based guidelines for minimum data requirements. Integrating auxiliary information, such as hydraulic modeling or remote sensing, could further enhance model robustness under data scarcity.

Hybrid rating curve models therefore represent a promising alternative to conventional approaches, particularly in basins with irregular morphology, dynamic flow regimes, heterogeneous measurements, or evolving channels. Their strength lies in combining physical principles with flexible learning, but reliable application will require continued attention to data quality, constraint design, and rigorous evaluation. Beyond the data-driven focus of this study, physically based Bayesian frameworks such as BaRatin (Le Coz et al. 2014; Kiang et al. 2018) demonstrate how hydraulic principles and uncertainty quantification can be jointly incorporated in rating-curve modeling. Implementing such models requires detailed cross-sectional and control-type information that was beyond the scope of our dataset, but their concepts are highly complementary to the machine-learning approaches tested here. Future work will therefore explore hybrid Bayesian-ML strategies that combine the interpretability and uncertainty propagation in tools such as BaRatin with the flexibility and generalization capacity of machine-learning models.

5. Conclusions

This study demonstrates that physics-informed machine learning (PIML) provides a systematic framework for developing reliable stage-discharge relationships under the data constraints typical of experimental catchments. By combining a physically based log-log baseline with machine learning residual corrections, the approach ensures monotonicity, non-negativity, and continuity while enhancing predictive skill. While acknowledging the constraints imposed by small sample sizes, the results show that the hybrid model consistently improves on conventional regression approaches, particularly in basins with irregular morphology or high measurement uncertainty.

Among the tested algorithms, the Hybrid MLP proved the most robust and generalizable across diverse hydrological settings, performing best at the majority of sites. The Hybrid SVR also delivered strong results and showed comparative advantages in highly irregular environments, suggesting that algorithm choice may be guided by site-specific conditions. In contrast, the traditional log-log regression performed best only in hydraulically simple profiles, highlighting the benefits of hybridization where nonlinearities dominate.

The strength of the proposed framework lies not in maximizing statistical fit with limited samples, but in providing a reproducible and physically constrained methodology that reduces the subjectivity of manual curve fitting. This makes it particularly relevant for experimental hydrology, where discharge estimation is often based on sparse, heterogeneous, and uncertain observations.

The MLP model demonstrated best overall performance (median R^2 and NSE > 0.98), achieving best results at 15 out of 20 evaluation sites, while the SVR model performed best at 4 sites, and the log-log baseline at only one site. Statistical analysis confirms MLP's consistency, with lower variability in performance metrics and higher median values across all metrics.

Limitations remain, notably the dependence on the quality of available hydrometric data and the limited interpretability of the machine learning components. Nevertheless, the approach establishes a proof-of-concept for integrating physical principles with data-driven learning in rating curve development. Future work should focus on refining uncertainty quantification, testing transferability under changing channel conditions, and integrating auxiliary data sources such as hydraulic modeling or remote sensing.

By bridging physics-based principles and machine learning techniques, this framework offers experimental hydrology a practical tool for objective and transferable rating curve development, extending the scope of reliable discharge estimation in data-scarce and hydraulically complex environments.

Acknowledgments

This research was supported by the Czech Science Foundation project 22-12837S: Hydrological and hydrochemical responses of montane peat bogs to climate change, and by the Technology Agency of the Czech Republic project SS02030040.

References

Adarsh, S., John, A. P., Anagha, R. N., Abraham, A., Afiya, M. P., Arathi, K. K., Azeem, A. (2018): Developing stage-discharge relationships using multivariate empirical mode decomposition-based hybrid modeling. Applied Water Science 8: 230, https://doi.org/10.1007/s13201-018-0874-8.

Ali, G., Maghrebi, M. F. (2023): A robust approach for the derivation of rating curves using minimum gauging data. Journal of Hydrology 623: 129609, https://doi.org/10.1016/j.jhydrol.2023.129609.

Belitz, K., Stackelberg, P. E. (2021): Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. Environmental Modelling & Software 139: 105006, https://doi.org/10.1016/j.envsoft.2021.105006.

Beven, K. (2019): How to make advances in hydrological modelling. Hydrology Research, 50(6), 1481–1494, https://doi.org/10.2166/nh.2019.134.

Bhasme, P., Vagadiya, J., Bhatia, U. (2022): Enhancing predictive skills in physically-consistent way: Physics Informed Machine Learning for hydrological processes. Journal of Hydrology 615: 128618, https://doi.org/10.1016/j.jhydrol.2022.128618.

Braca, G. (2008): Stage-discharge relationships in open channels: practices and problems, https://api.semanticscholar.org/CorpusID:53496646.

Di Baldassarre, G., Montanari, A. (2009): Uncertainty in river discharge observations: a quantitative analysis. Hydrology and Earth System Sciences 13(6), 913–921, https://doi.org/10.5194/hess-13-913-2009.

Dobrovolski, S. G., Yushkov, V. P., Vyruchalkina, T. Y., Sokolova, O. V. (2022): Are There Fundamental Laws in Hydrology? Pure and Applied Geophysics 179, 1475–1484, https://doi.org/10.1007/s00024-022-03003-1.

Esmaeilzadeh, M., Amirzadeh, M. (2024): Replication Study: Enhancing Hydrological Modeling with Physics-Guided Machine Learning. Computer Science. Machine Learning, arXiv:2402.13911v1, https://doi.org/10.48550/arXiv.2402.13911.

Feng, D., Beck, H., Lawson, K., Shen, C. (2023): The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. Hydrology and Earth System Sciences 27(12), 2357–2373, https://doi.org/10.5194/hess-27-2357-2023.

Herschy, R. W. (2019): Streamflow measurement (3rd ed.). CRC Press, https://doi.org/10.1201/9781482265880.

Hrafnkelsson, B., Sigurdarson, H., Rögnvaldsson, S., Jansson, A. Ö., Vias, R. D., Gardarsson, S. M. (2022): Generalization of the power-law rating curve using hydrodynamic theory and Bayesian hierarchical modeling. Environmetrics 33(2): e2711, https://doi.org/10.1002/env.2711.

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., Yang, H. (2019): Machine learning and artificial intelligence to aid climate change research and preparedness. Environmental Research Letters 14(12): 124007, https://doi.org/10.1088/1748-9326/ab4e55.

ISO (2020): Hydrometric uncertainty guidance (HUG), ISO 25377:2020.

ISO (2021): Hydrometry – Measurement of liquid flow in open channels – Velocity area methods using point velocity measurements, EN ISO 748:2021.

Kennedy, E. J. (1984): Discharge ratings at gaging stations. Techniques of Water-Resources Investigations 03A10, https://doi.org/10.3133/twri03A10.

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., Belleville, A., Sevrez, D., Sikorska, A. E., Petersen-Øverleir, A., Reitan, T., Freer, J., Renard, B., Mansanarez, V., Mason, R. (2018): A Comparison of Methods for Streamflow Uncertainty Estimation. Water Resources Research 54(10), 7149–7176, https://doi.org/10.1029/2018WR022708.

Kratzert, F., Klotz, D., Herrnegger, M. (2019a): Toward improved predictions in ungauged basins: Exploiting the power of machine learning. Water Resources 55(12), 11344–11354, https://doi.org/10.1029/2019WR026065.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G. (2019b): Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences 23(12), 5089–5110, https://doi.org/10.5194/hess-23-5089-2019.

Lane, S. N. (1998): Hydraulic modelling in hydrology and geomorphology: A review of high resolution approaches. Hydrological Processes 12(8), 1131–1150, https://doi.org/10.1002/(SICI)1099-1085(19980630)12:8<1131::AID-HYP611>3.0.CO;2-K.

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., Le Boursicaud, R. (2014): Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. Journal of Hydrology 509, 573–587, https://doi.org/10.1016/j.jhydrol.2013.11.016.

Liang, W., Chen, Y., Fang, G., Kaldybayev, A. (2023): Machine learning method is an alternative for the hydrological model in an alpine catchment in the Tianshan region, Central Asia. Journal of Hydrology, Regional Studies 49: 101492, https://doi.org/10.1016/j.ejrh.2023.101492.

Liu, X., Lu, D., Zhang, A., Liu, Q., Jiang, G. (2022): Data-Driven Machine Learning in Environmental Pollution: Gains and Problems. Environmental Science & Technology 56(4), 2124–2133, https://doi.org/10.1021/acs.est.1c06157.

McMillan, H. K., Westerberg, I. K. (2015): Rating curve estimation under epistemic uncertainty. Hydrological

Processes 29(7), 1873–1882, https://doi.org/10.1002/hyp.10419.

Mosavi, A., Ozturk, P., Chau, K.-W. (2018): Flood Prediction Using Machine Learning Models: Literature Review. Water 10(11): 1536, https://doi.org/10.3390/w10111536.

Nearing, G. S., Kratzert, F., Sampson, A. K. (2021): What role does hydrological science play in the age of machine learning? Water Resources Research 57(3): e2020WR028091, https://doi.org/10.1029/2020WR028091.

Poulinakis, K., Drikakis, D., Kokkinakis, I. W., Spottswood, S. M. (2023): Machine-Learning Methods on Noisy and Sparse Data. Mathematics 11(1): 236, https://doi.org/10.3390/math11010236.

Raissi, M., Perdikaris, P., Karniadakis, G. E. (2019): Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707, https://doi.org/10.1016/j.jcp.2018.10.045.

Roelofs, R., Shankar, V., Recht, B., Fridovich-Keil, S., Hardt, M., Miller, J., Schmidt, L. (2019): A Meta-Analysis of Overfitting in Machine Learning. Advances in Neural Information Processing Systems 32. Available online: https://dl.acm.org/doi/pdf/10.5555/3454287.3455110 (accessed on 10 May 2025).

Shortridge, J. E., Guikema, S. D., Zaitchik, B. F. (2016): Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences 20(7), 2611–2628, https://doi.org/10.5194/hess-20-2611-2016.

WMO (2010): Manual on stream gauging. Volume I – Fieldwork. WMO-No. 1044, World Meteorological Organization. Available online: https://library.wmo.int/records/item/35848-manual-on-stream-gauging-vol-i-fieldwork (accessed on 10 May 2025).

WMO (2020): Guide to Hydrological Practices I. Hydrology – From Measurement to Hydrological Information. WMO-No. 168, World Meteorological Organization. Available online: https://library.wmo.int/records/item/35804-guide-to-hydrological-practices-volume-i?offset=6 (accessed on 10 May 2025).

Xu, T., Liang, F. (2021): Machine learning for hydrologic sciences: An introductory overview. WIREs Water 8(5), https://doi.org/10.1002/wat2.1533.

Xu, W., Chen, J., Corzo, G., Xu, C.-Y., Zhang, X. J., Xiong, L., Liu, D., Xia, J. (2024): Coupling deep learning and physically based hydrological models for monthly streamflow predictions. Water Resources Research 60(2): e2023WR035618, https://doi.org/10.1029/2023wr035618.

Ying, X. (2019): An Overview of Overfitting and its Solutions. Journal of Physics, Conference Series 1168(2): 022022, https://doi.org/10.1088/1742-6596/1168/2/022022.