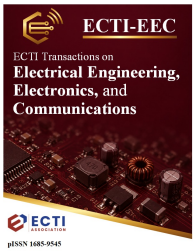
ECTI-EEC Transactions Pre-proof



Comparative Analysis of Machine Learning Models for Water Level Forecasting towards Flood Early Warning in the Mekong River

Tanapattara Wongkhamchan, Ariya Namvong, Jakub Michel, Chayada Surawanitkun, Wullapa Wongsinlatam, Tawun Remsungnen, Thalerngsak Wiangwiset, Wichian Tong-Gate, and Apirat Siritaratiwat

DOI: http://doi.org/10.37936/ecti-eec.2027251.262931

Reference: ECTI-EEC 262931

Received date: 22 December 2025

Accepted date: 15 October 2026

This is a PDF of an article that has been improved after acceptance, including the inclusion of a cover page and metadata, as well as the optimization of readability. Before its publication in its final form, this version will undergo further copyediting, typesetting, and review. As a result, this version is no longer the Accepted Manuscript, but it is not yet the definitive **Version of Record**. We are providing this early version to provide the article with early visibility.

Published by ECTI Association Thailand

ECTI Transactions on Electrical Engineering, Electronics, and Communications, Vol. 25, No. 1, pp. 22–31, February 2027

https://doi.org/10.37936/ecti-eec.2027251.262931
ISSN(Online): 1685-9545

Comparative Analysis of Machine Learning Models for Water Level Forecasting towards Flood Early Warning in the Mekong River

Tanapattara Wongkhamchan[1], Chayada Surawanitkun[2], Wullapa Wongsinlatam[3], Tawun Remsungnen[4], Jakub Michel[5], Thalerngsak Wiangwiset[6], Wichian Tong-Gate[7], Apirat Siritaratiwat[8] and Ariya Namvong[9†]

Abstract

Flooding is a recurring hazard in Thailand, particularly in Nong Khai, with substantial socioeconomic consequences. Traditional hydrological models can be computationally demanding and may not fully capture nonlinear dependencies among spatially distributed water-level stations. This study compares four data-driven models—Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Random Forest (RF), and Support Vector Machine implemented as Support Vector Regression (SVM/SVR)—for one-day-ahead water-level forecasting at Nong Khai. The models were trained and evaluated using 19 years (2006–2025) of daily observations from five monitoring stations along the Mekong River, with predictive performance assessed using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). All models reproduced the principal seasonal pattern; however, the recurrent models produced lower forecasting errors than the classical baselines. LSTM recorded the lowest forecasting error (RMSE = 0.3459; MAE = 0.1196), followed by GRU (RMSE = 0.4059; MAE = 0.1647), whereas RF and SVM/SVR showed substantially higher error values. The key contribution of this study is a consistent multi-station comparison suggesting that sequence-based learning may more effectively capture upstream-to-downstream hydrological dependencies in the Mekong River. These findings suggest LSTM as a promising forecasting component for a future hydrological Digital Twin- defined here as a continuously updated virtual representation of the river system that integrates monitoring data with predictive models- to support flood early warning and operational decision-making.

Keywords: Flood prediction, Mekong River, Water level forecasting, Long Short-Term Memory.

I. Introduction

Flooding is among the most frequent and economically damaging natural disaster in Thailand, causing significant disruption to the nation's socioeconomic infrastructure and agricultural sectors [1]. Historical data indicate that between 2009 and 2018, flooding impacted over 55 million people and generated economic losses exceeding 50 billion THB [2]. The severity of such events was illustrated by the 2011 "Mega Flood," which resulted in over 800 fatalities and economic losses estimated at USD 46.5 billion, underscoring the need for effective flood management systems [3].

Central Thailand often receives significant attention in disaster management. By contrast, the northeastern region—particularly provinces situated along the Mekong River—faces distinct and increasingly complex hydrological challenges driven by both climatic variability and upstream river regulation. The hydrological dynamics in this region are further influenced by irregular water level fluctuations in the Mekong River, which are often associated with water discharge from upstream hydroelectric dams in China and Lao PDR. These operations can trigger out-of-season floods or rapid rises in water level with little to no advance notice, leading to what are described as "floods in the dry season" that can significantly disrupt the livelihoods of riparian communities.

The severe flooding of September 2024, associated with Typhoon Yagi, exceeded the capacity of local defences and

Manuscript received December 22, 2025; revised on October 12, 2026; accepted October 15, 2026.
Recommended for publication by Associate Editor Chanon Warisarn.
†Corresponding Author: Ariya Namvong(ariyna@kku.ac.th)
[1,2,3,4,9]Center of Multidisciplinary Innovation Network Talent (MINT Center), Faculty of Interdisciplinary Studies, Department of Technology and Engineering, Khon Kaen University, Nong Khai Campus, Nong Khai 43000, Thailand.
[5]Department of Mathematical Sciences, High Point University, North Carolina, USA.
[6]Central Maintenance Sector, National Telecom Public Company Limited, Khon Kaen 40000, Thailand.
[7]Hydrology Office 1, Water Resources Office 3, Department of Water Resources, Nong Khai 43000, Thailand.
[8]Department of IoT and Information Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.

inundated Nong Khai's central economic district, causing significant damage to critical infrastructure and disrupting primary business zones [4]. This event also highlighted limitations in current disaster response mechanisms, particularly regarding the timeliness of warnings and the availability of systematic evacuation mapping.

Effective mitigation of such disasters depends on the ability to accurately predict water levels and identify high-risk zones with sufficient lead time. Traditional prediction methods, such as physical-based hydrological models, are often computationally intensive and require extensive data inputs, limiting their utility for real-time forecasting in complex river systems [5]. Data-driven approaches, specifically Machine Learning (ML) and Deep Learning (DL) architectures, have shown promising capabilities in modelling the nonlinear and complex relationships inherent in hydrological time-series data [6]. The literature suggests that recurrent neural networks (RNNs), particularly LSTM and GRU, are well-suited for capturing temporal dependencies, while RF and SVM provide useful benchmarks for regression tasks.

To enhance the robustness of the forecasting framework, this study incorporates multivariate time-series data from spatially distributed upstream monitoring stations. This approach is grounded in the hydrological principle that upstream discharge dynamics and flow rates are associated with downstream water-level changes, with characteristic lag times. Building on this, the study examines whether recurrent neural network (RNN) architectures, particularly the LSTM model, are well suited to modelling these complex spatiotemporal dependencies.

The LSTM's memory-gating mechanisms enable effective retention and processing of long-term sequential patterns derived from upstream data. This is expected to improve predictive accuracy at the downstream target station.

To address the need for improved early warning in Nong Khai, this study develops and compares the predictive performance of four machine learning models. Using a nineteen-year dataset (2006–2025) from five monitoring stations, it aims to identify a suitable forecasting architecture to support flood management in the Mekong River Basin, in a region that already leverages real-time infrastructure for public safety [7].

In this study, a hydrological Digital Twin refers to a continuously updated virtual representation of the river system that integrates real-time observations with predictive models to estimate current conditions, simulate possible future states, and support operational flood-management decisions.

II. LITERATURE REVIEW

Flood prediction methodologies are broadly categorized into physics-based hydrological models (such as HEC-HMS [8]) and data-driven machine learning techniques. While physics-based models simulate physical processes, data-driven approaches forecast water levels directly from historical records [9]. Recurrent neural networks (RNNs), specifically LSTM and GRU models, have received considerable attention for their ability to handle time-series data. These architectures have shown strong performance in capturing complex, long-term temporal dependencies in hydrological data and are frequently compared to identifying the most suitable predictive model for specific applications.

A. Deep Learning Architectures: Recurrent Neural Networks (RNNs)

Several studies have applied and evaluated LSTM and GRU models in flood forecasting contexts. Sharma *et al.* [10] and Hieu *et al.* [11] integrated these deep learning models with hydrological data, demonstrating improvements in the identification of inundation areas and prediction of river levels compared with computationally intensive physical models. Cho *et al.* [12] and Atashi *et al.* [13] compared various models, finding that LSTM and GRU generally produced lower RMSE values in predicting historical maximum water levels, with LSTM tending to outperform classical models such as RF.

The application of LSTM has been widely explored across diverse hydrological contexts. Zhang *et al.* [14] employed LSTM to reconstruct hydrological processes in the Mekong River, examining the regulatory effects of hydropower development. Chai *et al.* [15] improved daily runoff prediction accuracy by combining LSTM with a Firefly Algorithm for hyperparameter optimization, while Hop *et al.* [16] employed LSTM as a rapid surrogate for 1D/2D hydraulic models. However, limitations remain; Hasyimi *et al.* [17] noted that the accuracy of LSTM models can be adversely affected by a lack of variable diversity in the training dataset.

GRU has also received attention as a computationally efficient alternative, particularly in multivariate learning contexts. Fordjour *et al.* [18] and Gharehbaghi *et al.* [19] reported that GRU models can improve prediction accuracy when trained on multivariate data from spatially distributed observatories. Zhang *et al.* [20] and Manokij *et al.* [21] addressed lag-time preprocessing and data imbalance in precipitation forecasting, respectively, contributing to the development of hybrid LSTM–GRU architectures. Park *et al.* [22] and Pan *et al.* [23] evaluated the GRU for sequential hydrological processing, reporting favorable results in water-level prediction tasks. Widiasari *et al.* [24] developed a hybrid LSTM-GRU framework intended to improve robustness by simultaneously capturing long- and short-term dependencies.

B. Non-Linear Algorithms: Classical and Ensemble Approaches

Alongside deep learning approaches, classical non-linear algorithms such as RF and SVM continue to serve as important benchmarks in hydrological research due to their interpretability and robustness in high-dimensional spaces. Random Forest, an ensemble learning method, has been widely adopted for both regression and classification tasks. Soliman *et al.* [25] and Liao *et al.* [26] employed RF to generate rapid flood depth predictions and inundation maps, reporting computational speeds considerably faster than traditional hydrodynamic simulations. RF has also been applied in related hydrological contexts. Anggraini *et al.* [27] used RF to predict sea water levels from atmospheric parameters, while Mobley *et al.* [28] applied it to develop continuous flood hazard maps that identify risks outside designated zones. However, Schoppa *et al.* [29] noted

that while RF is competitive for low-to-medium magnitude floods, it may exhibit inaccuracies during extreme flood events.

SVM has also been applied in flood modelling research. Asaly *et al.* [30] applied SVM for the binary classification of flash flood likelihoods based on meteorological data, while Chang *et al.* [31] integrated Support Vector Regression into a multi-step forecasting system to predict flooding volumes with lead times up to 72 hours. Comparative studies frequently evaluate these classical models against one another. For example, Nhu *et al.* [32] compared decision tree-based algorithms, finding the M5 pruned (M5P) model to outperform RF for specific lake level predictions, whereas Wang *et al.* [33] evaluated six machine learning algorithms for Lake Erie, suggesting the efficacy of ML models over process-based alternatives. Hossoda *et al.* [34] further contributed to this area by developing a hybrid parametric RF framework for urban flood warning systems.

Recent literature indicates a general movement in flood forecasting toward data-driven methods as alternatives to computationally intensive physical simulations. Classical nonlinear models such as RF and SVM remain useful for fast, interpretable hazard mapping but may perform less reliably under the extreme volatility of severe floods. Deep learning architectures, particularly LSTM and GRU, have become prominent in hydrological time-series analysis for their capacity to capture complex temporal and nonlinear dynamics, though studies emphasize the need for careful hyper parameter tuning and hybrid designs to address data imbalance and variable diversity. This context supports a comparative analysis to identify a suitable architecture for specific river systems such as the Mekong River Basin.

III. EXPERIMENT

This section outlines the experimental framework, including a geographical and hydrological description of Nong Khai province—its location on the Mekong River and flood susceptibility—and the data sources and collection methods. It also summarizes the 19-year hydrological time-series dataset obtained from the Mekong River Commission (MRC) and its preliminary statistical analysis.

A. Study Area Description

Nong Khai province, covering an area of 3,275 km^2, is in north-eastern Thailand. It is situated on the southern bank of the Mekong River, which forms its northern border. Across the river lies Lao PDR, specifically Vientiane Prefecture, Vientiane Province, and Bolikhamsai Province. Nong Khai borders Bueng Kan Province to the east, Sakon Nakhon and Udon Thani to the south, and Loei to the west. Nong Khai's border is outlined in green on Fig. 1. The province experiences a tropical Savana climate with distinct wet and dry seasons. The monsoon season occurs from May to October, accounting for most of the region's annual precipitation; flood risk is highest during this period.

B. Data Sources and Collection

The data set used in this study consists of daily water-level measurements over a 19-year period (2006-2025), obtained from five monitoring stations in the MRC monitoring network.

These stations were selected to capture the upstream hydrological context and flow dynamics. The stations are from upstream to downstream:

1. Chiang Saen Station (Thailand)
2. Luang Prabang Station (Lao PDR)
3. Chiang Khan Station (Thailand)
4. Vientiane Station (Lao PDR)
5. Nong Khai Station (Thailand)

Fig. 1 illustrates the geographic locations and spatial distribution of these monitoring stations, in their upstream-to-downstream sequence. The Mekong River in this region is characterized by the following hydrological characteristics: an average annual discharge of 14,500 m^3/s at Nong Khai station;

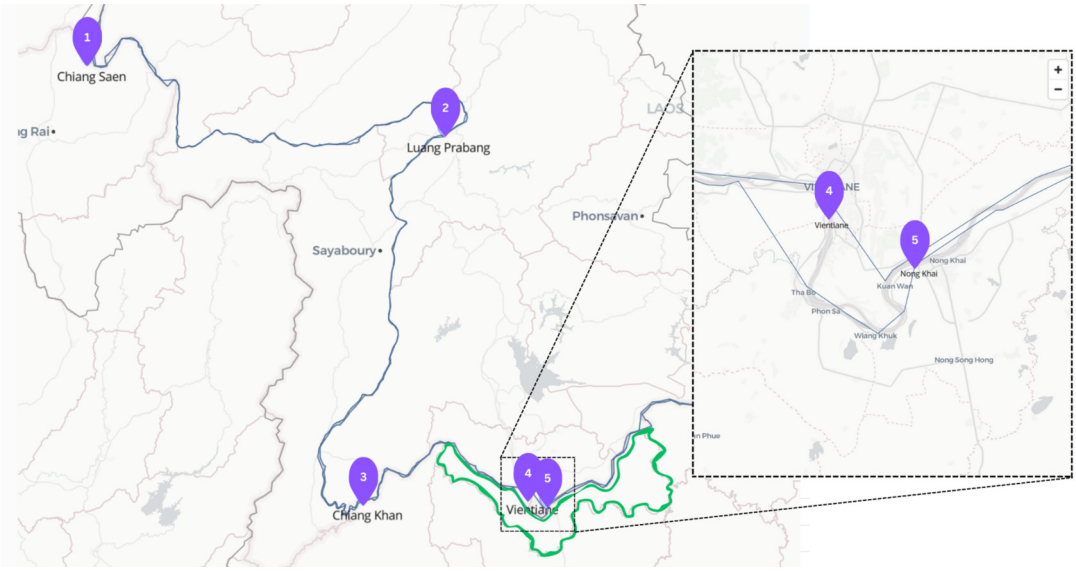


Fig. 1. location of Nong Khai province and Water Station along Mekong River(Source: Mekong River Commission, https://mrcmekong.org).

peak flows of 20,000-35,000 m^3/s during the flood season; and low season flows of 1,000-2,000 m^3/s. To contextualize the flood risk parameters relative to the physical river channel, Fig. 2 presents a schematic cross-section of the Mekong River at the Nong Khai station; this information was obtained from the Mekong River Commission on November 24, 2025, [34]. The vertical axis denotes elevation in meters above Mean Sea Level (MSL), while the horizontal axis represents the cross-sectional distance in meters from a fixed terrestrial reference point. Hydrometric measurements at this station are referenced to a standard zero-gauge elevation of 153.6 m above MSL. The coordinate at a cross-sectional distance of 820 meters marks the location where the riverbed elevation approximates this zero-gauge datum, serving as a baseline for the station's stage monitoring. Based on this vertical datum, operational safety thresholds are defined: the Alarm Level is established at 165.0 m (MSL), and the critical flood level is set at 165.8 m (MSL). While the figure shows a low-flow condition typical during the dry season, during the hydrological peak the water level often exceeds the indicated flood level, inundating adjacent urban and economic areas. These elevation markers are useful for relating hydrometric station measurements to potential inundation risks in the surrounding topography. Fig. 3 presents the distribution of monthly water levels at Nong Khai station, expressed as gauge height relative to the station's zero point. The operational alarm level corresponds to a gauge height of 11 m, and the flood level threshold to 12.2 m. The box-and-whisker plot illustrates the bimodal seasonality and variability differences between the dry and wet seasons. The data reveal that water levels exhibit considerable variability during peak monsoon months, and occasionally during transitional periods, with outliers frequently approaching or exceeding the 12.2 m flood threshold. These excursions above the critical levels highlight the station's susceptibility to severe flood events and support the need for predictive models capable of capturing such high-magnitude fluctuations.

Prior to model training, the raw hydrological data underwent a preprocessing phase to ensure data quality and model convergence. First, a missing value analysis was conducted; minor gaps (less than 24 hours) were filled using linear interpolation, while larger gaps were handled by excluding specific non-continuous periods. Second, to reduce scale differences between stations and facilitate gradient descent in the deep learning models, the data were normalized. Min-Max scaling

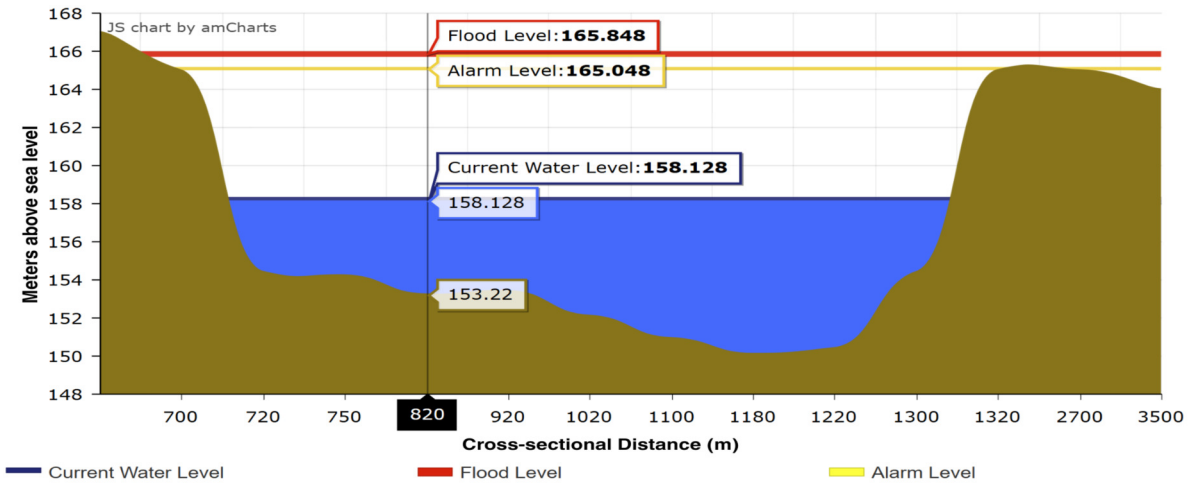


Fig. 2. Schematic cross section with actual water level on November 24, 2025 at Nong Khai station (Source: Mekong River Commission, https://ffw.mrcmekong.org).

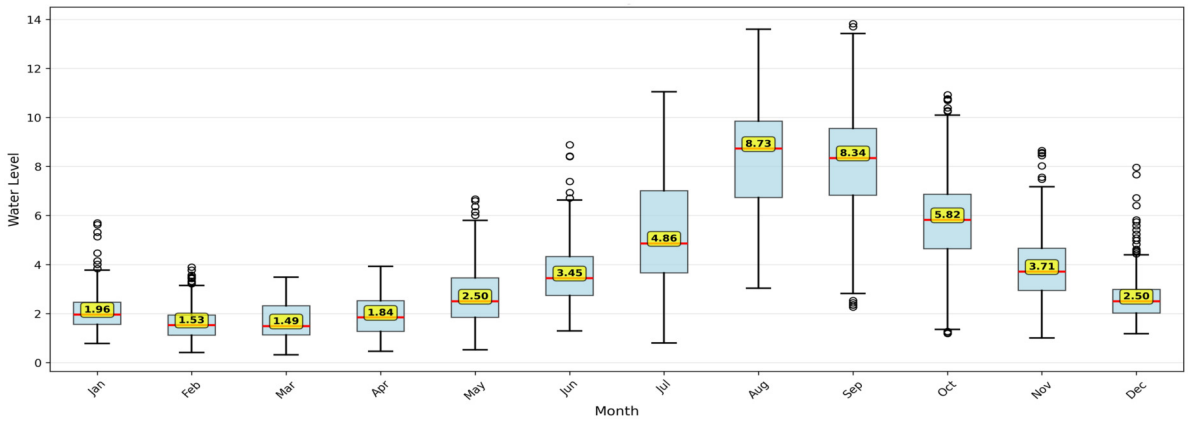


Fig. 3. Box and whisker plot of water-level data at Nong Khai station.

was applied to transform all water-level features into the range [0, 1] using equation (1),

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X_i is the observed water level, and X_{min} and X_{max} are the min and max recorded values in the training set, respectively. Finally, the time-series data were restructured into a supervised learning format using a sliding window approach, where a sequence of past observations (look-back window) was used to predict the water level at the next time step $(t+1)$.

IV. METHODOLOGY

A. Model Architecture and Design

This study compares four supervised learning models for one-day-ahead water-level forecasting. This constitutes a time-series regression problem, characterized by complex temporal dependencies and nonlinear patterns.

Two recurrent neural-network models, LSTM and GRU, are compared with two classical regression models, RF and SVR. The models were selected based on their established efficacy in hydrological modelling and time-series analysis

1) Input Feature Space and Spatiotemporal Structure: To capture the propagation of flood pulses along the Mekong River, the input feature vector is constructed from data collected at five locations. We define the monitoring network as a vector in equation (2).

$$S = [S_{CSA}, S_{LUA}, S_{CKH}, S_{VIE}, S_{NON}] \tag{2}$$

corresponding to the upstream-to-downstream sequence: Chiang Saen (CSA), Luang Prabang (LUA), Chiang Khan (CKH), Vientiane (VIE), and the target station, Nong Khai (NON).

For any given time step t, the input feature vector x_t represents the simultaneous water levels across all five stations as shown in equation (3).

$$x_t = [S^t_{CSA}, S^t_{LUA}, S^t_{CKH}, S^t_{VIE}, S^t_{NON}] \tag{3}$$

where S^t_i denotes the water level at station i at time t.

The objective is to predict the water level at Nong Khai at the next time step ($y_{t+1} = S^{t+1}_{NK}$). To achieve this, the models employ a sliding window of historical data with length k (the look-back period). The input matrix X_t for the deep learning models is thus defined as a sequence of these vectors, as shown in equation (4).

$$X_t = \{x_{t-k}, x_{t-k+1}, \ldots, x_t\} \tag{4}$$

This multivariate structure allows the models to learn local persistence and delayed influences from upstream stations as flood signals propagate from Chiang Saen down to Nong Khai.

2) Long Short-Term Memory (LSTM): The LSTM model is a specialised RNN architecture designed to mitigate the vanishing gradient problem, which limits standard RNNs from learning long-range temporal dependencies. This capability is particularly relevant in hydrological modelling, where water levels are influenced by antecedent meteorological events occurring over extended lag periods. The core component of the LSTM is its memory cell (C_t), which acts as a carrier of information across time steps, regulated by three distinct gating mechanisms: the forget gate (f_t), the input gate (i_t), and the output gate (o_t). The forget gate determines which information to discard from the previous cell state (C_{t-1}). The input gate controls the assimilation of new data. The output gate subsequently regulates the information passed to the hidden state (h_t) and the model's final prediction. This architecture enables LSTM to better capture the complex, long-term temporal dynamics inherent in water-level data. Mathematically, the LSTM transitions are defined by the following equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_{tf}] + b_f) \tag{5}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_i] + b_i) \tag{6}$$

$$\acute{C}_t = \tanh(W_{X0} \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t \odot c_{t-1} + i_t \odot \acute{C}_t \tag{8}$$

$$o_t = \sigma(W_0 \cdot [h_{t-1}, x_t] + b_0) \tag{9}$$

$$h_t = o_t \odot \tanh(C_t) \tag{10}$$

Where W denotes that weight and b represents the bias vectors, σ represents the sigmoid activation function, and $\odot represent selement - wise multiplication.$

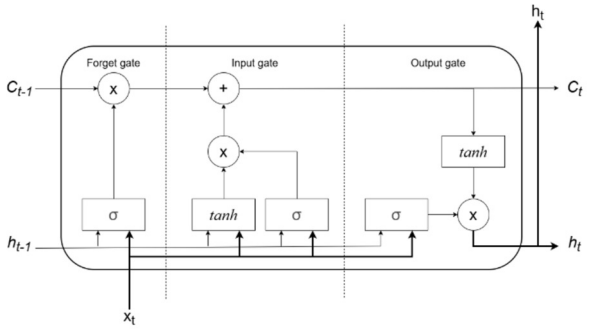


Fig. 4. Structure of a single LSTM cell.

The process, displayed in Fig. 4, begins with the forget gate equation (5) which determines which information from the previous flow state C_{t-1} is irrelevant (e.g., short-term noise) and should be discarded. Next equations (6) and (7) describe the input gate, which computes new candidate values, $\acute{C}_t$, from the current upstream data, x_t, and determines how much to update the cell state. Equation (8) updates the global cell state, effectively blending the long-term seasonal trend with new incoming flood signals. Equations (9) and (10) calculate the final hidden state h_t, which serves as the prediction output.

3) Gated Recurrent Units (GRU): The GRU is a simplified variant of the LSTM, designed to reduce computational

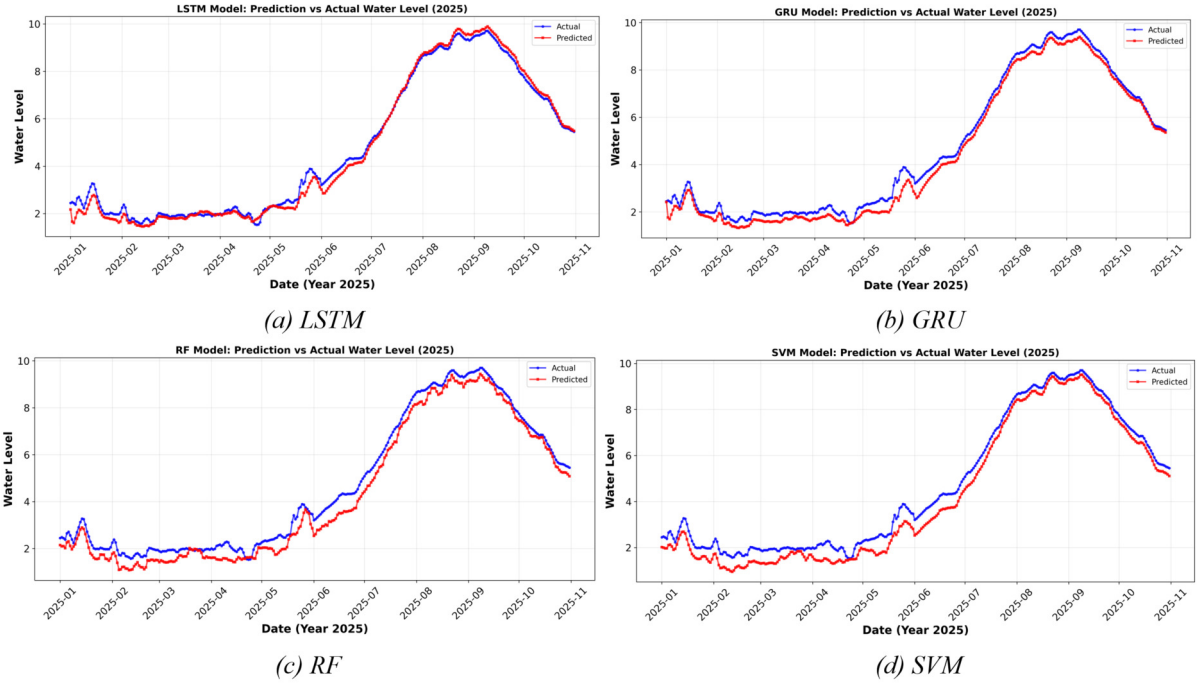


Fig. 5. Visual comparison using (a)LSTM, (b)GRU, (c)RF and (d)SVM method.

complexity while retaining predictive performance. The GRU simplifies the architectural complexity by consolidating the forget and input gates into a single update gate z_t and merging the cell state directly with the hidden state.

This architecture operates using two primary gates: the reset gate r_t and the update gate z_t. The reset gate determines the extent to which past information is ignored, while the update gate controls the balance between retaining previous memory and adding new information. This reduction in parameters can facilitate faster convergence and may mitigate the risks of overfitting on smaller datasets. The GRU operations are formalized as follows.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{11}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{12}$$

$$\tilde{h}_t = \tanh(W \cdot [r \odot h_{t-1}, x_t] + b_h) \tag{13}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{14}$$

Equation (11), the update gate, acts as a sliding mechanism, deciding how much of the past hydrological memory h_{t-1} to retain versus how much new information from the upstream stations x_t to accept. Equation (12) (Reset Gate) allows the model to "forget" past information that is no longer relevant for future prediction. Equation (14) computes the final weighted combination to produce the forecasted water level.

4) Random Forest Model: Unlike the sequence-based RNNs, the RF model treats the input vector X_t as a flattened set of features. It is an ensemble method that aggregates the predictions of N independent decision trees (T) constructed via bootstrap aggregating (bagging).

For a given input x, the final predicted water level $\hat{y}$ is the average of the outputs from all individual trees as shown in equation (15),

$$\hat{y} = \frac{1}{N} \sum_{n=1}^{N} T_n(x) \tag{15}$$

Each tree, T_n, splits the data based on feature thresholds. While RF lacks internal memory for temporal sequences, its ensemble nature helps reduce variance and mitigate overfitting.

5) Support Vector Machine: This study uses Support Vector Regression (SVR), a supervised learning algorithm derived from SVM, to forecast continuous water levels. Unlike standard regression, SVR employs the 'kernel trick,' specifically using the Radial Basis Function (RBF). This transformation maps the multivariate upstream input vector x_t into a higher-dimensional feature space where linear separation becomes feasible. The objective is to determine a continuous function $f(x)$ that predicts the water level (y) at Nong Khai station while maintaining a deviation less than a predefined threshold ϵ(epsilon) from the actual observed targets. The model constructs an optimal regression hyperplane, defined by equation (16),

$$f(x) = w \cdot \phi(x) + e \tag{16}$$

where w is the weight vector determining the orientation of the hyperplane. $\phi(x)$ is the non-linear transformation function mapping the input to a higher dimension, and e is the bias term. The SVR optimisation process minimizes model complexity, enabling it to capture the non-linear flow regimes and dam-regulated fluctuations of the Mekong River.

6) Configuration and Hyperparameters: All four models used daily water-level observations from five stations—Chiang Saen (CSA), Luang Prabang (LUA), Chiang Khan (CKH), Vientiane (VIE), and Nong Khai (NON)—to predict the Nong Khai water level one day ahead. A 60-day look-back window was applied; consequently, the LSTM and GRU received input sequences with dimensions of (60 × 5), while the same data were flattened into 300 input features for the Random Forest and SVR models. Historical Nonvalues were included to represent local water-level persistence, while the upstream stations provide information on downstream flow propagation.

The data were divided chronologically into a training set (2006–2023), a validation set (2024), and an independent test set (January–November 2025). All scaling parameters were estimated using the training data only.

The LSTM and GRU each comprised two recurrent layers with 64 hidden units, a dropout rate of 0.2, and one output neuron. Both models were trained using the Adam optimizer, a learning rate of 0.001, mean squared error loss, a batch size of 32, and a maximum of 100 epochs. Early stopping with a patience of 10 epochs and a ReduceLROnPlateau scheduler with a reduction factor of 0.5 and patience of 5 epochs were applied.

The Random Forest model used 100 trees, a maximum depth of 20, a minimum split size of 5, and a minimum leaf size of 2. The SVR model used an RBF kernel with C = 100, ε = 0.1, and γ = scale. The random seed was fixed at 42 to support reproducibility.

B. Performance Evaluation Metrics

The dataset was divided chronologically to prevent data leakage, as discussed in A.6 Model performance is reported on the independent test set. To evaluate model performance, this study uses two standard error metrics—RMSE (17) and MAE (18)—commonly applied in regression and hydrological forecasting. Together, they characterise the models' prediction errors and support a comprehensive assessment of accuracy and reliability. These metrics were selected based on their widespread use in regression analysis and hydrological forecasting literature. Each metric provides a distinct perspective on the model's error characteristics, and their combined use supports a more complete assessment of model performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i^2)}{N}} \tag{17}$$

$$MAE = \frac{\sum_{i=1}^{N}|O_i - P_i|}{N} \tag{18}$$

The O_i stands for observed value, P_i represents the predicted value and N is the number of observations, respectively. The results from all four models are presented in section V.

V. Results and Discussion

To assess the performance of the four forecasting models, this study uses a two-part evaluation: qualitative hydrograph reconstruction and quantitative error analysis. The performance of LSTM, GRU, RF, and SVM is assessed on the independent test set for the 2025 hydrological year.

A. Qualitative Assessment: Hydrograph Reconstruction

The qualitative evaluation focused on a visual comparison of the predicted water-level trajectories against the observed data. As shown in Fig. 5, all four models showed the ability to capture the primary seasonal dynamics of the Mekong River, including the transition between the low-variance dry season and the high-variance wet season. However, the models differed in their responses to short-term fluctuations and peak conditions.

Deep Learning Architectures (LSTM & GRU): The LSTM model, as shown in Fig. 5(a), produced a prediction trajectory that closely tracked the overall trend. The LSTM produced smoother predictions, suggesting that the model may have captured the underlying non-linear hydrological trends rather than memorizing stochastic noise. This smoothing effect may be operationally relevant, suggesting that LSTM's gating mechanisms, specifically the forget gate, tend to filter out transient variances that do not represent genuine flood pulses. The GRU model, as shown in Fig. 5(b) displayed comparable behavior, capturing the primary seasonal peaks and troughs while attenuating minor variability.

Classical Architectures (RF & SVM): The Random Forest and SVR predictions, shown in Fig. 5(c) and Fig. 5(d), followed short-term fluctuations in the observed water levels more closely than the recurrent models. This behaviour suggests that the classical models were more sensitive to local variations in the input data. Although such sensitivity may improve the fit during relatively stable periods, it may also reduce robustness when the observed fluctuations are caused by noise or short-lived disturbances. Therefore, the visual results should not be interpreted as direct evidence of overfitting; rather, they suggest that the RF and SVR models produced less-smoothed forecasts and may require further evaluation using training–validation error differences, extreme-event performance, and false-alarm rates.

B. Qualitative Assessment: Hydrograph Reconstruction

Table I provides performance metrics across all four models.

TABLE I
EVALUATION OF THE PERFORMANCE OF LSTM, GRU, RF AND SVM.

Model	RMSE	MAE
LSTM	0.3459	0.1196
GRU	0.4059	0.1647
RF	0.6123	0.3749
SVM	0.6099	0.3720

Regarding RMSE, the LSTM model recorded the lowest error value of 0.3459. The GRU followed with a moderate RMSE of 0.4059. By contrast, RF and SVM showed substantially higher RMSE values of 0.6123 and 0.6099, respectively, under the evaluation setting used in this study. A similar trend was observed in the MAE analysis. The LSTM recorded the lowest MAE of 0.1196, consistent with its superior RMSE performance relative to the classical benchmarks.

The LSTM's RMSE is approximately 43% lower than that of the RF and SVM models, suggesting that sequence-based memory architectures may more effectively represent the complex, non-stationary fluctuations of the Mekong River

in this dataset. In particular, the LSTM showed relatively low error rates during the high-variance wet season, suggesting that it may capture the overall trend in water levels even under rapidly changing conditions.

C. Discussion of Architectural Efficacy

LSTM's relatively strong performance may be related to its capacity to mitigate the vanishing gradient problem and to capture the Mekong River's extended hydrological lag periods. Although the GRU offers a more computationally efficient architecture, its slightly higher RMSE (0.4059) is consistent with the possibility that LSTM's additional parameterization may better accommodate the temporal complexities of this dataset. By contrast, the classical RF and SVM models showed higher error values, which may partly reflect their absence of sequential processing mechanisms.

VI. Conclusion

This study compared LSTM, GRU, RF, and SVR for one-day-ahead water-level forecasting at Nong Khai using daily observations from five monitoring stations along the Mekong River. Among the four models, LSTM recorded the lowest reported forecasting error, followed by GRU, while RF and SVR showed higher errors. These results suggest that recurrent models may be more suited for capturing the temporal and spatial relationships that influence downstream water levels in this setting.

The main contribution of this study is a systematic multi-station comparison that identifies LSTM as a potentially promising forecasting component for flood early-warning development in Nong Khai. When integrated with real-time monitoring and decision-support tools, the model may serve as a component of a hydrological Digital Twin—a continuously updated virtual representation of the river system—for the Mekong River Basin.

The findings are limited to using water-level data from five mainstream stations, a single target station, and a single evaluation year. Additional hydrological variables—including rainfall, discharge, tributary inflow, reservoir operations, soil moisture, and topographic data—were not incorporated in the current study. Accordingly, recalibration and independent validation would be advisable before applying the model to other river systems. Future work could incorporate additional hydrological variables, evaluate multiple forecasting horizons, quantify prediction uncertainty, and investigate advanced models for basin-scale forecasting and inundation mapping.

VII. Acknowledgement

The authors gratefully acknowledge the agencies that provided data for this study. The authors thank the Nong Khai Hydrological Division (Water Resource Office 3), the Nong Khai Irrigation Project (Regional Irrigation Office 5), and the Nong Khai Meteorological Station (Upper Northeastern Meteorological Centre) for sharing data and expertise on regional water levels and flood events. The authors also acknowledge the Centre of Multidisciplinary Innovation Network Talent (MINT Centre), Faculty of Interdisciplinary Studies, Department of Technology and Engineering, Khon Kaen University, Nong Khai Campus, Thailand, for providing research support, tools, equipment, and computing resources.

VIII. Author Contributions

Conceptualization, T.W., C.S., W.W., T.R., J.M., T.I., W.T., A.S., and A.N.; Methodology, T.W., C.S., W.W., T.R., T.I., W.T., A.S., and A.N.; Software, T.W., W.T.; Validation, T.W., C.S., W.W., T.R., J.M., and A.N.; Formal analysis, T.W., T.R., W.T. and A.N.; Investigation, T.W. and A.N.; Data curation, T.W., T.I., W.T.; Writing—original draft, T.W. and A.N.; Writing—review and editing, T.W., C.S., W.W., T.R., J.M., T.I., W.T., A.S., and A.N.; Supervision, A.N.; Project administration, A.N. All authors have read and agreed to the published version of the manuscript.

References

[1] T. Boonnuk, K. Poomphakwaen, and N. Kumyoung, "Application for simulating public health problems during floods around the Loei River in Thailand: The implementation of a geographic information system and structural equation model," *BMC Public Health*, vol. 22, no. 1, Art. no. 1651, 2022.

[2] S. Chantee and T. Mayakul, "Modeling flood susceptible areas using deep learning techniques with random subspace: A case study of the Mae Chan Basin in Thailand," *International Journal on Advanced Science Engineering Information Tecnology*, vol. 14, no. 1, pp. 344-357, 2024.

[3] H. H. Loc, A. Emadzadeh, E. Park, P. Nontikansak, and R. C. Deo, "The great 2011 Thailand flood disaster revisited: Could it have been mitigated by different dam operations based on better weather forecasts?," *Environ. Res.*, vol. 216, Art. no. 114493, 2023.

[4] W. Ariyasirichot, "The policy design on preparation and coping with flash flood Mekong River Basin entrance inequality deduction form disaster: A case study of Mueang Nong Khai District Nong Khai Province, Thailand," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 3, pp. 4106–4112, 2021.

[5] N. Tsumita, S. Piyapong, R. Kaewkluengklom, S. Jaensirisak, and A. Fukuda, "Flood susceptibility mapping of urban flood risk: Comparing autoencoder multilayer perceptron and logistic regression models in Ubon Ratchathani, Thailand," *Natural Hazards*, vol. 121, no. 15, pp. 17833–17867, 2025.

[6] W. Pratoomchai, C. Ekkawatpanit, N. Yoobanpot, and K. T. Lee, "A dilemma between flood and drought management: Case study of the Upper Chao Phraya flood-prone area in Thailand," *Water*, vol. 14, no. 24, Art. no. 4056, 2022.

[7] T. Wiangwiset *et al.*, "Design and implementation of a real-time crowd monitoring system based on public Wi-Fi infrastructure: A case study on the Sri Chiang Mai Smart City," *Smart Cities*, vol. 6, no. 2, pp. 987–1008, 2023.

[8] I. M. Chathuranika *et al.*, "Comparison of two hydrological models, HEC-HMS and SWAT in runoff estimation: Application to Huai Bang Sai tropical watershed, Thailand," *Fluids*, vol. 7, no. 8, Art. no. 267, 2022.

[9] M. N. Adli Zakaria *et al.*, "Exploring machine learning algorithms for accurate water level forecasting in Muda River, Malaysia," *Heliyon*, vol. 9, no. 7, Art. no. e17689, 2023.

[10] V. K. Sharma, A. K. Shukla, V. M. Chowdary, S. Saran, and S. K. Srivastav, "Integration of recurrent neural network models and flood inundation extent libraries for flood forecasting," *Remote Sensing Applications: Society and Environment*, vol. 39, Art. no. 101649, 2025.

[11] T. T. Hieu, T. Q. Chieu, D. N. Quang, and N. D. Hieu, "Water level prediction using deep learning models: A case study of the Kien Giang River, Quang Binh Province," *River*, vol. 2, no. 4, pp. 468–479, 2023.

[12] M. Cho, C. Kim, K. Jung, and H. Jung, "Water level prediction model applying a long short-term memory (LSTM)–gated recurrent unit (GRU) method for flood prediction," *Water*, vol. 14, no. 14, Art. no. 2221, 2022.

[13] V. Atashi, H. T. Gorji, S. M. Shahabi, R. Kardan, and Y. H. Lim, "Water level forecasting using deep learning time-series analysis: A case study of Red River of the North," *Water*, vol. 14, no. 12, Art. no. 1971, 2022.

[14] M. Zhang *et al.*, "Assessing hydropower impacts on flood and drought hazards in the Lancang–Mekong River using CNN-LSTM machine learning," *Water*, vol. 17, no. 15, Art. no. 2352, 2025.

[15] Q. Chai, S. Zhang, Q. Tian, C. Yang, and L. Guo, "Daily runoff prediction based on FA-LSTM model," *Water*, vol. 16, no. 16, Art. no. 2216, 2024.

[16] F. J. Hop, R. Linneman, B. Schnitzler, A. Bomers, and M. J. Booij, "Real time probabilistic inundation forecasts using a LSTM neural network," *Journal of Hydrology*, vol. 635, Art. no. 131082, 2024.

[17] M. Hasyimi, L. Nada, and F. Zain, "Model prediksi ketinggian air disemua pintu muka air Sungai Ciliwung dengan menggunakan metode long short-term memory," *Journal of Innovative and Creativity (Joecy)*, vol. 5, no. 2, pp. 21964–21973, 2025.

[18] G. K. Fordjour and A. Kalyanapu, "Implementing a GRU neural network for flood prediction in Ashland City," vol. 12, no. 3, 2024.

[19] A. Gharehbaghi, R. Ghasemlounia, F. Ahmadi, and M. Albaji, "Groundwater level prediction with meteorologically sensitive gated recurrent unit (GRU) neural networks," *J. Hydrol.*, vol. 612, Art. no. 128262, 2022.

[20] Y. Zhang, Z. Zhou, J. Van Griensven Thé, S. X. Yang, and B. Gharabaghi, "Flood forecasting using hybrid LSTM and GRU models with lag time preprocessing," *Water*, vol. 15, no. 22, Art. no. 3982, 2023.

[21] F. Manokij, K. Sarinnapakorn, and P. Vateekul, "Cascading models of CNN and GRU with autoencoder loss for precipitation forecast in Thailand," *ECTI Trans. Comput. Inf. Technol.*, vol. 15, no. 3, pp. 333–346, 2021.

[22] K. Park, Y. Seong, Y. Jung, I. Youn, and C. K. Choi, "Development of water level prediction improvement method using multivariate time series data by GRU model," *Water*, vol. 15, no. 3, Art. no. 587, 2023.

[23] M. Pan *et al.*, "Water level prediction model based on GRU and CNN," *IEEE Access*, vol. 8, pp. 60090–60100, 2020.

[24] I. R. Widiasari and R. Efendi, "Utilizing LSTM-GRU for IoT-based water level prediction using multi-variable rainfall time series data," *Informatics*, vol. 11, no. 4, Art. no. 73, 2024.

[25] M. Soliman, M. M. Morsy, and H. G. Radwan, "Generalized methodology for two-dimensional flood depth prediction using ML-based models," *Hydrology*, vol. 12, no. 9, 2025.

[26] Y. Liao, Z. Wang, C. Lai, and C. Y. Xu, "A framework on fast mapping of urban flood based on a multi-objective random forest model," *Int. J. Disaster Risk Sci.*, vol. 14, no. 2, pp. 253–268, 2023.

[27] A. P. K. Anggraini, T. Trismadi, and A. A. Supriyadi, "Predicting water levels from environmental parameters using random forest models," *Int. J. Eng. Continuity*, vol. 4, no. 1, pp. 39–53, 2025.

[28] W. Mobley, A. Sebastian, R. Blessing, W. E. Highfield, L. Stearns, and S. D. Brody, "Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: A pilot study in southeast Texas," *Nat. Hazards Earth Syst. Sci.*, vol. 21, no. 2, pp. 807–822, 2021.

[29] L. Schoppa, M. Disse, and S. Bachmair, "Evaluating the performance of random forest for large-scale flood discharge simulation," *Journal of Hydrology*, vol. 590, Art. no. 125531, 2020.

[30] S. Asaly *et al.*, "Predicting eastern Mediterranean flash floods using support vector machines with precipitable water vapor, pressure, and lightning data," *Remote Sensing*, vol. 15, no. 11, 2023.

[31] M.-J. Chang *et al.*, "Long-term flooding maps forecasting system using series machine learning and numerical weather prediction system," *Water*, vol. 14, no. 20, 2022.

[32] V. H. Nhu *et al.*, "Daily water level prediction of Zrebar Lake (Iran): A comparison between M5P, random forest, random tree and reduced error pruning trees algorithms," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 8, 2020.

[33] Q. Wang and S. Wang, "Machine learning-based water level prediction in Lake Erie," *Water*, vol. 12, no. 10, Art. no. 2654, 2020.

[34] D. H. Hossoda, R. F. Perez, J. R. B. Tercini, and J. I. G. Bonnecarrère, "Data-driven modeling for urban flood warning systems: A case study in the Guarará Basin, Brazil," *J. Flood Risk Manag.*, vol. 18, no. 3, 2025.



Tanapattara Wongkhamchan is currently a Lecturer in the Computer and Information Science Program, Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai Campus, Thailand. He is also pursuing a Ph.D. degree in Electrical and Computer Engineering at the same faculty. His research interests include artificial intelligence, data analytics, information systems, and smart city technologies.



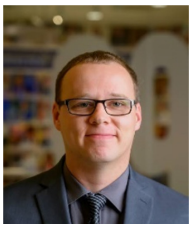
Chayada Surawanitkun received the B.Eng. degree in Electrical Engineering (1st class honor) and the Ph.D. degree in Electrical En gineering from Khon Kaen University, Thai land, in 2009 and 2013, respectively. She was granted the B.Eng. Scholarship from the Industrial/University Cooperative Research Center in HDD Component, supported by the Ministry of Science and Technology, and a Ph.D. scholarship from the Royal Golden Jubilee Ph.D. Program under The Thailand Research Fund. She is Program Chair of Master's and Doctoral programs in Electrical and Computer Engineering of the Faculty of Interdisciplinary Studies, Department of Technology and Engineering, Khon Kaen University, Nong Khai Campus, Thailand. Her research interests include power systems, data storage technology, magnetic random-access memory, thermal analysis of nanodevices, nanotech nology, renewable and sustainable energy, and energy management systems.



Wullapa Wongsinlatam received the B.Sc., M.Sc., and Ph.D. degrees in Applied Mathematics from Khon Kaen University, Thailand, in 2006, 2009, and 2013, respectively. She is currently an Associate Professor and Deputy Dean of the Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai, Thailand. Her research interests include applied mathematics, mathematical optimization, artificial neural networks, metaheuristic optimization, machine learning, and environmental systems modeling. She has published extensively in international journals and is the author of several textbooks and registered computer software copyrights related to intelligent neural networks, forecasting, and optimization.



Tawun Remsungnen received the B.Sc. degree in chemistry from Khon Kaen University, Thailand, in 1997, the M.Sc. degree in computational science from Chulalongkorn University, Thailand, in 1999, and the Dr.rer.nat. degree in computational chemistry from Innsbruck University, Austria, in 2002. He is currently an Assistant Professor with the Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai, Thailand. His research interests include mathematical modeling, molecular dynamics simulations, and scientific computation. He has published numerous research articles in prestigious international journals.



Jakub Michel recieved his PhD in Statis tics from University of New Hampshire in 2023. His current position is an Assistant Professor at High Point University. His current interests are in applied statistical and machine-learning methods that support evidence-based decisions in energy, health, and other real-worldsystems, withanempha sis on rigor, interpretability, and impact. Dr. Michel is a member of SigmaXi,TheScientific Research Honor Society and a MAA Project.



Thalerngsak Wiangwiset received the B.Eng. degree in electrical and computer engineering from Naresuan University, Thailand, in 2004, the M.Sc. degree in computer science from Khon Kaen University, Thailand, in 2008, and the Ph.D. degree in electrical and computer engineering from Khon Kaen University, Thailand, in 2025. He is currently a System Analyst and Developer with the National Telecom Public Company Limited (NT), Thailand. His research interests include cybersecurity, data science, agentic AI, and software engineering. He has received multiple innovation and outstanding research awards from NT and Khon Kaen University.



Wichian Tong-Gate is currently the Director of Hydrology Division 1 (Nong Khai), Water Resources Office 3, Department of Water Resources, Thailand. His work focuses on hydrological monitoring, water resources management, and flood monitoring, particularly along the Mekong River in Nong Khai Province, Thailand.



Apirat Siritaratiwat received the B.Eng. degree in Electrical Engineering from Khon Kaen University, Thailand, in 1992. After working in industry for a few years, he joined the Department of Electrical Engineer ing, Khon Kaen University (KKU) in 1994 and continued his postgraduate study at the University of Manchester, U.K., between 1995 and 1999. He has done research in HDD and is one of the pioneer researchers of ESD/EOS and EMI in recording heads with the KKU Seagate Cooperation Research Laboratory, Faculty of Engineering, Khon Kaen University. Currently, He works at Department of IoT and Information Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang, Bangkok. His research interests are ESD/EOS, EMI, renewable energy and energy management system.



Ariya Namvong received the B.Sc. and M.Sc. degrees in Computer Science and the Ph.D. degree in Computational Science. He is currently a faculty member of the Computer and Information Science Program, Department of Technology and Engineering, Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai Campus, Thailand. He also serves on the Program Committee for the Master's and Doctoral programs in Electrical and Computer Engineering. His research interests include artificial intelligence, data analytics, computational modeling, smart city analytics, data governance, wireless network analytics, and smart energy systems. His research addresses evidence-based decision-making, public-sector digital transformation, sustainability, and quality-of-life improvement.