A Hybrid Muskingum–Machine Learning Flood Forecasting Model: Application and Evaluation in the Tarim River Basin

Pengyang Wang , Ling Zhang * , Donglin Li * , Fengzhen Tang , Xin Wang , Yuanjian Wang

Posted Date: 22 April 2026

doi: 10.20944/preprints202604.1577.v1

Keywords: flood forecasting; Muskingum model; machine learning; residual correction; Tarim River

 doi:10.20944/preprints202604.1577.v1

Article

A Hybrid Muskingum–Machine Learning Flood Forecasting Model: Application and Evaluation in the Tarim River Basin

Pengyang Wang [1,2], Ling Zhang [2,3,*], Donglin Li [2,3,*], Fengzhen Tang [2,3],Xin Wang [2,3] and Yuanjian Wang [2,3]

[1] School of Hydraulic Engineering, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, China
[2] Yellow River Institute of Hydraulic Research, Yellow River Conservancy Commission, Zhengzhou 450003, China
[3] Key Laboratory of Yellow River, MWR, Zhengzhou 450003, China
* Correspondence: zhanglyrcc@foxmail.com (L.Z.); lidonglinl@163.com (D.L.)

Highlights

What are the main findings?

- A hybrid Muskingum-machine learning framework significantly improves flood forecast accuracy in the arid Tarim River Basin.
- Linear regularized models (Ridge, LASSO) outperform complex models (LSTM, RF) for multi-step-ahead residual correction.
- The study reveals the Muskingum model's residuals are linearly correlated with discharge, justifying simple linear correctors.
- A lead-time-dependent strategy is optimal: Ridge for short leads (<12h), LASSO for longer leads (16-24h).

What are the implications of the main findings?

- Demonstrates that increasing model complexity does not guarantee better performance in hybrid hydrological forecasting.
- Offers a practical, lead-time-adaptive modeling strategy for real-time flood management in complex, data-scarce arid regions.
- Proves that coupling simple physical models with machine learning effectively corrects systematic biases like peak underestimation.

Abstract

The traditional Muskingum model has difficulty representing complex hydraulic behaviors under high-flow conditions because it relies on simplified assumptions and fixed parameters. To address the pronounced nonlinearity and non-stationarity of flood routing in the arid Tarim River Basin, a hybrid forecasting framework was developed by coupling the Muskingum method with multiple machine-learning algorithms (Ridge, LASSO, RF, and LSTM) to predict and correct Muskingum residuals. Global Muskingum parameters were identified using the L-BFGS-B algorithm to represent basin-scale routing characteristics. For rolling forecast, a multidimensional feature space was constructed by integrating routing gradients and hydraulic interaction terms. The results indicated that all hybrid models outperformed the traditional Muskingum method across lead times. The Ridge-based hybrid model achieved the best performance at short lead times, with the Nash–Sutcliffe efficiency (NSE) at a 4-h lead time increasing from 0.56 for the physical baseline to 0.977. For longer lead times (12–24 h), the LASSO-based hybrid model demonstrated higher robustness, which was attributed to L1-regularization-based feature selection. The key scientific contribution of this work lies in proposing a lead-time-dependent adaptive modeling strategy, revealing the structural characteristics of the residuals of the Muskingum model, and demonstrating that, in the study basin,

 doi:10.20944/preprints202604.1577.v1

simple linear models outperform complex models in multi-step correction. Overall, the proposed framework alleviates systematic underestimation during high-flow periods and provides a predictive scheme for arid-region rivers that preserves physical interpretability while improving forecasting accuracy.

Keywords: flood forecasting; Muskingum model; machine learning; residual correction; Tarim River

1. Introduction

Floods are among the most destructive natural disasters globally, and accurate forecasting is vital for disaster prevention and mitigation, adaptive water resource regulation, and social safety [1,2]. River flood routing is a core component of flood forecasting systems, and its physical basis lies in the representation of unsteady flow evolution in complex river channels [3].

The Muskingum method, a classical linear hydrological routing model, has been widely applied due to its clear physical mechanism and computational efficiency [4]. However, its reliance on simplified assumptions—such as regular channel cross-sections and linear storage–discharge relationships—limits its ability to capture nonlinear dynamics induced by floodplain inundation and backwater effects under high-flow conditions [5,6]. Moreover, its parameters are typically assumed time-invariant, neglecting the influence of changing underlying surfaces and lateral inflows, which often results in systematic magnitude bias when applied to highly non-stationary time series. These limitations highlight the need for hybrid forecasting approaches that integrate deterministic physical mechanisms with stronger nonlinear representation capabilities.

Recent studies have explored coupling physical routing mechanisms with data-driven algorithms to overcome the limitations of traditional models. Gauch et al. [7] applied LSTM networks for multiscale runoff simulations and found that deep learning could capture seasonal storage–release patterns missed by physical models. Klotz et al. [8] demonstrated that incorporating physical constraints improved the generalization of machine learning models in data-scarce alpine regions. Akbari et al. [9] proposed a nonlinear Muskingum model with segmented variable parameters, showing improved accuracy through hybrid optimization. Lin et al. [10] and Li et al. [11] developed hybrid deep learning frameworks that integrate physical mechanisms with neural networks to enhance forecasting reliability.

In other hydrological contexts, studies have shown that deep learning models can achieve superior performance, particularly in watersheds with strong nonlinear rainfall-runoff relationships. For instance, in the Clear Creek and Upper Wapsipinicon River watersheds in Iowa, LSTM-based models significantly outperformed linear models, achieving NSE values of 0.85–0.93 for 24-hour lead time predictions [12]. Similarly, in the Dongjiang River Basin, LSTM achieved optimal performance (NSE = 0.95), while the competitive performance of linear models was attributed to the high autocorrelation of daily streamflow and the effective linearization of the prediction problem through antecedent flow variables [13].

While these studies have advanced the field, most focus on either replacing the physical model with complex learners or developing increasingly sophisticated nonlinear variants of Muskingum. Few have systematically investigated the structural characteristics of the residuals produced by the physical model, nor have they examined how these characteristics should guide the selection of machine learning algorithms for correction—especially in a rolling multi-step forecasting context. The contrasting performance of linear versus deep learning models across different basins underscores a key principle: model selection should be guided by the structural characteristics of the residuals, not by model complexity alone—a principle that this study systematically explores in the context of the Tarim River Basin.

This study addresses this gap by demonstrating that, for the Tarim River Basin, the residuals exhibit a strong linear correlation with discharge, and that simple linear regularized models (Ridge, LASSO) outperform complex models (RF, LSTM) for multi-step residual correction. This finding

 doi:10.20944/preprints202604.1577.v1

challenges the conventional assumption that increasing model complexity invariably yields better forecasting performance.

The Tarim River Basin, an arid inland system, serves as an exceptionally challenging test case due to its pronounced hydrological non-linearity and non-stationarity. Specifically, its flood processes are driven by a volatile combination of high-mountain snow/ice melt—accounting for over 40% of total runoff—and extreme localized rainfall [14]. This dominance of cryospheric melt makes the hydrograph highly sensitive to rapid temperature fluctuations, resulting in extreme flow variability where peak discharge can exceed baseflow by several orders of magnitude. Furthermore, intensive human interventions, such as large-scale irrigation diversions and reservoir operations, introduce artificial perturbations that decouple natural flow patterns [15]. These factors, combined with the stochastic nature of "flashy" floods in desert reaches, pose significant challenges for traditional hydraulic and hydrological forecasting techniques. Analysis of historical flood routing in the Tarim River reveals that the forecasting residuals of the Muskingum model exhibit correlations with upstream flow gradients, routing intensity, and the spatiotemporal characteristics of the river channel, rather than behaving as random white noise—demonstrating a relationship between nonlinear disturbances not captured by the mechanistic model and the basin's hydrodynamic response. This observation motivated the development of a hybrid forecasting framework, in which the Muskingum model provides a baseline routing phase, while machine learning algorithms with temporal memory capabilities are employed for nonlinear approximation and real-time residual correction.

The main contributions of this study are as follows: first, a baseline routing layer was constructed based on global optimization, where the L-BFGS-B algorithm was employed to identify Muskingum parameters representing the intrinsic basin characteristics across the entire time series, thereby alleviating the generalization issues associated with event-specific parameter heterogeneity; second, a multidimensional feature-coupled residual correction strategy was proposed, which, by constructing an eight-dimensional feature space incorporating baseline discharge and historical memory, enabled the model to capture dynamic disturbances beyond those described by physical equations; and third, the performance boundaries of mechanistic–data fusion models were systematically evaluated by comparing the residual correction performance of ridge regression (Ridge), LASSO regression, random forest (RF), and LSTM across multiple lead times, thereby identifying the performance characteristics of models with different levels of complexity under physical mechanism constraints.

2. Materials and Methods

2.1. Study Area and Data

This study focused on key river reaches within the Tarim River Basin, the largest inland river system in China, commonly referred to as the "three sources and one main stream" (Figure 1). The Tarim River Basin is in southern Xinjiang, China, deep within the interior of the Eurasian continent, and represents a typical arid inland river basin. The formation, transformation, and consumption of water resources in this basin exhibit pronounced complexity and distinct regional characteristics[18,19]. The term "three sources and one main stream" refers to the three major headwater tributaries in the upper reaches of the Tarim River, including the Aksu River, the Yarkant River, and the Hotan River, as well as the upper main stem formed following their confluence. This study focused on the confluence region of the Aksu River, the Yarkant River, and the Hotan River near Alar. Detailed information for the hydrological stations used in this study, including geographic coordinates, catchment area, and elevation, is provided in Table S1 in the Supplementary Materials.

 doi:10.20944/preprints202604.1577.v1

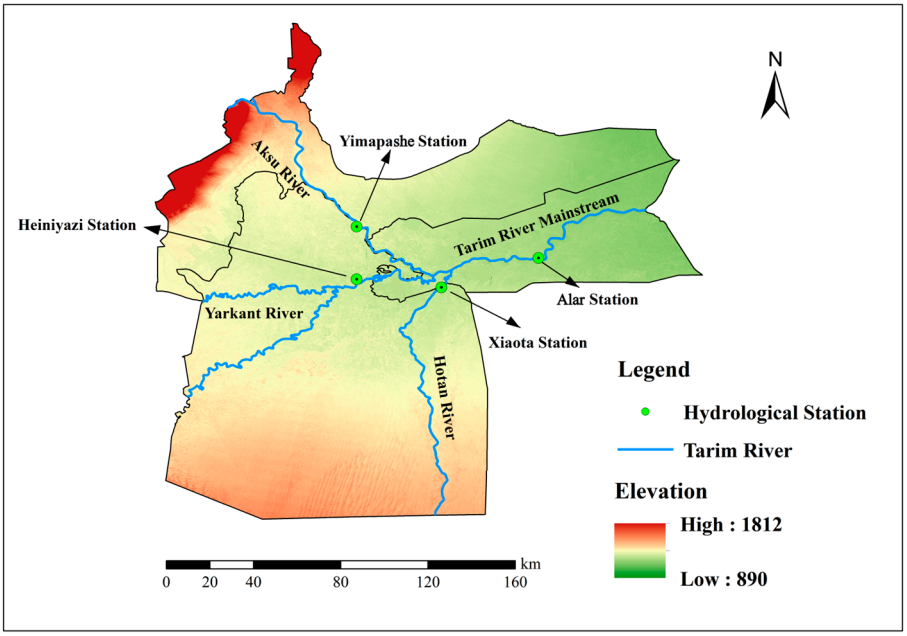


Figure 1. Study area.

Hydrological data were obtained from representative stations in this region, including the Yimapashe Station on the Aksu River (representing inflow from the Aksu River), Xiaota Station (representing inflow after the confluence of the Yarkant and Hotan Rivers), and Alar Station on the main stem (representing downstream outflow).

The original dataset comprised discharge measurements at two temporal resolutions (Table 1): 4-hour and daily intervals. For the three stations (Xiaota, Yimapashe, and Alar), 4-hour data accounted for 46.4–47.5% of the 1,747-day monitoring period, daily data for 52.4–53.4%, and completely missing days totaled only 3–4 per station (missing rate <0.23%). To construct a continuous time series at a consistent temporal scale suitable for flood routing, the data were resampled to a uniform 4-hour interval. Missing daily values were filled using linear interpolation for the few completely missing days, while daily-scale data (approximately half the dataset) were interpolated to 4-hourly values using the piecewise cubic Hermite interpolating polynomial (PCHIP) method, which preserves the monotonicity of the runoff series and minimizes interpolation-induced oscillations [16]. Forward and backward filling ensured data completeness. The reliability of the resampled dataset is supported by the strong performance of the baseline Muskingum model during training (NSE > 0.85 for most high-flow events) and the consistent improvements achieved by the hybrid models, indirectly validating the preprocessing approach.

Table 1. Summary of data availability and temporal resolution for hydrological stations.

Station Name	Xiaota	Yimapashe	Alar
Start Time	2021/1/1 8:00	2021/1/1 8:00	2021/1/1 8:00
End Time	2025/10/13 16:00	2025/10/13 16:00	2025/10/13 16:00
Total Days	1747	1747	1747
4-hour Data (days)	811	812	829
4-hour Proportion (%)	46.42	46.48	47.45
Daily Data (days)	932	932	915
Daily Proportion (%)	53.35	53.35	52.38
Completely Missing Days	4	3	3

The dataset was divided into two parts in chronological order.

 doi:10.20944/preprints202604.1577.v1

The training set includes complete data from January 1, 2021, to December 31, 2024, and is used for model calibration and machine learning training. This period was chosen to encompass the existing well-documented historical flood events in the Tarim River Basin, allowing for robust calibration of the physical model and comprehensive training of the machine learning models.

The test set involves the flood period from 08:00 on July 10, 2025, to 04:00 on August 24, 2025, and is used for independent evaluation of the flood forecasting performance of the model. The selection of this period is mainly based on the following considerations: First, starting from 08:00 on July 10, 2025, the measured flow at Alar Station exceeded 1000 m^3 /s for the first time, marking the beginning of a significant flood event. Subsequently, the flow receded below 1000 m^3/s by 04:00 on August 24, 2025, with the entire flood process lasting approximately 45 days. During this period, the station experienced multiple independent flood peaks, with the highest peak flow reaching 1970 m^3/s, representing a typical multi-peak flood event. Since this study focuses on the accuracy of flood peak forecasting, selecting a flood period that contains multiple representative peaks with distinct magnitude variations allows for a more rigorous and comprehensive evaluation of the proposed hybrid model’s predictive capability under flood events of varying intensities, thereby enhancing the reliability of model assessment and the generalizability of the conclusions.

2.2. Model Framework: Hybrid Forecasting System

This study developed and implemented a hybrid runoff forecasting framework that integrated physical routing mechanisms with data-driven approaches. The framework was designed to improve runoff forecasting accuracy and generate prediction results at a temporal resolution of 4 h with a maximum lead time of 24 h (six forecasting steps)[17]. The framework consists of three primary modules.

Baseline module (Muskingum method): The Muskingum model parameters were calibrated using training period data to construct a mechanistic routing model for segmented flow propagation. The upstream inflows from the Aksu River (Yimapashe Station) and Yarkant and Hotan Rivers (Xiaota Station) were used as inputs to produce baseline discharge simulations at the downstream Alar Station, denoted as , which serves as a physical reference for subsequent residual correction.

Intelligent correction module (machine learning-based residual forecasting): The residual series between the baseline simulations and the observed discharge was calculated as . Subsequently, a corresponding feature set was constructed. Four machine learning models, including ridge regression (Ridge), LASSO, RF, and LSTM, were trained in parallel. These models learned the mapping between the constructed features at the current time step and the residuals and output the corresponding residual correction .

Rolling forecasting and evaluation module: During the testing phase, a strict time-rolling forecasting strategy was implemented for multi-step forecasting. For each time step (t) in the test sequence, only the observational information available at that time was applied to drive the models and compute features (with the true values used at the initial step and predictions from the previous step used thereafter). Residual corrections were subsequently obtained using the trained machine learning models. The final forecast was expressed as . The model performance was evaluated using metrics including the Nash–Sutcliffe efficiency (NSE) and the root mean square error (RMSE).

2.3. Muskingum Method

In this study, the classical Muskingum segmented continuous routing scheme was adopted as the physical core of the proposed hybrid framework. The formulation was based on the channel water balance equation combined with a linear storage–discharge relationship[4]. The routing equation is expressed as follows:

$$Q_2=C_0I_2+C_1I_1+C_2Q_1 \quad (1)$$

where I_1 and I_2 denote the inflow discharges at the beginning and end of the time interval (m^3/s), respectively; Q_1 and Q_2 denote the corresponding outflow discharges (m^3/s); and C_0, C_1, and C_2 are

 doi:10.20944/preprints202604.1577.v1

the routing coefficients determined jointly by the parameter K (storage constant, representing flood wave travel time), x (flow weighting factor, reflecting channel storage capacity), and the computational time step Δt (4 h in this study). The coefficients are calculated as follows:

$$C_0=\frac{0.5\Delta t-Kx}{0.5\Delta t+K-Kx},\quad C_1=\frac{0.5\Delta t+Kx}{0.5\Delta t+K-Kx},\quad C_2=\frac{K-Kx-0.5\Delta t}{0.5\Delta t+K-Kx} \tag{2}$$

To compensate for the intermediate inflow or losses not explicitly considered by the model, a gain coefficient b was introduced to apply a multiplicative correction to the upstream inflow[18]. For each tributary, the adjusted inflow $I_{adj}(t)$ was calculated as:

$$I_{adj}(t) = I_{raw}(t) \times (1 + b) \tag{3}$$

where $I_{raw}(t)$ is the original observed inflow discharge (m^3/s), and b is the gain coefficient with a value range of [−0.5, 0.5]. A positive b (b>0) represents the unmonitored areal rainfall contributions or tributary inflows, whereas a negative b (b<0) represents the losses caused by channel evaporation, infiltration, or agricultural water withdrawals.

A two-stage calibration strategy was adopted for parameter estimation, with both stages using the L-BFGS-B (Limited-memory Broyden–Fletcher–Goldfarb–Shanno with bound constraints) optimization algorithm to minimize the mean squared error. This algorithm applies a limited-memory quasi-Newton method to handle constrained parameter optimization problems and exhibits high convergence efficiency[19]:

$$\min_{\theta} f\,(K, x, b) = \frac{1}{N}\sum_{i=1}^{N}(\,Q_{obs,i} - Q_{sim,i})^2 \tag{4}$$

Flood event calibration: Independent flood events were identified from the training dataset (2021–2024) based on the criteria of discharge exceeding 500 m^3/s, duration ≥ 24 h, and inter-event separation ≥ 72 h. For each identified event, the model parameters were independently calibrated using the L-BFGS-B algorithm with MSE minimization as the objective.

Global calibration: This was subsequently conducted using the continuous time series of the entire training period to obtain a set of globally optimal Muskingum parameters representing the basin-scale routing characteristics for the baseline simulation in the subsequent hybrid forecasting framework.

Routing procedure: After calibration, segmented continuous routing calculations were performed for the Aksu River (from Yimapashe to Alar) and the Hotan River (from Xiaota to Alar). The initial downstream discharge at the Alar Station was set as a weighted sum of 60% of the observed discharge at the Yimapashe Station and 40% of the observed discharge at the Xiaota Station, based on their long-term average flow contribution ratios. The routing results from the two river segments were then combined to obtain the baseline downstream discharge simulation at the Alar Station, denoted as .

2.4. Machine Learning Models

After the baseline discharge simulations from the physical routing component, residual learning was performed using the four adopted machine learning models.These models were selected to represent a spectrum of complexity, from simple linear regularized models (Ridge, LASSO) to a non-linear ensemble model (RF) and a complex sequential model (LSTM), allowing for a systematic investigation into how model complexity affects residual correction performance. For each model, hyperparameter optimization was performed using grid search with 5-fold cross-validation on the training set to ensure fair and optimal comparison. The final hyperparameters used for each model are detailed in Table S2 in the Supplementary Materials.

Ridge regression: Residual forecasting was first implemented using a linear regression model incorporating L2 regularization, in which the regularization parameter alpha was set to 1.0. This

 doi:10.20944/preprints202604.1577.v1

formulation enables the construction of a stable linear mapping from the input feature space to the residual response variable[20].

LASSO regression: An alternative linear regression model with L1 regularization was then employed, with alpha specified as 0.01. Owing to the sparsity-inducing property of the L1 penalty term, this model facilitated the implicit selection of influential predictors within the constructed feature set[21].

RF: Nonlinear residual structures were represented using an ensemble learning model composed of 100 decision trees (n_estimators = 100), with the depth of each individual tree restricted to 8 (max_depth = 8). This configuration allowed the interactions among input variables and nonlinear dependencies to be approximated within the ensemble forecasting framework[22].

LSTM: Temporal dependencies in the residual series were modeled using a recurrent neural network architecture. The implemented network consisted of an input layer matched to the feature dimension, a single LSTM layer containing 64 hidden units, a Dropout layer with a dropout rate of 0.2, a fully connected layer with 32 neurons, and an output layer. The model parameters were optimized using the Adam algorithm (learning rate = 0.001) with a mean squared error loss function over 50 training epochs. This architecture enables the sequential memory effects in hydrological time series to be represented through recurrent state transitions[23].

2.5. Feature Engineering and Model Training

Feature engineering was performed to provide effective information inputs for the machine learning models to predict the residuals[24]. Based on domain knowledge, an eight-dimensional feature set was constructed as follows:

Muskingum simulated discharge $Q_{(musk)}$: The physical baseline discharge at the current time step.

Total upstream inflow in_{sum}: The sum of inflows from the Aksu River and the Yarkant and Hotan Rivers at the current time step.

Lagged downstream observations obs_{lag_1}, obs_{lag_2}, and obs_{lag_4}: Lagged discharges at the Alar Station from 1, 2, and 4 time steps earlier, respectively (during rolling forecasting, true observations were used in the initial step, and forecasts from the previous step were used subsequently).

Lagged residuals res_{lag_1}, res_{lag_2}, and res_{lag_4}: Residual values from 1, 2, and 4 time steps earlier, respectively.

The rationale for selecting these specific features is as follows: the Muskingum simulated discharge and total upstream inflow represent the current physical state of the hydrological system; the lagged observations capture the "memory" of the flow process, which is critical for flood forecasting due to the high autocorrelation of streamflow; and the lagged residuals provide a mechanism for the model to correct systematic biases by leveraging information from previous forecasting errors. These features were chosen based on domain knowledge and have been commonly adopted in hybrid hydrological modeling studies [25]. While other meteorological variables such as precipitation, temperature, and snowmelt indicators could potentially enhance model performance, they were not included in this study due to the unavailability of long-term, high-quality observational data at the required temporal resolution for the study area.

Training procedure: The globally calibrated Muskingum parameters were adopted to compute the baseline simulations ($Q_{(musk)}$) and residual series (res) for the entire training period. The feature engineering procedure was applied to construct the feature matrix (X) and the target vector (res). Both features (X) and target residuals (res) were standardized, and the scalers were retained for use during forecasting. The standardized features and residuals were subsequently used to train the Ridge, LASSO, RF, and LSTM models in parallel.

2.6. Rolling Forecasting Method

Rolling forecasting was adopted to simulate the real-time operational forecasting by ensuring that only information reaching the current time step was utilized, thereby preventing the application of future observations[26].

 doi:10.20944/preprints202604.1577.v1

1. Initialization: At the initial time step of the test dataset, historical observations of upstream inflow and downstream discharge were available.
2. Multi-step rolling procedure: The s-th forecasting (s = 1, 2, …, 6) corresponding to lead times of 4–24 h was utilized as an example.
3. Physical baseline routing: The calibrated Muskingum model was first applied using upstream inflows reaching the current time to generate the baseline discharge $Q_{(musk)}(t)$. In operational practice, these observed inflows should be replaced with upstream inflow forecasts. The downstream discharge sequence was initialized using observations and subsequently updated with the prediction results from the previous step.
4. Feature construction: Based on all information available at the current time step, a multidimensional feature vector was constructed following the same feature engineering rules used during training.
5. Residual forecasting: After standardization, the constructed features were input into the trained machine learning models to obtain the predicted residual $\widehat{res}(t)$.
6. Integrated output: The final discharge forecasting was determined by $\widehat{Q}(t)=Q_{musk}(t)+\widehat{res}(t)$.
7. State update: The predicted discharge $\widehat{Q}(t)$ at the current time step was subsequently stored as an observation for computing lagged features at the next time step, and the residual history sequence was updated accordingly. The procedure was iteratively repeated until the end of the test dataset.

This strategy ensured that only continuous information reaching the corresponding time was used at each lead time step, thereby simulating realistic real-time forecasting conditions and preventing data leakage[27].

3. Results

3.1. Muskingum Method Baseline Model

In a multi-model coupled forecasting framework, the robustness of the physical baseline layer determines the physical plausibility of subsequent residual corrections. Based on 14 representative flood events during 2021–2024, the characteristics of the Muskingum routing parameters, full-sequence transformation behavior, and test-period performance over the Alar reach of the Tarim River main stem were analyzed.

The event-specific calibration results (Table 2) demonstrated that different flood events exhibited pronounced heterogeneity in the Muskingum routing parameters. For most high-flow events (e.g., 2022062700 and 2023090120), the Nash–Sutcliffe efficiency (NSE) remained above 0.85. This indicates that under routing conditions dominated by linear storage–discharge relationships, the traditional Muskingum method can adequately capture the time-lag characteristics of flood waves.

However, certain events (e.g., 2024091016) exhibited negative NSE values, signaling a severe mismatch between the model structure and the actual hydrological processes. These negative metrics do not merely reflect statistical error but represent a structural breakdown of the linear routing hypothesis under extreme conditions. Specifically, the Tarim River's "flashy" floods often trigger complex hydraulic phenomena, such as overbank flow and floodplain inundation, which fundamentally alter the storage-attenuation relationship (K and x cease to be constants). Furthermore, the 2024 event was characterized by intense, localized cryospheric melting coupled with unexpected upstream dam regulations, creating a non-stationary hydrological anomaly that violates the mass balance assumptions of the standard Muskingum equation. The failure of the physical model to track these abrupt shifts justifies the necessity of the machine learning-based residual module, which is designed to compensate for such mechanistic deficiencies by learning the non-linear "memory" of these anomalies.

 doi:10.20944/preprints202604.1577.v1

Table 2. Local Muskingum parameter calibration results.

Event ID	K1_local	x1_local	K2_local	x2_local	RMSE	NSE	PE
2021070916	7.98	0.01	44.93	0.49	38.96	0.77	−6.61
2021071916	17.21	0.49	4.00	0.01	100.66	0.83	−2.57
2022052908	24.10	0.30	72.00	0.01	32.83	0.88	−2.05
2022061008	13.49	0.04	22.85	0.01	48.54	0.82	−7.17
2022062700	18.74	0.26	20.86	0.01	72.06	0.92	−10.35
2022091004	37.45	0.11	72.00	0.01	6.23	0.89	−1.49
2023071712	72.00	0.02	34.45	0.45	30.25	0.03	−7.12
2023072412	20.09	0.01	9.38	0.09	73.94	0.77	−8.13
2023082312	4.00	0.01	4.00	0.01	75.64	0.63	−1.29
2023082912	30.94	0.31	8.65	0.49	13.26	0.92	0.42
2023090120	36.47	0.26	37.21	0.49	16.69	0.94	−2.3
2024071008	15.61	0.01	4.37	0.49	47.09	0.77	6.08
2024071620	14.70	0.37	7.71	0.01	110.51	0.82	−1.45
2024091016	20.36	0.22	17.27	0.01	15.08	−0.20	−2.44

To avoid bias introduced by simple parameter averaging, the global parameter optimization over the entire training period (2021–2024) was performed using the L-BFGS-B algorithm to extract representative parameters reflecting the intrinsic physical properties of the river reach. For Tributary 1 (Yimapashe), the optimized parameters were K = 17.37 and x = 0.31, indicating routing characteristics typical of natural channels. The relatively high x value suggests a greater weighting of inflow on channel storage, implying strong regulation and storage capacity along this reach. In contrast, Tributary 2 (Xiaota) exhibited a smaller K value of 8.24 h, with the corresponding x value approaching the lower physical boundary of 0.01. This parameter combination demonstrated strong hydraulic connectivity, rapid flood-wave propagation, and minimal wave deformation along the Xiaota–Alar reach, with channel regulation effects primarily manifested as reservoir-type storage–release behavior. Such differentiated parameter identification reflected the pronounced differences among upstream tributaries in terms of channel slope and cross-sectional morphology.

The application of global parameters to the 2025 test period (Figure 2) indicated that the Muskingum model performed well in phase capture, accurately detecting the rising and turning points of multiple flood peaks. However, a systematic underestimation in magnitude was observed, particularly in the high-flow range above 1500 m^3/s, where the observed discharges were significantly higher than model predictions. The analysis of the hydrographs of observed discharge, model-predicted discharge, and residuals indicated that the residual series exhibited significant linear correlations with both observed and simulated discharges. Residuals increased during high-flow periods and decreased during low-flow periods, suggesting that the Muskingum model exhibited a magnitude-dependent proportional bias. This residual structure can provide physical validation for the subsequent application of linear machine learning models for efficient correction.

 doi:10.20944/preprints202604.1577.v1

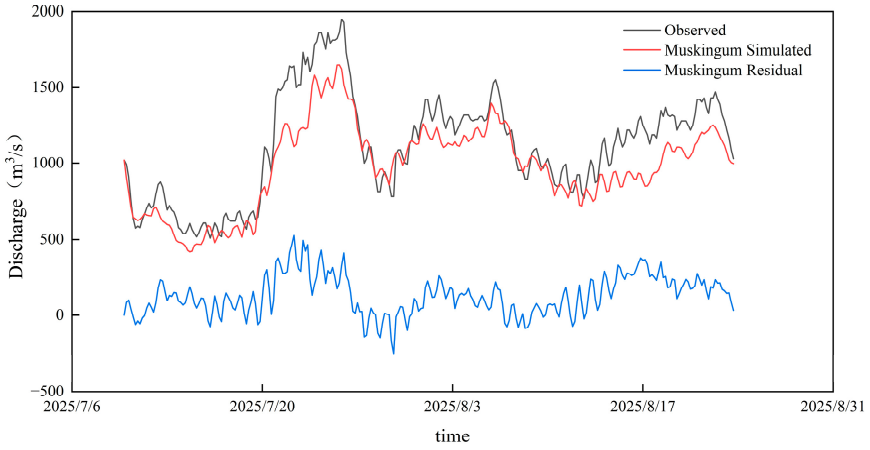
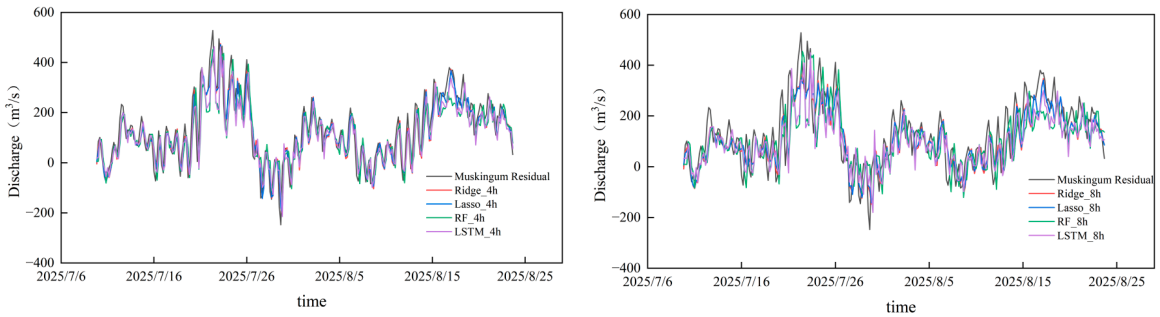


Figure 2. Discharge comparison using the Muskingum method.

3.2. Comparative Analysis of Residual Forecasting

The residual forecasting performance of four machine learning models, including Ridge, LASSO, random forest (RF), and LSTM, was evaluated under different lead times ranging from 4 to 24 h.

At short lead times (4 and 8 h), all four models exhibited excellent residual-capturing capability, with the predicted residual curves showing a high degree of overlap with the original residual series derived from the Muskingum simulations (Figure 3). This indicates that these models can effectively identify residual features not captured by the physical routing model. As the lead time increased (12–24 h), the forecasting difficulty increased markedly. Linear models (Ridge and LASSO) demonstrated strong robustness, maintaining stable overall trends despite reduced correction performance at extreme peaks. As illustrated in Figure 2 by the hydrographs of streamflow and residuals, the Muskingum residuals exhibit a certain linear relationship with the observed streamflow. The Ridge regression model provides a stable and well-regularized fit to the short-term linear correlation between residuals and current features, which underpins its superior performance at short lead times. The advantage of the LASSO model at longer lead times stems from its L1 regularization, which enables effective feature selection by retaining only the most relevant lagged information, thereby reducing error accumulation and mitigating overfitting. In contrast, the nonlinear RF model exhibited phase deviations in residual forecasting at medium-to-long lead times, with predicted peaks deviating from observed residual peaks. For the LSTM model, pronounced high-frequency oscillations appeared in the predicted curves at longer lead times (20 and 24 h), which may be attributed to error accumulation during rolling forecasting. Because LSTM relies on the predictions from the previous step as inputs, small deviations at one step may be amplified through nonlinear activation functions during recursive propagation, resulting in numerically unstable oscillations lacking physical interpretability. Overall, while all models achieved good residual fitting performance at short lead times, traditional linear regularized models exhibited a higher practical value for medium-to-long lead time flood forecasting beyond 8 h owing to their numerical stability.



(**a**) (**b**)

 doi:10.20944/preprints202604.1577.v1

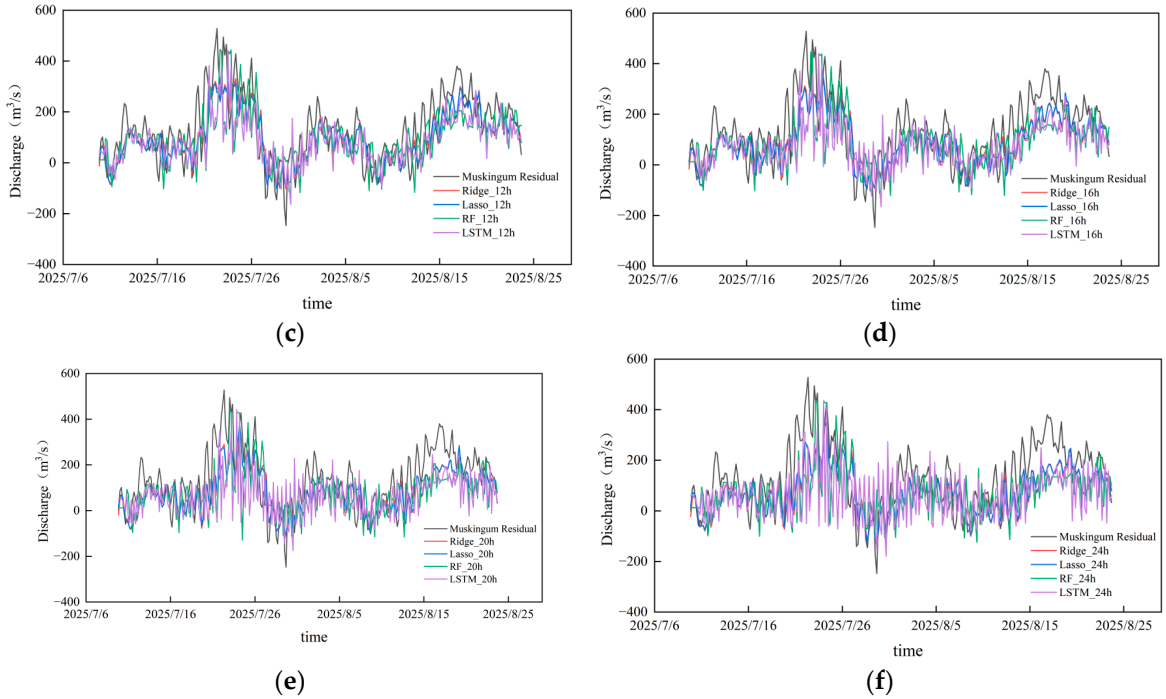


Figure 3. Residual comparison across different lead times: (a) 4-hour lead time; (b) 8-hour lead time; (c) 12-hour lead time; (d) 16-hour lead time; (e) 20-hour lead time; (f) 24-hour lead time. The black line represents the original residual series derived from Muskingum simulations, and the colored lines represent the residuals predicted by different machine learning models.

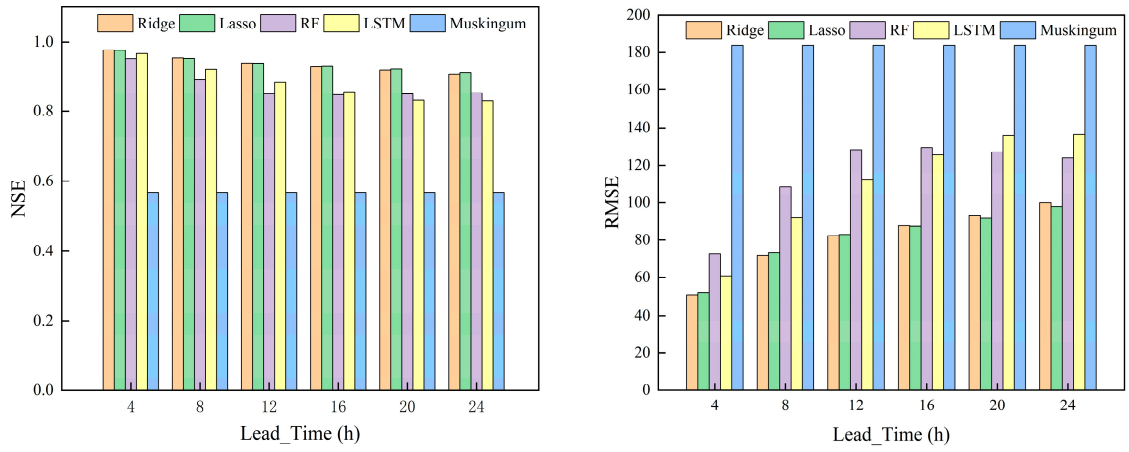
3.3. Comprehensive Performance Evaluation of Models

To comprehensively evaluate the forecasting performance of the physics-guided residual compensation hybrid framework, the overall performance of different models was compared across lead times ranging from 4 to 24 h in three dimensions, including overall fitting accuracy, error distribution, and hydrological characteristics.

3.3.1. Overall Accuracy and Error Analysis

As shown in Figure 4, all hybrid models significantly outperformed the traditional Muskingum method in terms of NSE and RMSE (baseline NSE = 0.567 and RMSE = 183.94 m^3/s). As the lead time increased, the model performance exhibited a systematic degradation pattern, with NSE gradually decreasing and RMSE correspondingly increasing. Linear regularized models (Ridge and LASSO) demonstrated excellent robustness and remained leading performers across all lead times. At the short lead time of 4 h, the Ridge model achieved the best performance (NSE = 0.977 and RMSE = 51.04 m^3/s). At the long lead time of 24 h, the LASSO model presented the best performance (NSE = 0.911 and RMSE = 97.97 m^3/s). In contrast, the performance of nonlinear models (RF and LSTM) declined more markedly at longer lead times. For the 24 h lead time, the NSE of the LSTM model decreased to 0.829, whereas the RMSE increased to 136.43 m^3/s.

 doi:10.20944/preprints202604.1577.v1



(**a**) (**b**)

Figure 4. Comparison of accuracy characteristics among different models: (a) NSE for lead times of 4 – 24 h; (b) RMSE for lead times of 4 – 24 h.

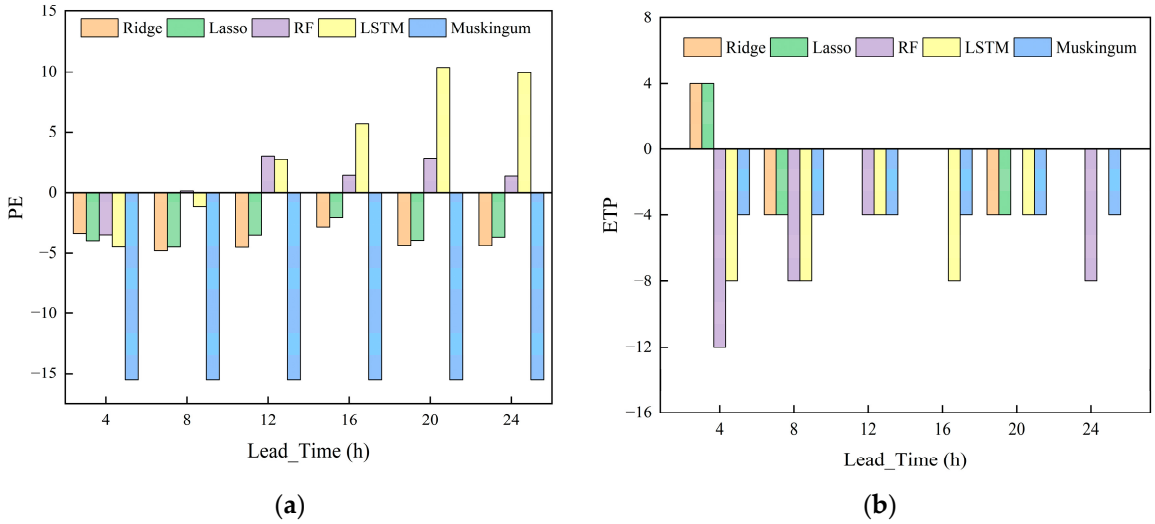
3.3.2. Comparison of Hydrological Feature Forecasting Capability

With respect to flood peak forecasting, the hybrid models substantially reduced the systematic underestimation observed in the Muskingum simulations (PE = −15.55%) (Figure 5). Peak discharge errors produced by linear models remained constrained within −5% across all lead times, with only minor fluctuations as forecasting horizons increased, indicating strong correction stability. Among the nonlinear approaches, the LSTM model exhibited a transition from underestimation to overestimation at longer lead times, resulting in a peak overestimation of 10.38% at the 20 h lead time. In comparison, the RF model produced near-neutral peak magnitude errors over medium-to-long lead times (12–24 h) but with reduced stability.

The evaluation of peak timing errors further revealed critical differences in the models' phase-capturing capabilities, which carry profound implications for operational hydraulic safety. The linear models maintained consistent deviations within ±4 h across all horizons. From a physical perspective, this stability indicates that linear architectures effectively preserve the kinematic characteristics of the flood wave established by the Muskingum baseline. In arid river channels like the Tarim, the flood wave propagation speed is primarily governed by the river's gradient and geometry; by applying a constrained linear scaling, the models ensure that the hydrodynamic momentum and translation time remain physically consistent.

In contrast, the nonlinear models (LSTM and RF) exhibited larger and more irregular fluctuations in peak timing. This phenomenon, often termed "phase drift," poses a significant risk in real-time flood management. While deep learning models excel at minimizing statistical residuals in discharge magnitude, their lack of explicit physical constraints can lead to a decoupling of the flow-velocity relationship during multi-step rolling forecasts. In practical engineering applications, a "magnitude-accurate but phase-shifted" forecast can be more hazardous than a slight magnitude error, as it may mislead the timing of critical interventions—such as the operation of diversion sluices or the reinforcement of dikes. The superior phase stability of the linear hybrid framework thus provides more actionable intelligence for downstream emergency response, ensuring that flood defense measures are synchronized with the physical arrival of the wave front.

 doi:10.20944/preprints202604.1577.v1



(a) (b)

Figure 5. Comparison of hydrological characteristics among different models: (a) PE for lead times of 4 – 24 h; (b) ETP for lead times of 4 – 24 h.

3.4. Optimal Model Analysis

Based on the comprehensive performance evaluation, the optimal model selection varied across forecasting horizons (Table 2). For short lead times (4–12 h), the Ridge model achieved the best overall performance, with NSE improvement rates of 65.65–72.33%, RMSE reductions of 55.26–72.25%, and flood peak error improvements of 69.13–78.14%. As the lead time increased to 16–24 h, the LASSO model became the preferred configuration owing to its inherent feature selection capability. Despite the performance degradation at extended horizons, the LASSO model maintained an NSE improvement of 60.70% and a flood peak error reduction of 76.33% at the 24 h lead time. Notably, at the 16 h lead time, it achieved a peak error improvement of 86.75%, representing the highest correction efficiency among all lead times and highlighting its effectiveness in capturing key signals under complex, long-lead forecasting conditions.

Table 2. Performance improvement summary.

Lead time	Optimal model	Muskingum NSE	Optimal model NSE	Muskingum RMSE (m^3/s)	Optimal model RMSE (m^3/s)	Muskingum PE (%)	Optimal model PE (%)
4 h	Ridge	0.567	0.977	183.94	51.04	−15.55	−3.4
8 h	Ridge	0.567	0.954	183.94	71.9	−15.55	−4.8
12 h	Ridge	0.567	0.94	183.94	82.3	−15.55	−4.52
16 h	Lasso	0.567	0.932	183.94	87.37	−15.55	−2.06
20 h	Lasso	0.567	0.924	183.94	91.58	−15.55	−3.94
24 h	Lasso	0.567	0.911	183.94	97.97	−15.55	−3.68

The comprehensive evaluation results indicated that linear regularized models achieved the most favorable balance among forecasting accuracy, numerical stability, and interpretability. Accordingly, a lead-time-dependent model selection strategy is recommended for practical operational forecasting. Specifically, the Ridge model was applied for lead times within 12 h, whereas the LASSO model was implemented for lead times of 16 h and beyond to ensure an appropriate trade-off between predictive accuracy and robustness. A comparison of discharge forecasts produced by the selected optimal models across different lead times is presented in Figure 6.

 doi:10.20944/preprints202604.1577.v1

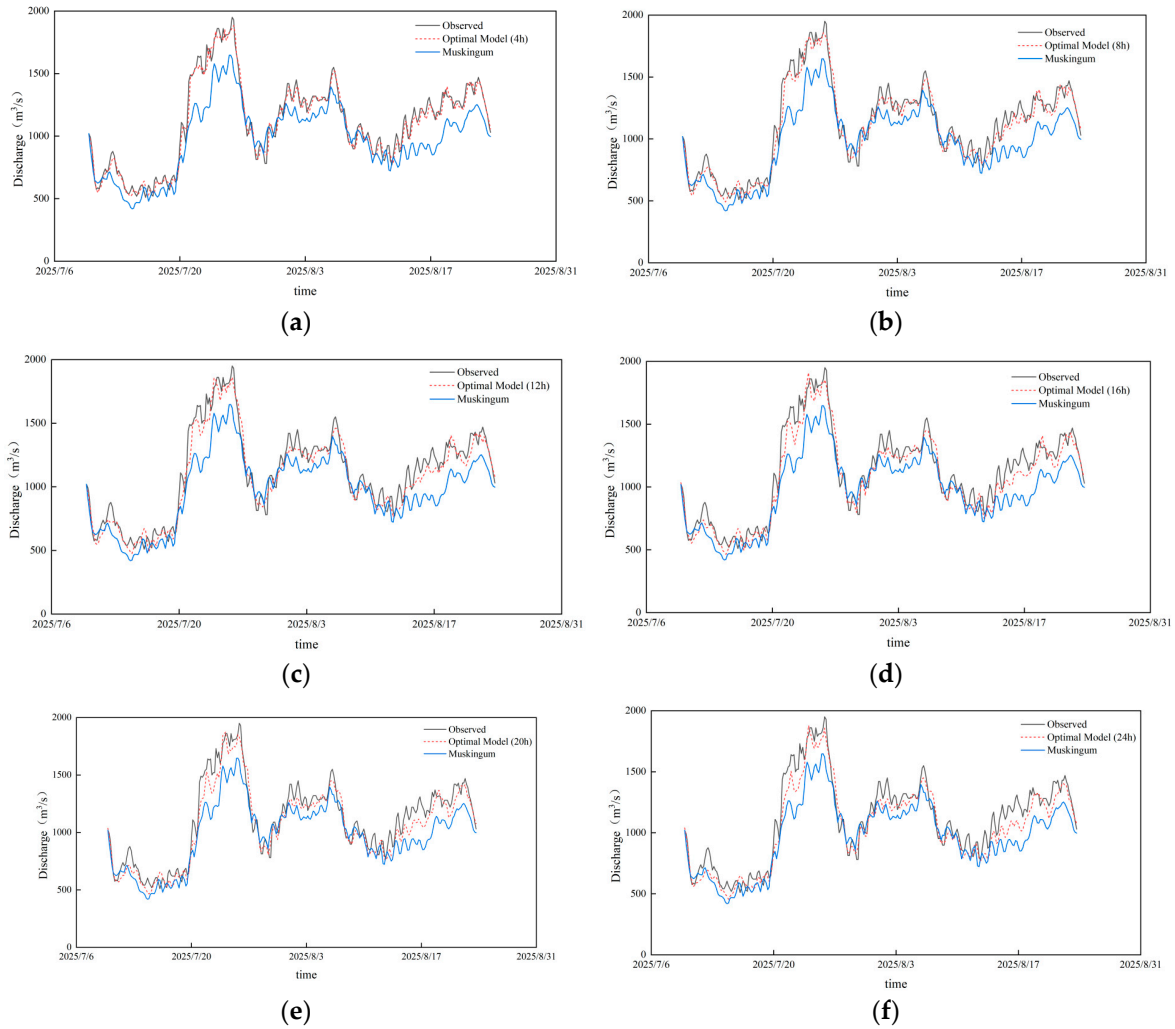


Figure 6. Discharge hydrographs predicted by the optimal models for lead times of 4–24 h: (a) 4 h; (b) 8 h; (c) 12 h; (d) 16 h; (e) 20 h; (f) 24 h. The black line represents observed discharge, the blue line represents Muskingum simulations, and the red dashed line represents forecasts by the optimal hybrid models.

4. Discussion

4.1. Comparison with Other Decision Methods

The hybrid Muskingum-machine learning framework proposed in this study achieved significantly improved forecasting performance over traditional methods in the Tarim River Basin, with linear regularized models (Ridge, LASSO) outperforming nonlinear models (RF, LSTM) across all lead times. This finding contrasts with studies conducted in the Clear Creek and Upper Wapsipinicon River watersheds in Iowa, where deep learning models such as LSTM significantly outperformed linear models, achieving NSE values of 0.85–0.93 for 24-hour lead time predictions[12]. The underlying reason for this difference lies in the distinct hydrological characteristics of the two basins: in the Tarim River Basin, the residuals of the Muskingum model exhibit a strong linear correlation with discharge, whereas the rainfall-runoff relationship in the Iowa watersheds displays stronger nonlinear features, making deep learning models more suitable for capturing such patterns.

In the study conducted at the Boluo station in the Dongjiang River Basin, LSTM also achieved optimal performance (NSE = 0.95), while linear models such as LR and tree-based models like ETR ranked lower (with NSE values as low as 0.89)[13]. The competitive performance of linear models in that study was attributed to the high autocorrelation of daily streamflow time series and the effective linearization of the prediction problem through the incorporation of antecedent flow variables. These results indicate that the choice of machine learning model for residual correction should not be

 doi:10.20944/preprints202604.1577.v1

determined by model complexity alone, but rather by the structural characteristics of the residuals themselves. In the Tarim River Basin, where residuals are predominantly linear, simple linear models are sufficient and even superior; in basins where residuals exhibit complex nonlinear dynamics, deep learning architectures offer greater advantages. Furthermore, appropriate feature engineering can, to some extent, "linearize" nonlinear problems, thereby reducing the demand for model complexity.

4.2. Research Advantages and Implications

The main advantages and implications of this study are reflected in the following four aspects.

(1) The linear structural characteristics of Muskingum residuals are revealed, providing a theoretical basis for model selection. The study found that the residuals of the Muskingum model in the Tarim River Basin exhibit a strong linear correlation with observed discharge, indicating that the systematic bias of the traditional model is primarily a proportional bias related to flow magnitude rather than random noise. This finding provides solid physical support for using simple linear models for residual correction.

(2) It is demonstrated that simple linear models outperform complex models in multi-step rolling forecasting, challenging conventional wisdom. In multi-step rolling forecasts with lead times ranging from 4 to 24 hours, Ridge regression and LASSO consistently outperformed LSTM and Random Forest across all lead times, achieving an NSE of 0.911 at the 24-hour lead time and a peak flow error improvement rate of 76.33%. This result indicates that increasing model complexity does not necessarily lead to improved performance; rather, the key lies in the linear structure of the residuals themselves.

(3) A lead-time-dependent adaptive model selection strategy is proposed, enhancing operational practicality. Based on systematic evaluation of model performance, a dynamic strategy is proposed: Ridge regression is adopted for short lead times (≤12 h), while LASSO is used for long lead times (≥16 h), providing a clear and actionable decision-making framework for operational forecasting.

(4) A novel hybrid modeling approach is constructed, with the Muskingum physical model as the core and machine learning as a supplementary component, balancing accuracy and interpretability. The framework first establishes a stable physical baseline using the Muskingum method, and then corrects residuals through lightweight linear models. This approach significantly improves forecasting accuracy while preserving the physical interpretability of the hydrological process, and its high computational efficiency makes it suitable for real-time flood warning applications in data-scarce arid regions.

From an operational implementation perspective, the proposed framework can be readily deployed by water management authorities in the Tarim River Basin or similar arid regions. The computational efficiency of the linear models (Ridge and LASSO) enables forecast updates to be generated in seconds, meeting the latency requirements of real-time warning systems. The lead-time-dependent model selection rule provides forecasters with a clear, interpretable decision criterion—Ridge for short leads, LASSO for longer leads—eliminating the need for subjective parameter tuning. Furthermore, the framework relies only on streamflow data from upstream and downstream stations, which are routinely monitored, avoiding dependence on high-resolution meteorological data that may be unavailable in data-scarce regions. This combination of low data requirements, computational simplicity, and operational interpretability positions the framework as a practical tool for enhancing flood early warning capabilities in arid inland river basins. This methodology offers a transferable reference for hybrid hydrological modeling in other basins.

4.3. Limitations and Future Directions

Despite the significant forecasting performance achieved in this study, several limitations remain:

A notable limitation concerns the generalizability of the results, as the test set includes only one flood event, from July to August 2025. Although this event contained multiple independent flood peaks, with a maximum flow of 1970 m^3/s, representing a typical multi-peak complex flood event, a

 doi:10.20944/preprints202604.1577.v1

single test event cannot fully represent all possible hydrological conditions within the basin. Future research should evaluate the framework's performance across a wider range of flood types and magnitudes to further establish its robustness and generalizability.

Another limitation relates to the data resampling process, which may have some influence on the results. The original dataset comprised two temporal resolutions: 4-hour intervals and daily intervals, with daily-scale data accounting for approximately 52–53% of the dataset, which required conversion to 4-hour intervals using the PCHIP interpolation method. Although PCHIP preserves the monotonicity of the runoff series and minimizes interpolation-induced oscillations, and missing data accounted for less than 0.23%, the interpolation process itself may still have minor effects on flood peak timing and magnitude. Future research could consider using higher temporal resolution raw data for validation.

The feature set employed in this study also presents a limitation, as it is primarily based on lagged terms of streamflow and residuals, and does not incorporate meteorological and hydrological variables such as precipitation, temperature, or snowmelt indicators. Although snow and ice melt constitute an important component of runoff in the Tarim River Basin, this study was unable to directly utilize temperature or snowmelt indices as input features due to the difficulty of data acquisition. Future research could consider integrating multi-source real-time meteorological and remote sensing data to further enhance the interpretability of the model.

Finally, the conclusions of this study are necessarily context-specific, as they are based on the hydrological conditions of the Tarim River Basin, where the residuals exhibit a strong linear correlation. In other basins with different residual structures, the conclusion that "linear models outperform complex models" may not be applicable. Therefore, when applying this framework to other basins, it is essential to first analyze the structural characteristics of the residuals before selecting the appropriate machine learning algorithm.

5. Conclusions

This study developed a hybrid flood forecasting framework by coupling the Muskingum method with machine learning models. The core scientific findings and their broader implications are summarized as follows.

(1) A robust physical baseline is established using L-BFGS-B-based global calibration. By optimizing Muskingum parameters across the entire training period rather than event-specifically, the framework captures the intrinsic routing characteristics of the Tarim River's "three sources and one main stream" system. This physically consistent baseline, combined with upstream multi-station inflow data, provides a reliable foundation for subsequent residual correction.

(2) The residuals of the traditional Muskingum model in this basin exhibit a strong linear correlation with observed discharge. Analysis reveals that the systematic bias is predominantly proportional to flow magnitude rather than random noise. This structural characteristic is the key to understanding why simple linear models suffice for correction.

(3) Simple linear regularized models (Ridge, LASSO) consistently outperform complex nonlinear models (RF, LSTM) in multi-step rolling forecasting. The superior performance of linear models—achieving NSE values above 0.91 even at 24-hour lead times—indirectly confirms the linear nature of the residual structure. This finding challenges the conventional assumption that increasing model complexity always yields better forecasting results.

(4) The optimal machine learning model varies across basins depending on residual structure. The contrasting performance of linear versus deep learning models in the Tarim River Basin versus other watersheds (e.g., Iowa, Dongjiang) underscores a key principle: model selection should be guided by the structural characteristics of the residuals, not by model complexity alone. This principle offers a transferable reference for hybrid hydrological modeling in other regions.

In summary, the proposed framework demonstrates that by first establishing a physically robust baseline and then selecting a machine learning algorithm based on residual structure, accurate and interpretable flood forecasting can be achieved even in data-scarce arid regions. The framework's

 doi:10.20944/preprints202604.1577.v1

lead-time-dependent strategy further enhances operational practicality, providing a clear decision rule for real-time deployment by water management authorities.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: P.W.: conceptualization, methodology, formal analysis, writing—original draft and editing. L.Z. and D.L.: supervision, validation, and writing—review. F.T.: resources and data curation. X.W.:Software and investigation. Y.W.: project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: The work was funded by the Natural Science Foundation of Henan Province (252300421559), the Zhongyuan Leading Talents Funding Program for Scientific and Technological Innovation (254000510046), and the Special Project for Basic Research Business Fees of the Yellow River Institute of Hydraulic Research (HKY-JBYW-2024-01).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors sincerely thank the Tarim River Basin Administration for providing the hydrological data essential to this study. Additional support was provided by the Yellow River Institute of Hydraulic Research, Yellow River Conservancy Commission, and the Yellow River Laboratory.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
L-BFGS-B	Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bound Constraints
LSTM	Long Short-Term Memory
RF	Random Forest
NSE	Nash–Sutcliffe Efficiency
RMSE	Root Mean Square Error
PE	Peak Error
ETP	Error in Time to Peak

References

1. Portalés-Julià, E.; Mateo-García, G.; Purcell, C.; Gómez-Chova, L. Global flood extent segmentation in optical satellite images. *Scientific reports* **2023**, *13* (1), 20316.
2. Meresa, H.; Tischbein, B.; Mekonnen, T. Climate change impact on extreme precipitation and peak flood magnitude and frequency: observations from CMIP6 and hydrological models. *Natural Hazards* **2022**, *111* (3), 2649-2679.
3. Nevo, S.; Morin, E.; Gerzi Rosenthal, A.; Metzger, A.; Barshai, C.; Weitzner, D.; Voloshin, D.; Kratzert, F.; Elidan, G.; Dror, G. Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences* **2022**, *26* (15), 4013-4032.
4. Salvati, A.; Moghaddam Nia, A.; Salajegheh, A.; Shirzadi, A.; Shahabi, H.; Ahmadisharaf, E.; Han, D.; Clague, J. J. A systematic review of Muskingum flood routing techniques. *Hydrological Sciences Journal* **2024**, *69* (6), 810-831.
5. Bindas, T.; Tsai, W. P.; Liu, J.; Rahmani, F.; Feng, D.; Bian, Y.; Lawson, K.; Shen, C. Improving river routing using a differentiable Muskingum-Cunge model and physics-informed machine learning. *Water Resources Research* **2024**, *60* (1), e2023WR035337.
6. Yang, D.; Yan, B.; Gu, D.; Chang, J.; Du, S. A hybrid deep learning-Muskingum framework for enhanced runoff prediction: Model coupling and hydrological process integration. *Journal of Hydrology: Regional Studies* **2026**, *63*, 103077.

 doi:10.20944/preprints202604.1577.v1

7. Gauch, M.; Kratzert, F.; Klotz, D.; Nearing, G.; Lin, J.; Hochreiter, S. Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network. *Hydrology and Earth System Sciences* **2021**, *25* (4), 2045-2062.
8. Klotz, D.; Kratzert, F.; Gauch, M.; Keefe Sampson, A.; Brandstetter, J.; Klambauer, G.; Hochreiter, S.; Nearing, G. Uncertainty estimation with deep learning for rainfall–runoff modeling. *Hydrology and Earth System Sciences* **2022**, *26* (6), 1673-1693.
9. Akbari, R.; Hessami-Kermani, M.-R.; Shojaee, S. Flood routing: improving outflow using a new non-linear Muskingum model with four variable parameters coupled with PSO-GA algorithm. *Water Resources Management* **2020**, *34* (10), 3291-3316.
10. Lin, Y.; Wang, D.; Wang, G.; Qiu, J.; Long, K.; Du, Y.; Xie, H.; Wei, Z.; Shangguan, W.; Dai, Y. A hybrid deep learning algorithm and its application to streamflow prediction. *Journal of Hydrology* **2021**, *601*, 126636.
11. Li, W.; Liu, C.; Xu, Y.; Niu, C.; Li, R.; Li, M.; Hu, C.; Tian, L. An interpretable hybrid deep learning model for flood forecasting based on Transformer and LSTM. *Journal of Hydrology: Regional Studies* **2024**, *54*, 101873.
12. Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water resources research* **2020**, *56* (1), e2019WR025326.
13. Zhang, Z.; Xiao, Y.; Chen, R.; Long, K.; Deng, H.; Zheng, Z.; Liao, J. Comparative assessment of machine learning models for daily streamflow prediction in a subtropical monsoon watershed. *Scientific Reports* **2026**.
14. Fang, G.; Li, Z.; Yang, J.; Chen, Y.; Duan, W.; Amory, C.; Wang, Y. Changes in flooding in the alpine catchments of the Tarim River Basin, Central Asia. *Journal of Flood Risk Management* **2023**, *16* (1), e12869.
15. Qianjuan, S.; Hongbo, L.; Yong, L.; Guangpeng, Z.; Lianchao, Z.; Aihua, L. Extreme climate and human interventions driving flood dynamics in arid inland rivers: regional and global perspectives. *Journal of Environmental Management* **2025**, *395*, 127788.
16. Fritsch, F. N.; Carlson, R. E. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis* **1980**, *17* (2), 238-246.
17. Frame, J. M.; Kratzert, F.; Raney, A.; Rahman, M.; Salas, F. R.; Nearing, G. S. Post-processing the national water model with long short-term memory networks for streamflow predictions and model diagnostics. *JAWRA Journal of the American Water Resources Association* **2021**, *57* (6), 885-905.
18. Farzin, S.; Singh, V. P.; Karami, H.; Farahani, N.; Ehteram, M.; Kisi, O.; Allawi, M. F.; Mohd, N. S.; El-Shafie, A. Flood routing in river reaches using a three-parameter Muskingum model coupled with an improved bat algorithm. *Water* **2018**, *10* (9), 1130.
19. Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing* **1995**, *16* (5), 1190-1208.
20. Khorram, S.; Jehbez, N. Improving the streamflow prediction accuracy in sparse data regions: a fresh perspective on integrated hydrological-hydrodynamic and hybrid machine learning models. *Engineering Applications of Computational Fluid Mechanics* **2024**, *18* (1), 2387051.
21. Kumar, V.; Kedam, N.; Sharma, K. V.; Mehta, D. J.; Caloiero, T. Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. *Water* **2023**, *15* (14), 2572.
22. Wahba, M.; Essam, R.; El-Rawy, M.; Al-Arifi, N.; Abdalla, F.; Elsadek, W. M. Forecasting of flash flood susceptibility mapping using random forest regression model and geographic information systems. *Heliyon* **2024**, *10* (13).
23. Kumar, V.; Azamathulla, H. M.; Sharma, K. V.; Mehta, D. J.; Maharaj, K. T. The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. *Sustainability* **2023**, *15* (13), 10543.
24. Galelli, S.; Castelletti, A. Tree-based iterative input variable selection for hydrological modeling. *Water Resources Research* **2013**, *49* (7), 4295-4310.
25. Talei, A.; Chua, L. H. Influence of lag time on event-based rainfall–runoff modeling using the data driven approach. *Journal of hydrology* **2012**, *438*, 223-233.
26. Liang, Z.; Tang, T.; Li, B.; Liu, T.; Wang, J.; Hu, Y. Long-term streamflow forecasting using SWAT through the integration of the random forests precipitation generator: case study of Danjiangkou Reservoir. *Hydrology Research* **2018**, *49* (5), 1513-1527.

 doi:10.20944/preprints202604.1577.v1

27. He, M.; Xu, X.; Wu, S.; Kang, C.; Huang, B. Multi-step ahead forecasting of daily streamflow based on the transform-based deep learning model under different scenarios. *Scientific Reports* **2025**, *15* (1), 5451.