Predictive Modelling of Flood Dynamics in Malaysia's East Coast Using an NARX Model

Nur Nabilah Zakaria[1], Azlee Zabidi[2]*, Mahmood Alsaadi[3], Mohd Izham Mohd Jaya[4]
Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Pekan, Pahang, Malaysia[1, 2, 4]
Department of Computer Science, University of Al-Maarif, Al Anbar, 31001, Iraq[3]

Abstract—**Flood forecasting is critical for improving early warning systems in Malaysia's East Coast region, particularly in flood-prone Pekan. This study develops a Nonlinear Autoregressive with Exogenous Inputs (NARX) model to predict river water levels using data from four stations: Sungai Pahang, Sungai Pahang Tua, Sungai Paloh Hinai, and Sungai Mentiga (2020–2024). The dataset was preprocessed through short-gap interpolation, removal of long missing segments, and segmentation into continuous sequences to ensure high-quality inputs for modeling. A total of 75 NARX configurations were evaluated using different lag values, hidden neuron counts, and training epochs. Model performance was assessed using Mean Squared Error (MSE) and residual diagnostics. The best model—lag = 6 and 300 hidden units—achieved a validation loss of 0.102, demonstrating stable convergence and strong generalization. Prediction results showed close alignment with actual river levels. The findings confirm that the NARX approach effectively captures nonlinear hydrological dynamics and provides reliable short-term water level forecasts for Pekan, addressing an existing gap in localized flood prediction studies.**

Keywords—Flood prediction; NARX model; hydrological modelling; Pekan

I. Introduction

The East Coast region of Peninsular Malaysia frequently experiences heavy rainfall during the Northeast Monsoon season, which often leads to flooding in low-lying and riverine areas. These flood events result in significant damage to infrastructure, disruption of transportation systems, and substantial economic losses for affected communities. Accurate and timely flood forecasting is crucial to support early warning systems and minimize flood-related impacts. Reliable river water level prediction enables authorities to make proactive decisions and implement preventive measures to safeguard lives and reduce property damage.

Despite ongoing advancements in hydrological monitoring, flood prediction in Malaysia remains challenging due to the nonlinear, dynamic, and spatially heterogeneous nature of river systems, particularly in monsoon-driven environments. These challenges are more pronounced at the district level, where complex upstream–downstream interactions and data continuity issues limit the effectiveness of generic basin-scale forecasting models.

In recent years, data-driven modelling has gained increasing attention for flood forecasting due to its ability to capture nonlinear and complex relationships within hydrological systems. This study aims to develop a predictive model using the Nonlinear Autoregressive model with Exogenous Inputs (NARX) for river water level forecasting in the East Coast region of Peninsular Malaysia. The NARX model utilizes historical hydrological data to learn temporal patterns, enabling more accurate predictions. By implementing the NARX-based approach, this research seeks to improve the accuracy and reliability of flood forecasting systems to enhance early warning capabilities and reduce flood-related risks.

Although NARX models have been widely applied in hydrological forecasting, most existing studies emphasize major river basins or regional-scale analysis, with limited focus on district-level river systems experiencing recurrent flooding. In particular, the Pekan district in Pahang—despite its high flood vulnerability—has received comparatively little attention in terms of localized, data-driven river level prediction models. Furthermore, prior work rarely investigates how multiple upstream river stations collectively influence downstream water level dynamics within an integrated modelling framework.

This study presents a systematically optimized NARX-based river water level prediction model tailored for the Pekan river system. The model integrates water level data from multiple upstream and downstream stations, including Sungai Pahang Tua, Sungai Paloh Hinai, and Sungai Mentiga, to predict downstream levels at Sungai Pahang (Pekan). A structured preprocessing and segmentation strategy is employed to ensure model training is conducted only on continuous, fully observed temporal segments, thereby preserving hydrological realism and avoiding artefacts introduced by long-gap imputation. In addition, a comprehensive exploration of lag length, hidden layer capacity, and training duration is performed to identify an optimal NARX configuration that balances prediction accuracy, convergence stability, and generalization performance. Through this approach, the study provides a localized and transferable modelling framework that supports district-level flood early warning applications.

The remainder of this study is organized as follows: Section II reviews related work on flood forecasting and data-driven hydrological modelling. Section III describes the dataset, preprocessing strategy, and NARX model architecture. Section IV presents the experimental results and discussion. Finally, Section V concludes the study and outlines limitations and directions for future work.

*Corresponding author.

II. Literature Review

Flood disasters remain a major environmental and socio-economic concern in many regions, particularly in Malaysia, where seasonal monsoon patterns heavily influence hydrological conditions [1]. Researchers have employed various forecasting approaches to understand and predict flood behavior, ranging from traditional statistical and hydrological models to more recent machine learning and deep learning techniques. This literature review provides an overview of flooding characteristics in the Pantai Timur region, followed by a review of conventional flood forecasting methods. It then discusses the evolution of data-driven methods, including Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Random Forest (RF), before examining advanced deep learning techniques such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and hybrid intelligent models. Finally, this chapter emphasizes the significance of the Nonlinear Autoregressive with Exogenous Input (NARX) model in hydrological forecasting and identifies the research gap motivating the present study.

A. Flooding in Malaysia and the Pantai Timur Region

Malaysia experiences frequent flooding due to its tropical climate and high-intensity rainfall patterns. The Pantai Timur region — consisting of Kelantan, Terengganu, and Pahang — is the most flood-prone zone in Peninsular Malaysia, particularly during the Northeast Monsoon season [1]. The region's river basins, low-lying settlements, and coastal geography increase vulnerability to overflow and prolonged inundation [2]. The Northeast Monsoon (NEM) typically occurs between November and March, bringing strong winds and heavy rainfall from the South China Sea [3]. The Malaysian Meteorological Department (2023) states that this seasonal event produces prolonged rainfall that causes river water levels to rise rapidly, triggering severe flooding across Pantai Timur [4]. The intensity and duration of monsoon rainfall directly affect river discharge, making water level prediction a crucial preventive measure. Several major flood events have occurred in Pantai Timur over the past decades, including those in Kelantan (2014) [5], Pahang (2021) [6], and Terengganu (2022) [7]. The 2014 Kelantan flood, widely referred to as "Bah Kuning" [8], affected more than half a million people, resulting in extensive property damage and infrastructure collapse. Such recurring events highlight the importance of accurate forecasting systems to mitigate disaster impacts.

B. Importance of River Level Forecasting

River level forecasting plays a crucial role in early warning systems, evacuation planning, and resource mobilization. Government agencies such as the Department of Irrigation and Drainage (DID) and the National Disaster Management Agency (NADMA) rely on real-time hydrological forecasting for decision-making [9]. Therefore, forecasting models must be accurate, reliable, and capable of capturing nonlinear hydrological behaviors.

C. Conventional Flood Forecasting Methods

1) Statistical time-series models (ARIMA, Regeression): The Autoregressive Integrated Moving Average (ARIMA) model is one of the most widely used statistical methods in river flow forecasting [10]. However, ARIMA assumes linearity and struggles to model complex nonlinear patterns that often characterize hydrological systems [11]. Regression-based models face similar limitations, particularly when dealing with high-variability rainfall-runoff relationships.

2) Physically-based hydrological models (SWAT, HEC-HMS): Physically-based models such as Soil and Water Assessment Tool (SWAT) and Hydrologic Engineering Center – Hydrologic Modeling System (HEC-HMS) simulate watershed hydrology using environmental and land-use parameters [12]. While these models provide insight into hydrological processes, they require extensive data input and calibration, making them difficult to apply in regions with limited monitoring networks [12]. Furthermore, parameter uncertainties can significantly reduce model reliability during extreme flooding events [13].

D. Machine Learning Approaches in Flood Forecasting

With advancements in computational power and data availability, machine learning has become an attractive approach for modeling hydrological processes. Artificial Neural Networks (ANNs) are capable of learning complex nonlinear relationships from data and have shown promising performance in flood forecasting applications. However, standard feed-forward ANNs are limited in handling long-term temporal dependencies in time-series data [14]. Support Vector Regression (SVR) offers good generalization performance but requires careful tuning of kernel functions and regularization parameters [15]. Random Forest (RF) performs well in capturing nonlinear patterns but may become computationally intensive with large datasets [16]. Both methods demonstrate improvement over linear models, yet are still constrained in modeling sequential dependencies over time.

E. Deep Learning Models in Hydrological Time-Series Forecasting

Recurrent Neural Networks introduced time-sequence modeling capability. However, traditional RNNs suffer from gradient vanishing issues, limiting their ability to learn long-term dependencies [16]. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models were specifically developed to address memory retention issues and have demonstrated superior performance in predicting streamflow and river water levels [17]. These models capture temporal dependencies more effectively than simple ANNs [18]. Hybrid models that integrate convolutional and recurrent architectures or attention mechanisms have further improved forecasting accuracy by extracting temporal and spatial patterns [19]. However, these models require large datasets and high computational resources, limiting their real-time deployment in some regions [20].

F. NARX Model for Hydrological and Flood Forecasting

The Nonlinear Autoregressive with Exogenous Input (NARX) model is a dynamic neural network that uses previous output values and external variables (e.g., rainfall) to predict future water levels [21]. This architecture enables the model to represent nonlinear hydrological dynamics effectively. NARX models converge faster and provide higher prediction accuracy than traditional ANN and RNN due to their explicit memory mechanism [22]. They are particularly effective in modeling systems where external forcing influences the target variable. Studies demonstrate that NARX achieves high accuracy in river level forecasting across Southeast Asian river basins [23]. Its performance is notably superior when dealing with short-term and multi-step forecasting [23].

G. Research Gap

Although a variety of hydrological and data-driven flood forecasting models have been applied in Malaysia, most previous research has focused on river basins in Kelantan and Terengganu [24], while significantly less attention has been given to Pahang, particularly the Pekan district. Existing flood prediction frameworks do not adequately represent the hydrological and rainfall–runoff characteristics of Pahang's river systems. As a result, there is currently no dedicated flood prediction model specifically developed for Pekan, despite its recurring seasonal floods and high socio-economic vulnerability. Furthermore, previous studies rarely examine flood behavior along the entire river course, from upstream to downstream, which is critical for understanding how rainfall and runoff propagate through the watershed before causing overflow in low-lying coastal areas like Pekan. Developing a tailored model would therefore not only improve prediction accuracy but also enable a better understanding of flood dynamics along the river continuum, supporting more effective local early-warning and disaster preparedness strategies. Hence, this study focuses on designing and evaluating an NARX-based prediction model specifically for the Pekan region, to address this gap.

H. Summary of the Literature Review

While conventional statistical and physically-based models provide foundational hydrological understanding, their performance declines due to the nonlinear and dynamic characteristics of flood behavior. Machine learning and deep learning methods such as ANN, SVR, RF, LSTM, and hybrid architectures have shown improved predictive ability by capturing complex temporal dependencies. The NARX model, in particular, has demonstrated strong performance in multi-step hydrological forecasting due to its capability to incorporate external variables and maintain memory of past system states.

However, the review shows that little research has focused on flood prediction in Pahang, and almost none has been conducted specifically for Pekan, even though the district is highly affected by recurring monsoon floods [1]. Additionally, there is limited study on flood progression along the river system from upstream to downstream, which is essential for understanding how floodwater develops and travels before reaching populated flood-prone areas. Therefore, the present study aims to develop a NARX-based river level prediction model specifically tailored for the Pekan river system, providing more localized and accurate early-warning insights to support flood management and preparedness.

III. METHODOLOGY

A. Dataset

The dataset consists of hourly hydrological measurements collected over five years (2020–2024) from four river stations: Sungai Pahang, Sungai Pahang Tua, Sungai Paloh Hinai, and Sungai Mentiga. The target station for prediction is Sungai Pahang at Pekan, which represents the confluence point of the other three rivers.

Each dataset contains several attributes, including Station ID, Station Code, Station Name, Date, Time, Water Level (m), Rainfall (mm), Cumulative Daily Rainfall (mm), and Cumulative Annual Rainfall (mm). For this study, only the Water Level (m) parameter was retained to ensure model focus and reduce noise from unrelated variables.

B. Data Preprocessing

To standardize the temporal resolution, the hourly data were aggregated into daily averages, resulting in one representative water level value per day for each station. The daily average water level, $\bar{H}_d$, was computed as:

$$\bar{H}_d = \frac{1}{n}\sum_{i=1}^{n} H_i \tag{1}$$

where, $H_i$represents the water level recorded at the i^{th}hour of a given day, and $n$is the total number of hourly observations within that day (typically $n = 24$).

C. Interpolation of Short Gaps

Short gaps, defined as missing periods of fewer than 5 consecutive days, were imputed using linear interpolation. For a missing value at time t, the estimate was calculated as:

$$x_t = \frac{(t-t_{prev})*x_{next}+(t_{next}-t)*x_{prev}}{t_{next}-t_{prev}} \tag{2}$$

where, x_{prev} and x_{next} represent the nearest observed values surrounding the gap. This approach is appropriate for short-duration gaps where hydrological changes are typically gradual and continuous.

Large gaps (≥5 consecutive days) were not imputed. Instead, the entire row corresponding to that timestamp was removed across all stations. This ensures that each training instance contains complete values for Sungai Pahang, Sungai Pahang Tua, Sungai Paloh Hinai, and Sungai Mentiga, thereby avoiding inconsistencies caused by high-uncertainty imputations.

The removal of rows containing long gaps is motivated by several scientific and statistical considerations:

- Non-linear and event-driven hydrological behavior: Water levels in rivers respond to rainfall, upstream releases, tidal influences, and runoff processes that evolve non-linearly. Over multi-day periods, such dynamics cannot be reliably reconstructed using linear or low-order imputation techniques. Long-gap interpolation introduces unrealistic smoothing and distorts natural fluctuations.
- Autocorrelation decay in river-level time series: Hydrological time series typically exhibit rapid

autocorrelation decay within a few days. When observations are separated by long gaps, their statistical dependence becomes weak, making interpolated values unreliable. Avoiding long synthetic intervals preserves the natural temporal structure of the data.

- Uncertainty amplification: Interpolation error grows with gap duration. Long-gap estimates carry high uncertainty that may propagate into model training, degrading predictive performance and weakening the model's ability to capture extreme events.
- Avoidance of model bias and artefact learning: Machine learning models are sensitive to artificial patterns caused by imputation. Long imputed segments may lead the model to learn interpolation artefacts rather than genuine hydrological relationships among the four stations.

After preprocessing, the dataset was divided into continuous segments—defined as sequences of dates in which none of the stations (Sungai Pahang, Sungai Pahang Tua, Sungai Paloh Hinai, Sungai Mentiga) exhibits a large missing gap. Separate regression models were trained on these fully complete segments. Training regression models on continuous, fully observed segments is scientifically valid and methodologically justified for several reasons:

- Preservation of underlying hydrological dynamics: Continuous segments represent uninterrupted periods during which the inter-station relationships reflect real hydrological behavior. This allows the model to learn authentic temporal patterns rather than artefacts arising from imputed values.
- Assumption of local stationarity: River systems often exhibit local stationarity, meaning their statistical properties remain stable within reasonably short continuous periods. Segment-wise modelling respects this property and allows consistent parameter estimation.
- Avoidance of cross-regime contamination: Long missing gaps often coincide with unusual hydrological states (e.g., floods, sensor outages, maintenance). Segmentation prevents mixing incompatible regimes that could introduce non-physical relationships into the model.
- Reduction of structural bias: Restricting training to fully observed data ensures that regression parameters reflect true input–output relationships rather than smoothed or synthetic signals.
- Hypothetical validity via conditional independence: If each continuous segment is assumed to originate from the same underlying hydrological process governing the four stations, then each segment provides an independent sample from the same distribution. This satisfies the theoretical assumptions required for regression modelling.
- Improved generalization: Models trained only on high-quality, continuous data are better able to generalize to unseen periods, particularly during rapidly changing hydrological conditions such as rising or receding river stages.

These grouped datasets were then utilized for the Nonlinear AutoRegressive model with eXogenous inputs (NARX) during the modeling stage.

The NARX model was implemented using Python with TensorFlow and Keras frameworks. The dataset was divided into training and testing sets, with the training phase used to optimize network weights.

The loss function used to evaluate model performance is the Mean Squared Error (MSE), calculated as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

where, y_i and $\hat{y}_i$ are the observed and predicted water levels, respectively, and N is the total number of samples. Lower MSE values indicate higher prediction accuracy and better model generalization.

D. NARX Model Structure

The general form of the NARX model is expressed as:

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-n_y), x(t-1), x(t-2), \dots, x(t-n_x)) \tag{4}$$

where,

- $y(t)$ is the predicted river water level at time t,
- $x(t)$ represents the exogenous input variable (e.g., rainfall or other influencing factors),
- n_y and n_x denote the feedback (lag) orders, and
- $f(\cdot)$ is a nonlinear function approximated using an MLP network.

E. Model Architecture

Table I shows that the predictive model used a feedforward neural network with one hidden layer, following the structure:

TABLE I. THE STRUCTURE OF THE MODEL

Layer	Description
Input Layer	Dimension determined by the lag size and number of variables.
Hidden Layer	Fully connected, with variable neuron counts (150, 300, or 600 units).
Activation Function	Rectified Linear Unit (ReLU).
Dropout Layer	20% dropout for regularization and overfitting prevention.
Output Layer	A single linear neuron producing the next-step forecast.

The model was defined using the Sequential API in TensorFlow Keras. The network was compiled with the Adam optimizer (learning rate = 1×10^{-3}, weight decay = 1×10^{-5}) and Mean Squared Error (MSE) as the loss function.

F. Training Procedure

Each configuration was trained under varying combinations of:

- Lags: {3,4,5,6,7}
- Hidden Units: {150,300,600}
- Epochs: {100,300,500,1000,5000}

This yielded a total of 45 experimental runs.

Training was executed with a batch size of 128 and data shuffling enabled to prevent temporal bias.

During each experiment:

- The model was trained on the training set and validated on the test set.
- The training and validation loss histories were recorded.
- Predictions were generated for both training and test datasets.
- Diagnostic visualizations were saved, including:
  - Loss curve (training vs. validation loss),
  - Residual plots (train/test residuals),
  - Prediction plots (actual vs. predicted values).

G. Evaluation Metrics

The model's predictive performance was primarily evaluated using Mean Squared Error (MSE) as the loss function. Additional diagnostic analyses included:

- Residual distribution to assess error randomness and bias.
- Train vs. Validation loss comparison to detect overfitting.
- Actual vs. Predicted plots to visually assess alignment between predictions and ground truth.

IV. Implementation, Results and Discussion

A series of experiments were conducted to evaluate the performance of a neural network model trained with different lag values and hidden unit configurations for time-series forecasting. The parameter grid covered lags {3, 4, 5, 6, 7}, hidden units {150, 300, 600}, and epochs {100, 300, 500, 1000, 5000}. The primary evaluation metric was validation loss (Mean Squared Error), with train loss also monitored to assess potential overfitting or underfitting.

A. Best Validation Performance by Lag and Hidden Units

The heatmap in Fig. 1 illustrates the best validation loss achieved for each combination of lag and hidden unit count.

Fig. 1. Heatmap.

Fig. 1 shows that the model achieved its lowest validation loss (0.102) for Lag = 6 and Hidden Units = 300, indicating that this configuration provides an optimal balance between temporal dependency capture and model complexity. Generally, lags between 3 and 6 produced comparable results, while performance deteriorated at lag = 4 and lag = 5, especially when the hidden layer size was small (150). Increasing the hidden units to 300 consistently improved performance for most lags, whereas using 600 units did not always yield further benefits — suggesting possible overparameterization for the dataset size.

B. Relationship Between Training and Validation Loss

The scatter plot in Fig. 2 compares the final train loss against validation loss across all configurations.

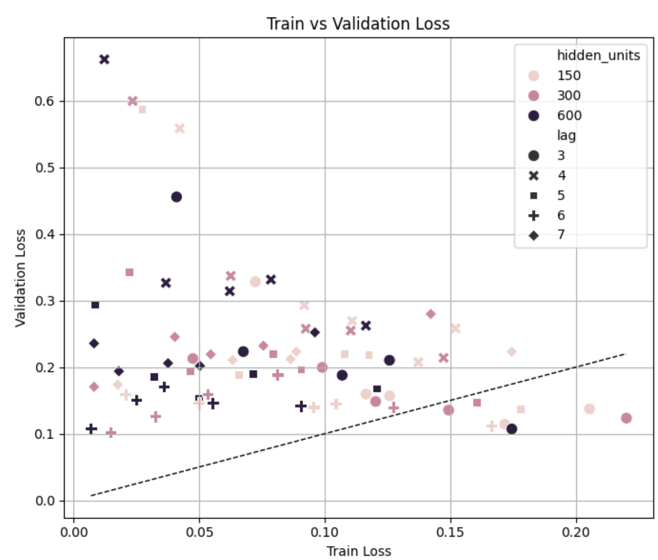

Fig. 2. Scatter plot.

As shown in Fig. 2, most points lie close to or above the diagonal reference line, indicating that validation losses were typically higher than training losses — a standard indicator of slight overfitting. However, several configurations (particularly those with hidden units = 300 and moderate lags) exhibited a close alignment between train and validation losses, suggesting good generalization. Notably, models trained with Lag = 4 and Lag = 5 showed larger gaps between train and validation losses, reflecting unstable learning dynamics or insufficient temporal context.

C. Validation Loss Trends Across Epochs

Fig. 3 to Fig 7 show validation loss progression across training epochs for each lag value.

Lag = 3: Validation loss decreased sharply up to 1000 epochs for hidden units = 150 and 300, followed by an increase at 5000 epochs, indicating early convergence and later overfitting.

Lag = 4: All configurations exhibited diverging losses beyond 1000 epochs, implying unstable training or excessive learning rates for deeper models.

Lag = 5: The validation loss initially improved but then rose steeply for higher epochs, confirming overfitting at extended training durations.

Lag = 6: The trend showed consistent improvement as training progressed, achieving the lowest validation loss (≈0.10) at 5000 epochs, particularly with hidden units = 300. This indicates a robust generalization pattern.
- Lag = 7: The loss curves demonstrated oscillation and mild overfitting beyond 2000 epochs, though the overall performance remained stable within a 0.18–0.25 MSE range.

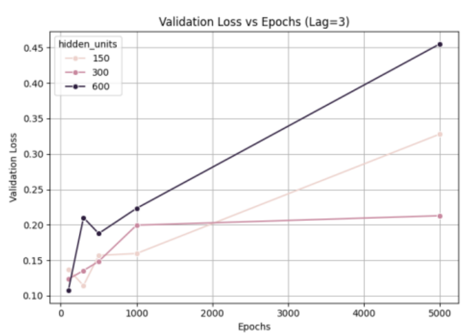

Fig. 3. Lag 3 validation loss.

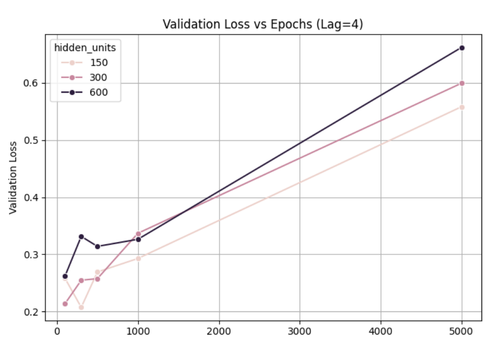

Fig. 4. Lag 4 validation loss.

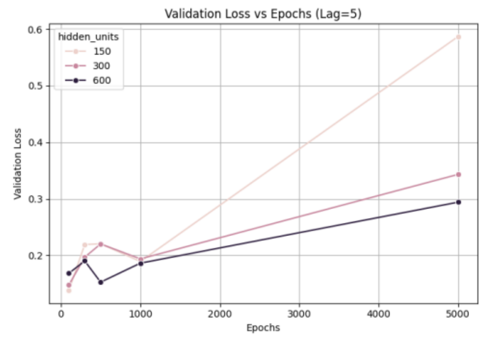

Fig. 5. Lag 5 validation loss.

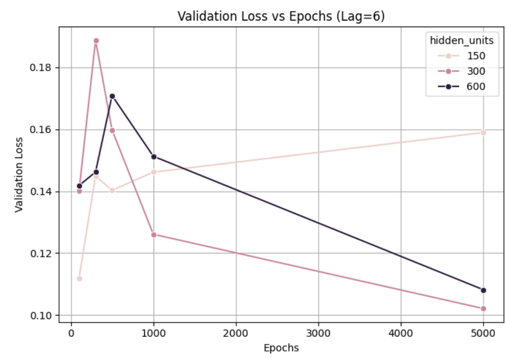

Fig. 6. Lag 6 validation loss.

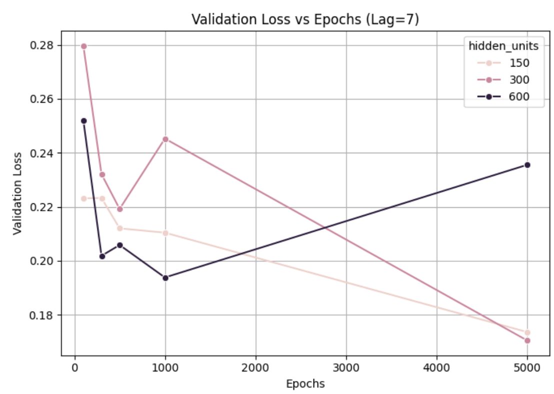

Fig. 7. Lag 7 validation loss.

D. Cross-Lag Performance Interpretation

Comparing across lags, the results suggest that:

- Short lags (3–4) capture limited temporal dependencies, causing early saturation in validation performance.
- Intermediate lag (6) offers an optimal trade-off — long enough to model dependencies, but not too large to introduce noise.
- Longer lag (7) yields diminishing returns, potentially due to redundant temporal information or vanishing gradients.

This aligns with the hypothesis that increasing input sequence length does not guarantee improved forecasting accuracy beyond a certain point.

E. Summary of Key Observations

From Table II, the observed behavior underscores the importance of balancing model capacity, temporal depth, and training duration. While larger models (600 units) have higher representational power, they are also prone to overfitting when training data are limited. Conversely, too few hidden units (150) restrict learning capability. Hence, medium-sized architectures (300 units) paired with sufficient historical context (Lag = 6) represent an optimal trade-off for this dataset. These findings can guide future hyperparameter selection for similar neural forecasting tasks, where data periodicity and temporal correlation length are critical factors.

TABLE II. THE BEST RESULT FOR EACH LAG

Lag	Hidden Units	Best Val Loss	Notes
3	300	0.114	Stable and generalizable
4	300	0.214	Poor stability beyond 1000 epochs
5	600	0.153	Unstable, possibly overfitted
6	300	0.102 (best)	Strong generalization and stable learning
7	150	0.174	Gradual improvement but limited gain

F. Detailed Evaluation of the Best Model (Lag = 6, Hidden Units = 300)

1) Loss curve analysis: The loss curve in Fig. 8 depicts the evolution of training and validation losses across 5000 epochs. The training loss exhibits a steep decline during the early epochs, stabilizing near zero after approximately 1000 epochs.

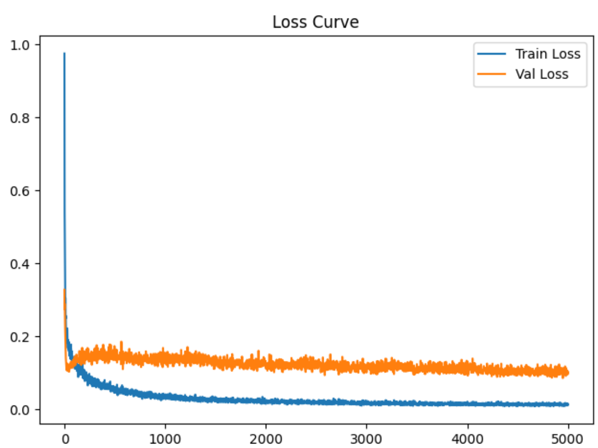

Fig. 8. Training and validation loss curves showing convergence behavior for the best model configuration (Lag = 6, Hidden Units = 300).

The validation loss follows a similar trend but converges to a higher steady-state value around 0.10, consistent with the best validation loss recorded in the summary. The relatively small gap between training and validation losses indicates effective generalization and minimal overfitting. Both curves maintain stability without significant oscillation or divergence, suggesting that the model converged smoothly. This steady behavior reflects robust learning dynamics, where the model captures relevant temporal dependencies without memorizing noise from the training set.

2) Residual diagnostics: The residual plots in Fig. 9 illustrate the difference between predicted and actual values for both training and test sets.

Fig. 9. Residual plots for training and test datasets, showing randomness and unbiased prediction errors.

In the training residuals, values fluctuate randomly around zero with no visible pattern or trend, implying that the model's predictions are unbiased and errors are normally distributed. The magnitude of residuals remains small (within ±0.1), confirming that the model fits the training data accurately without systematic error. The test residuals display slightly larger fluctuations, ranging within ±0.25, which is expected due to unseen data. However, no clear autocorrelation or drift is present, supporting the conclusion that the model generalizes well beyond the training sample. This diagnostic outcome supports the statistical soundness of the model and its ability to perform consistent predictions across both datasets.

3) Prediction performance on training data: The training prediction plot in Fig. 10 demonstrates an almost perfect overlap between actual and predicted values. The consistency across the full sequence confirms the model's high accuracy in reconstructing training samples while maintaining generalization, as verified in the test results. This balance reinforces that the selected architecture (Lag = 6, Hidden Units = 300) achieves optimal expressiveness without overfitting. The model appears to have successfully extracted dominant temporal patterns and relationships within the data.

4) Prediction performance on test data: Fig. 11 compares actual and predicted values for the test set. The predicted curve closely tracks the actual target trend, capturing both peaks and troughs of the series effectively. Minor deviations occur at extreme points, which may be attributed to local variations not captured by the lag window of six.

The close alignment between both curves indicates that the model successfully learns the non-linear temporal dependencies within the data. The prediction continuity and minimal phase lag between actual and predicted signals further validate the model's capacity to anticipate short-term dynamics with good precision.

5) Summary of findings for optimal model: The results highlight the importance of appropriate lag selection in temporal modeling. With a lag of 6, the network accesses sufficient historical context to infer meaningful sequential dependencies without introducing redundant or noisy information. Moreover, the moderate hidden layer size of 300 provides adequate representational capacity to capture non-linear relations while preserving regularization balance. The residual randomness and low validation loss collectively confirm that the model achieved an optimal bias–variance tradeoff. This configuration will be considered as the final candidate for deployment or further comparative experiments with other forecasting or hybrid deep learning approaches (see Table III).

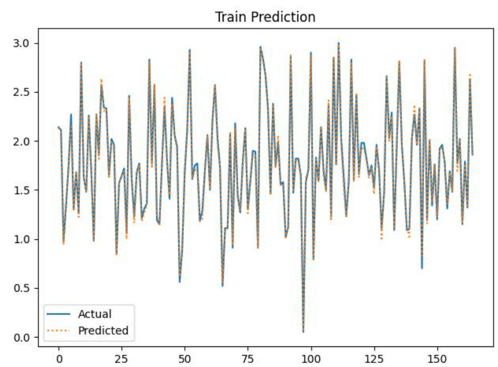

Fig. 10. Training prediction performance showing close alignment between predicted and actual series.

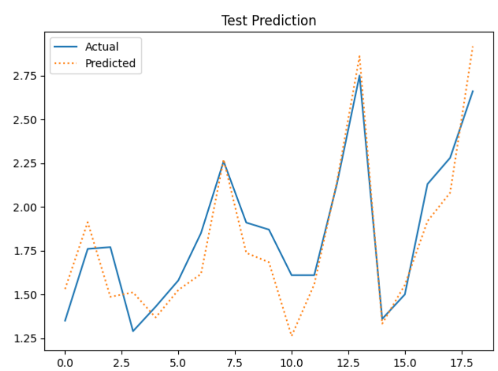

Fig. 11. Test prediction performance comparison between actual and predicted values.

TABLE III. FINDING SUMMARY

Metric	Observation
Lag	6
Hidden Units	300
Best Validation Loss	0.102
Training Stability	Excellent (Smooth convergence)
Generalization	Strong (Low validation gap)
Residual Behavior	Random, low magnitude
Test Prediction Fit	High correlation with actual series

V. CONCLUSION

This study set out to develop a reliable and accurate river water level forecasting model for the Pekan region using the Nonlinear Autoregressive model with Exogenous Inputs (NARX). The motivation derived from recurring monsoon-driven floods in the East Coast of Peninsular Malaysia, where existing forecasting systems lack models tailored specifically to Pahang's hydrological characteristics. The literature review highlighted that while conventional statistical and physically-based models offer foundational insights, their performance often deteriorates under nonlinear and dynamic flood conditions. Modern machine learning and deep learning techniques provide improved predictive capabilities, yet most previous research remains concentrated in Kelantan and Terengganu, leaving a significant gap for Pahang—particularly the low-lying Pekan district.

A comprehensive dataset of daily water levels from 2020 to 2024 was preprocessed, cleaned, and segmented into continuous temporal sequences to ensure robust model training. Using this dataset, a series of 75 experiments was conducted across varying lag values, hidden unit sizes, and training epochs to identify the optimal NARX configuration. Performance evaluation relied primarily on Mean Squared Error (MSE), supported by residual diagnostics and actual-versus-predicted comparisons for both training and test sets.

The experimental results demonstrated that Lag = 6 and Hidden Units = 300 provided the most effective and stable predictive performance, achieving a minimum validation loss of 0.102. This configuration offered a balanced combination of temporal memory and model complexity, outperforming smaller architectures, which lacked representational strength, as well as larger models, which tended to overfit. Loss curve analyses further confirmed smooth convergence and minimal divergence between training and validation losses, indicating strong generalization. Residual analyses showed randomness without bias, while prediction plots revealed close alignment with actual water level fluctuations, successfully capturing trend variations and peak events.

From a scientific perspective, this study contributes a structured and reproducible NARX modelling framework that integrates multi-station river dynamics with systematic lag–capacity optimization for district-level hydrological forecasting. Practically, the proposed model offers actionable value for local flood early warning systems by providing accurate short-term water level predictions that can support decision-making by authorities in flood-prone areas such as Pekan.

Despite these encouraging results, several limitations of this study should be acknowledged. First, the model relies solely on historical water level data and does not explicitly incorporate rainfall, tidal influence, or reservoir release information, which may further enhance predictive accuracy during extreme hydrological events. Second, model training was conducted using daily-aggregated data, which may limit responsiveness to rapid intra-day water level changes during flash flood conditions.

Future work will focus on extending the proposed framework by incorporating additional hydrometeorological variables and evaluating multi-step-ahead forecasting performance for longer early warning lead times. Integration with real-time sensor networks and comparative evaluation against alternative data-driven models may further strengthen the operational applicability of the approach.

In conclusion, the NARX-based approach developed in this work offers a valuable contribution toward strengthening flood preparedness and decision-support mechanisms in Pahang. With continued refinement and integration into operational forecasting systems, it can play a significant role in mitigating the socio-economic impacts of seasonal flooding and enhancing resilience in vulnerable communities.

ACKNOWLEDGMENT

This study was made possible by the generous funding provided by the international grant UIC241504. The authors are deeply indebted to Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) and Al Maarif University College for their invaluable contributions in terms of facilities, expertise, and collaboration. The authors also wish to express their sincere appreciation to the Department of Irrigation and Drainage, Pahang State (JPS Pahang) for their support, cooperation, and provision of relevant information throughout the course of this study.

REFERENCES

[1] A. N. Abd Rahman, F. Othman, W. J. Wan Jaafar, and A. H. K. Ahmed Elshafie, "An assessment of floods' characteristics and patterns in Pahang, Malaysia," Larhyss J., vol. No. 55, no. Sept 2023, pp. 89–105.

[2] I. Usman Kaoje et al., "Physical Flood Vulnerability Assessment using Geospatial Indicator-Based Approach and Participatory Analytical Hierarchy Process: A Case Study in Kota Bharu, Malaysia," Water, vol. 13, no. 13, p. 1786, Jan. 2021, doi: 10.3390/w13131786.

[3] A. Albani, M. Z. Ibrahim, and K. H. Yong, "Influence of the ENSO and Monsoonal Season on Long-Term Wind Energy Potential in Malaysia," Energies, vol. 11, no. 11, p. 2965, Nov. 2018, doi: 10.3390/en11112965.

[4] "(PDF) Review of the North-East Monsoon 2020/2021 in Malaysia," ResearchGate. Accessed: Nov. 13, 2025. [Online]. Available: https://www.researchgate.net/publication/387669899_Review_of_the_North-East_Monsoon_20202021_in_Malaysia

[5] K. A. Baharuddin et al., "The Record-Setting Flood of 2014 in Kelantan: Challenges and Recommendations from an Emergency Medicine Perspective and Why the Medical Campus Stood Dry," Malays. J. Med. Sci. MJMS, vol. 22, pp. 1–7, May 2015.

[6] "2022/26 'Malaysia's Floods of December 2021: Can Future Disasters be Avoided?' by Serina Rahman." Accessed: Nov. 13, 2025. [Online]. Available: https://www.iseas.edu.sg/articles-commentaries/iseas-perspective/2022-26-malaysias-floods-of-december-2021-can-future-disasters-be-avoided-by-serina-rahman/

[7] "Malaysia Floods 2022, DREF Application (MDRMY009) - Malaysia | ReliefWeb." Accessed: Nov. 13, 2025. [Online]. Available: https://reliefweb.int/report/malaysia/malaysia-floods-2022-dref-application-mdrmy009

[8] N. Alias, H. mohamad, W. Y. Chin, and Z. Yusop, "Rainfall analysis of the Kelantan big yellow flood 2014," J. Teknol., vol. 78, pp. 83–90, May 2016, doi: 10.11113/jt.v78.9701.

[9] "Communication, Education and Public Awareness Program (CEPA) in Flood Forecasting and Warning Program (PRAB) | Journal of Water Resources Management." Accessed: Nov. 13, 2025. [Online]. Available: https://journal.water.gov.my/index.php/jowrm/article/view/6

[10] B. Ghimire, "Application of ARIMA Model for River Discharges Analysis," J. Nepal Phys. Soc., vol. 4, p. 27, May 2017, doi: 10.3126/jnphyssoc.v4i1.17333.

[11] [S. Gegenleithner, M. Pirker, C. Dorfmann, R. Kern, and J. Schneider, "Long short-term memory networks for enhancing real-time flood forecasts: a case study for an underperforming hydrologic model," Hydrol. Earth Syst. Sci., vol. 29, pp. 1939–1962, Apr. 2025, doi: 10.5194/hess-29-1939-2025.

[12] M. Aliye, A. Aga, T. Danbara, and P. Aga, "Evaluating the Performance of HEC-HMS and SWAT Hydrological Models in Simulating the Rainfall-Runoff Process for Data Scarce Region of Ethiopian Rift Valley Lake Basin," Open J. Mod. Hydrol., vol. 10, pp. 105–122, Oct. 2020, doi: 10.4236/ojmh.2020.104007.

[13] N. W. Chaney, J. D. Herman, P. M. Reed, and E. F. Wood, "Flood and drought hydrologic monitoring: the role of model parameter uncertainty," Hydrol. Earth Syst. Sci., vol. 19, no. 7, pp. 3239–3251, July 2015, doi: 10.5194/hess-19-3239-2015.

[14] "(PDF) Efficacy of Feedforward and LSTM Neural Networks at Predicting and Gap Filling Coastal Ocean Timeseries: Oxygen, Nutrients, and Temperature." Accessed: Nov. 13, 2025. [Online]. Available: https://www.researchgate.net/publication/351284531_Efficacy_of_Feedforward_and_LSTM_Neural_Networks_at_Predicting_and_Gap_Filling_Coastal_Ocean_Timeseries_Oxygen_Nutrients_and_Temperature

[15] "A Feature-Weighted SVR Method Based on Kernel Space Feature." Accessed: Nov. 13, 2025. [Online]. Available: https://www.mdpi.com/1999-4893/11/5/62

[16] A. Chernodub and D. Nowicki, "Sampling-based Gradient Regularization for Capturing Long-Term Dependencies in Recurrent Neural Networks," arXiv.org. Accessed: Nov. 14, 2025. [Online]. Available: https://arxiv.org/abs/1606.07767v3

[17] L. A. De la Fuente, M. R. Ehsani, H. V. Gupta, and L. E. Condon, "Toward interpretable LSTM-based modeling of hydrological systems," Hydrol. Earth Syst. Sci., vol. 28, no. 4, pp. 945–971, Feb. 2024, doi: 10.5194/hess-28-945-2024.

[18] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Dec. 2018, pp. 1394–1401. doi: 10.1109/ICMLA.2018.00227.

[19] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks," Hydrol. Earth Syst. Sci., vol. 22, no. 11, pp. 6005–6022, Nov. 2018, doi: 10.5194/hess-22-6005-2018.

[20] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing, "Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets," Hydrol. Earth Syst. Sci., vol. 23, no. 12, pp. 5089–5110, Dec. 2019, doi: 10.5194/hess-23-5089-2019.

[21] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," IEEE Trans. Neural Netw., vol. 5, no. 6, pp. 989–993, Nov. 1994, doi: 10.1109/72.329697.

[22] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175, Jan. 2003, doi: 10.1016/S0925-2312(01)00702-0.

[23] G. Hayder, M. Iwan Solihin, and M. R. N. Najwa, "Multi-step-ahead prediction of river flow using NARX neural networks and deep learning LSTM," H2Open J., vol. 5, no. 1, pp. 43–60, Jan. 2022, doi: 10.2166/h2oj.2022.134.

[24] S. L.m, B. H, Y. M.r, bin H. M.h, binti M. D. S.a, and S. T.a./L, "Hydrological Analysis for Flood Forecasting at Sg Golok River Basin Malaysia," 2022, doi: 10.1007/978-981-16-7397-9_27.