"A Data-driven Machine Learning Approach for Reservoir Water Level Forecasting"

Seubsuang Kachapornkul[1],
Rangsarit Vanijjirattikhan[1],
Jittiwut Suwatthikul[1],
Kanokvate Tungpimolrut[1],
Toshiyuki Miyachi[2],
Shinsuke Miwa[2]

National Electronics and Computer Technology Center[1] (NECTEC) THAILAND

National Institute Of Information And Communications Technology[2] (NICT) JAPAN

Introduction

- **Dams** play a crucial role in addressing poverty, drought, floods, and water supply issues in ASEAN countries.
- In Thailand, 14 major dams managed by Electricity Generating Authority of Thailand (**EGAT**) contribute significantly to water management and **hydropower generation**.
- **Dam failures** can be caused by hydrological hazards, seismic activity, climate change, aging infrastructure, and inadequate safety management.

Dam Operation

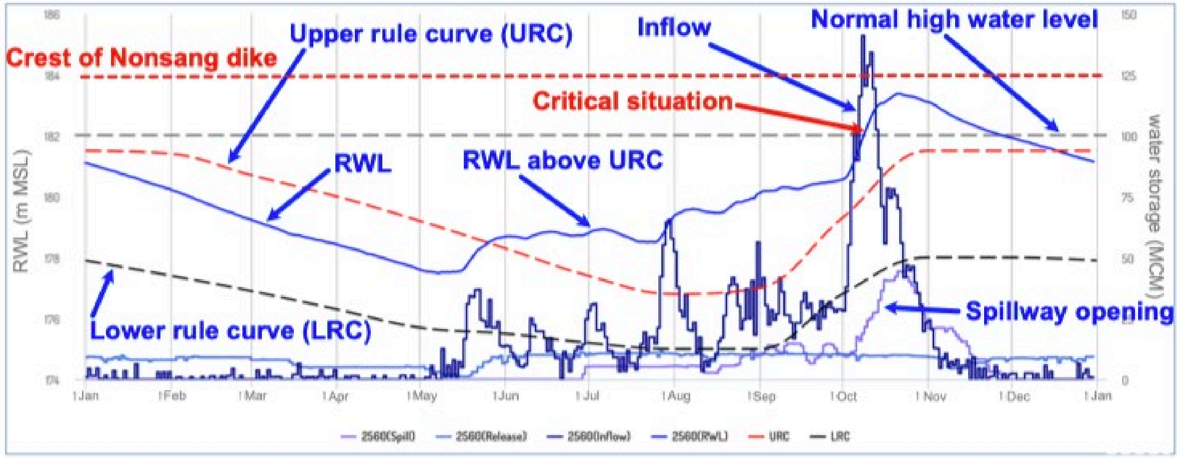
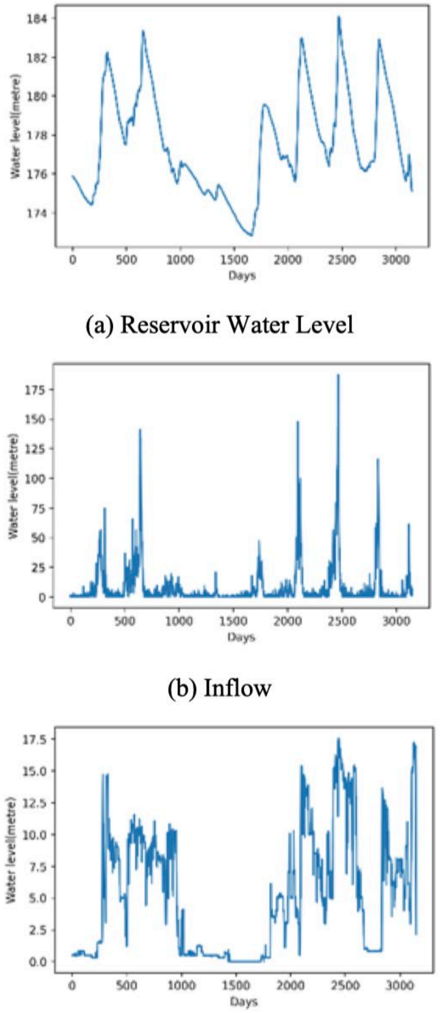


Fig. 2. Water situation at Ubol Ratana dam in 2017.

- **Ensuring dam safety** requires continuous monitoring, maintenance, and a comprehensive operation plan.
- **Managing upstream water storage** effectively helps avoid dam failures caused by excessive or insufficient reserves.
- Maintaining **the Reservoir Water Level (RWL)** between the Upper Rule Curve (URC) and Lower Rule Curve (LRC) is crucial for balancing dam safety and effective water resource management.
- Exceeding the URC can lead to **potential flood risks**, while dropping below the LRC may cause **drought-related issues**, necessitating precise monitoring and controlled water release strategies.

Problem Statement



(a) Reservoir Water Level

(b) Inflow

(c) Outflow

Fig. 3. The relation among RWL, inflow and outflow (Ubol Ratana dam).

- Key hydrological parameters like **rainfall**, **water inflow**, and **reservoir water level (RWL)** are essential for predicting and preventing flood and drought disasters.
- Traditional statistical and machine learning techniques have **limitations** in accurately **predicting Reservoir Water Levels (RWL)** due to the complexity of hydrological systems and extreme weather events making them inadequate for effective dam operation and disaster prevention.
- The **Long Short-Term Memory (LSTM) network** is proposed to leverage historical data and rainfall information for more reliable and extended forecasting periods (5-15 days), aiding proactive dam operations and disaster mitigation.

ASEAN IVO Project

- The **ASEAN IVO** (ASEAN ICT Virtual Organization) is an initiative that fosters regional collaboration among ASEAN countries and Japan. It supports joint research projects focused on ICT-driven innovations to address societal challenges.
- The “**Cyber to Real World Integrated Testbed for Dam Safety Management and Water Governance System**”, an ASEAN IVO project, aims to enhance dam safety and water governance in ASEAN countries The project aims to enhance dam safety awareness, optimize water resource management, and minimize disaster risks and financially supported by **NICT**.
- **NICT** (National Institute of Information and Communications Technology) is a leading research organization in Japan that develops advanced ICT technologies, including high-speed networking, computer network simulations, and cyber-physical systems. NICT supports regional initiatives like ASEAN IVO.

CyReal Framework by NICT

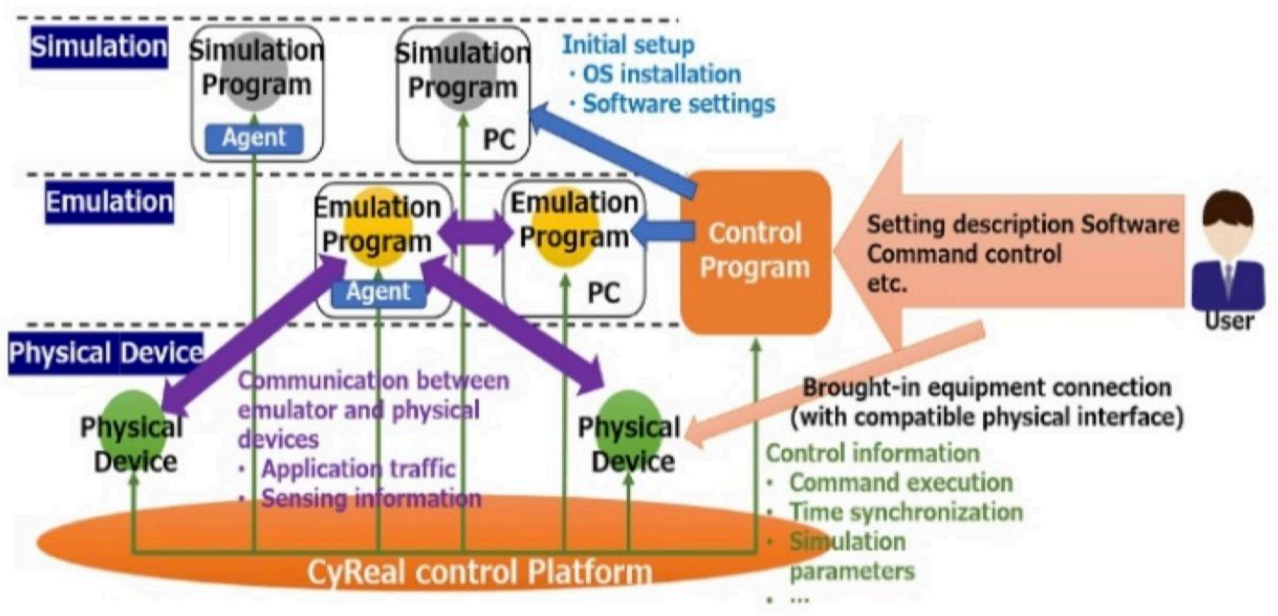


Fig. 1. A sample structure of CyReal environment.

- **Integrated Simulation and Emulation** – CyReal enables seamless integration of simulations, emulations, and real-world systems for testing ICT applications efficiently.
- Cost-Effective R&D – It **minimizes research and development costs** by providing a virtual testbed, reducing the need for real-world infrastructure setup.
- **Disaster Prediction and Response** – The platform supports applications like flood damage prediction and evacuation advisory systems, enhancing disaster preparedness.

Reservoir Water Level Forcasting

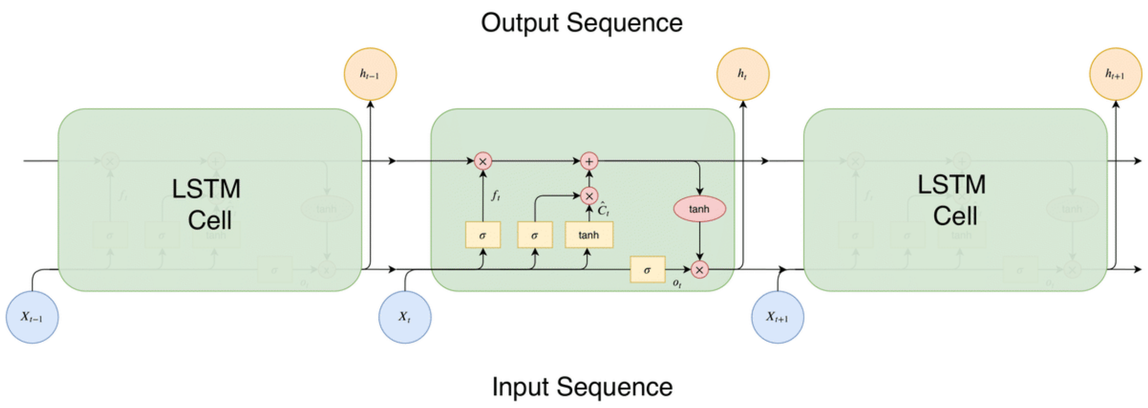
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

$$CC = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \quad (2)$$

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2} \quad (3)$$

- Employs a **Long Short-Term Memory (LSTM) network**, a powerful deep learning approach suitable for time-series forecasting, to predict reservoir water levels effectively.
- The model is trained using **historical data** collected from the Dam Safety Remote Monitoring System (**DS-RMS**), which provides real-time insights into key hydrological factors such as **rainfall**, inflow, and **reservoir water levels**.
- The **evaluation** of the forecasting model is performed using three primary metrics:
  - Root Mean Square Error (RMSE)
  - Correlation Coefficient (CC)
  - Standard Deviation (SD)

Long Short-Term Memory (LSTM) network parameters



$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)$$
$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, X_t] + b_C)$$

$$C_t = i_t \cdot \hat{C}_t + f_t \cdot C_{t-1}$$
$$h_t = o_t \cdot \tanh(C_t)$$

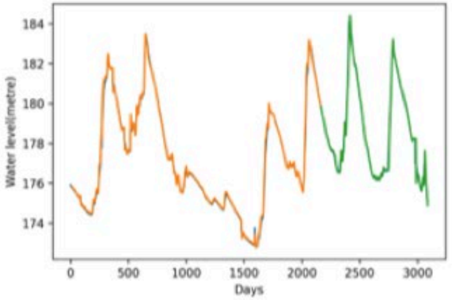
With courtesy from thorirmar.com.

- Training Configuration
  - Dataset Split:

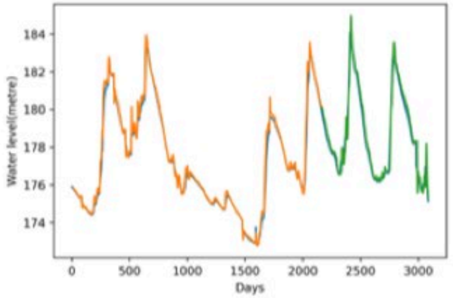
    70% for training

    30% for testing
- Hyperparameters Used:
  - Number of LSTM units: 50
  - Backsteps (time steps): 15
  - Batch size: 25
  - Epochs: 100
- Dataset Characteristics
  - Data collected every 4 hours from Remote Terminal Units (RTUs) at reservoir sites for **6 records per day**.
  - Includes reservoir water level (RWL), rainfall, outflow, and calculated inflow.
  - Total dataset: **3,090 days** of historical records from Ubol Ratana dam between January 1, 2016, to August 20, 2024.

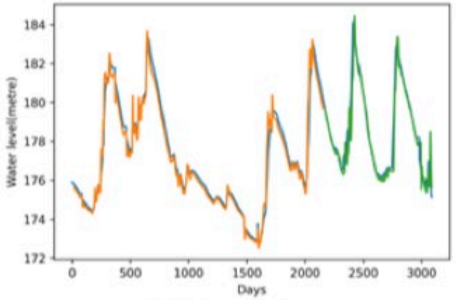
Simulation Results



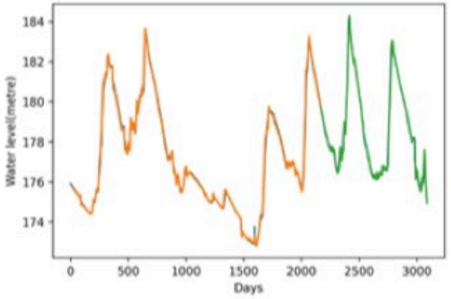
(a) Next 5 days



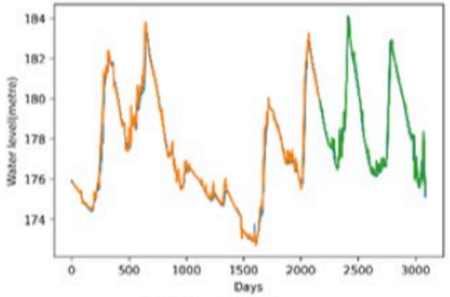
(b) Next 10 days



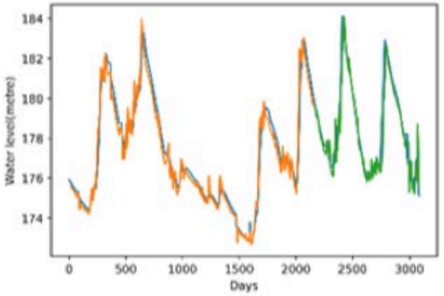
(c) Next 15 days



(a) Next 5 days



(b) Next 10 days



(c) Next 15 days

Only RWL

Using RWL and Rainfall

TABLE I. EVALUATION OF THE PERFORMANCE OF EACH MODEL.

	RMSE	CC	SD
Next 5 days	0.25709	0.99450	2.19423
Next 5 days (with rainfall)	0.24409	0.99383	2.19423
Nest 10 days	0.47525	0.97960	2.19598
Next 10 days (with rainfall)	0.46201	0.97924	2.19598
Next 15 days	0.63887	0.95634	2.19688
Next 15 days (with rainfall)	0.61607	0.96015	2.19688

- Compare the prediction results for the next 5, 10, 15 days of Reservoir Water Level (RWL) with and without the information of rainfall.
- **RMSE, SD** are **increased** and **CC** is **decreased** as the prediction horizon for RWL extends further into the future.
- **Additional Rainfall information** helps reduce RMSE.

Conclusion

- The paper emphasizes the importance of **balancing reservoir water levels** to ensure dam safety and prevent flood and drought-related disasters, highlighting the need for accurate long-term forecasting.
- A Long Short-Term Memory (LSTM) network, utilizing historical data from real-time monitoring, was investigated for **Reservoir Water Level forecasting**, demonstrating promising performance metrics for 5-day, 10-day, and 15-day predictions.
- The model's accuracy improved when rainfall data was incorporated, suggesting that incorporating **additional environmental parameters** could further enhance forecasting performance.

Thank you for your attention