Spatial-Temporal Flood Hazard Mapping Using Integration of Telemetry Data and Prediction Model

Pornnapa Panyadee
OASYS Research Group,
Department of Computer Engineering,
Faculty of Engineering, Chiang Mai University
Chiang Mai, Thailand
pornnapa_panya@cmu.ac.th

Paskorn Champrasert
OASYS Research Group,
Department of Computer Engineering,
Faculty of Engineering, Chiang Mai University
Chiang Mai, Thailand
paskorn@eng.cmu.ac.th

2024 International Conference on Electronics, Information, and Communication (ICEIC) | 979-8-3503-7188-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICEIC61013.2024.10457113

***Abstract*—The flood early warning system can help mitigate the resulting damages by predicting future events. This is achieved through the utilization of data obtained from telemetry stations to predict the values of water levels in the future. Flood hazard maps are considered a tool for representing the potential flood events that occur in an area. This paper proposed a framework to apply the spatial and temporal data to generate flood hazard mapping using the integration of interpolation telemetry station data and a temporal prediction model. The framework consists of two components: 1) the temporal prediction model is applied to water level prediction on hourly and daily scales, and 2) the interpolation of spatial data to generate a flood hazard map. The evaluation results show that the hourly and daily temporal prediction models can predict the water level with an average of MAPE using 500 iterations are 3.17% and 4.88% of training, and 3.48% and 4.72% of testing. Then, the flood hazard map is generated. The accuracy is 70.90% and F1-score is 81.50% compared to the observation flood event.**

***Index Terms*—flood hazard map, temporal prediction**

I. Introduction

The prolonged heavy rainfall often leads to destructive flash floods, causing significant harm to the environment, economy, and people [1]. However, flood losses can be reduced through appropriate management before a flood occurs. Flood early warning systems are essential to facilitate proactive protection in an event. Furthermore, those involved in crisis management rely on real-time alert systems with high expectations for the reliability of the warning alerts. Accurate spatial and temporal prediction is crucial for effective crisis management, minimizing the likelihood of erroneous decision-making in emergencies as much as possible.

The flood early warning system is designed to monitor, predict, warn, and respond to flood events. Data monitoring is typically collected from telemetry stations installed along rivers, including water level, rainfall, and weather [2]. These data are utilized for prediction purposes to estimate flood events. However, the data provides only localized water levels at the installed station, which might not represent the overall water level across the entire area. Therefore, the interpolation of spatial data is the key to providing an overview of the entire area to identify the areas at risk of flooding.

The flood hazard maps are created based on historical flood events and are displayed as spatial data. However, the spatial data is not enough to represent the flood event for a long period. Thus the temporal data are needed to accurately, suitability, and responsible flood prediction [3]. The creation of an effective flood hazard map requires the integration of spatial and temporal data.

Thus, this paper proposes a framework to apply the spatial and temporal data to generate flood hazard mapping using interpolation telemetry station data and a temporal prediction model. The framework is composed of two key components: the temporal prediction model and the utilization of interpolation spatial data to generate a flood hazard map. The temporal prediction model involves the Long Short-Term Memory (LSTM) technique, a type of machine learning, to enhance flood event prediction [4] [5]. The temporal prediction model is divided into hourly and daily temporal prediction models. This is necessary because the time taken for water to traverse the river from upstream to downstream locations typically exceeds one hour. The spatial data will utilize data from the daily temporal prediction model to interpolate the water level values specifically in the river area. The inverse distance weight (IDW) [6] is applied in the proposed to interpolate the water level data with spatial data in the river line. This is compared with the Digital Elevation Model (DEM) in the surrounding areas to estimate the water outflow of the riverbank. The proposed framework can be applied to create a flood map for the responsible authorities to use in planning measures to mitigate potential flood losses

II. Flood Basics

A. Flood Early Warning Systems

Flood early warning systems provide real-time data from telemetry stations to provide timely and effective warnings, and to help risk areas in preparing for and responding to floods. Water level data are considered an important factor in the flood prediction model because they are the basic data for estimating future water levels. Monitored against predefined warning and critical thresholds which water levels the issuance of flood warnings when these thresholds are breached [2]. Moreover, the lag time of river water levels between two stations is another significant factor in the prediction model. The lag represents the time takes for the water to travel from the upstream peak to the downstream. This calculation helps determine the duration it takes for the water to reach its destination. The water levels can be combined with other data (e.g., discharge, rainfall, and flood history) to enhance the accuracy of the prediction model in terms of temporal data.

B. Flood Hazard Maps

Flood hazard maps are maps that show the flood event displayed on a geolocation map. Generally, the flood hazard mapping methods can be divided method into two categories: (1) deterministic methods and (2) non-deterministic methods.

Deterministic methods are used by fundamental principles to simulate and predict flooding through a set of mathematical equations. In [7], the model was used to simulate floodplain inundation. However, this method was usually designed under specific assumptions and cannot fit other real-world conditions. In addition, the models require detailed data that are difficult to collect for large areas.

Non-deterministic methods are the statistical method and machine learning method. The model applied historical flood to determine areas of flooding. the method consists of the water level rainfall and environmental data on flooding. The statistical method, For example, [8] logistic regression (LR), and [9] frequency ratio (FR). Machine learning methods, For example [8] artificial neural network (ANN), [5] long short-term memory (LSTM). In summary, two reasons prove why the performance of machine learning models is better than the statistical models [10]. First, machine learning models can perform complex tasks with limited information; second, machine learning models can represent large complex non-linear systems in a computationally efficient. The principles of LSTM are designed to effectively capture and learn from sequenced data. This paper applies the long short-term memory (LSTM) method in the temporal prediction model.

The resulting temporal data also requires conversion into spatial data using spatial interpolation methods. As mentioned previously, there are several spatial interpolation models such as IDW, Spline, and Kriging. [11] comparison of interpolation methods for depth to groundwater is presented. The results show that IDW has the lowest Relative Error Coefficient compared to other techniques. The IDW technique is simple and effective, especially when the number of data points is small. Thus, this study applied the IDW to interpolate water level data in a river line in terms of spatial data.

III. Spatial Temporal Flood Mapping Framework

This paper proposes a framework to apply the temporal and spatial data to generate flood hazard mapping using the interpolation of telemetry station data and a temporal prediction model. The framework integrates the temporal prediction model and the spatial data to generate flood hazard mapping. The prediction model consists of hourly and daily temporal prediction models. These models are applied to predict multiple outputs of the future water at each telemetry station. The water level verifies the warning level from the telemetry station at each location. When the water level is higher than the warning level, the daily temporal prediction model is applied to predict the water level for the next day. The flood hazard map is created from the prediction values that are converted to spatial data with the IDW interpolation model, and it includes a digital elevation model for predicting flood events in this area next 3 days represented on the map, as shown in Fig. 1.

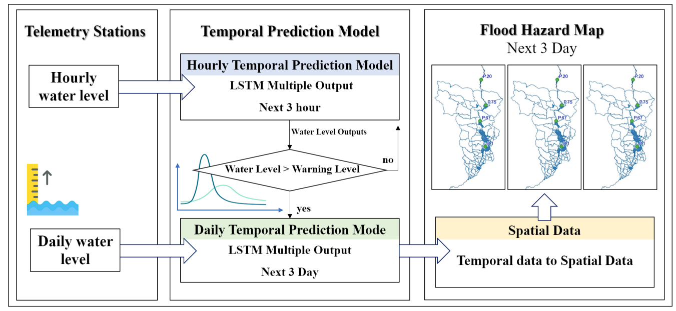


Fig. 1. The Proposed Spatial Temporal Flood Hazard Mapping Framework

A. The Temporal Prediction Model

The LSTM which is a type of recurrent neural network that can learn long-term dependencies in time series data has been applied in the proposed framework. The LSTM consists of cell states and gates. The cell state is stored as the state of the memory cell. The gate is the controller of the data flow which is analogous to values control when to read, write, or forget. Thus, The LSTM is the temporal prediction model which is the process of predicting the future multiple output water level. The temporal prediction model consists of hourly and daily temporal prediction models, as shown in Fig. 1.

The hourly temporal prediction model takes data inputs from two sources: the water level from the telemetry station outside the watershed area which is the upstream station and within the watershed area. It is applied the data from the previous three hours, represented as WL_{t-i}^{Up} and WL_{t-i}^{N}, where Up is the upstream telemetry station, N represents the number of telemetry stations in the watershed area, and i is the previous time. The outputs consist of water level values for the next three hours from the telemetry stations within the watershed area. The output of this method is taken as one of the inputs in daily temporal prediction models.

The daily temporal prediction model takes data inputs from the telemetry stations within the watershed area. The inputs of the model are the water level of the last three values to predict the next three days. The input is the output from the hourly temporal prediction model which is higher than the warning level of the telemetry station at each location. Then, the output is converted to spatial data to generate the flood hazard map.

B. The Spatial Module Evaluation

Flood hazard maps are generated using the outputs of the daily temporal prediction model. The water level at each telemetry station is interpreted as the water level along the river to create spatial data for the river water levels within the watershed. In this paper, the Inverse Distance Weighting (IDW) model is applied for interpolation. The model

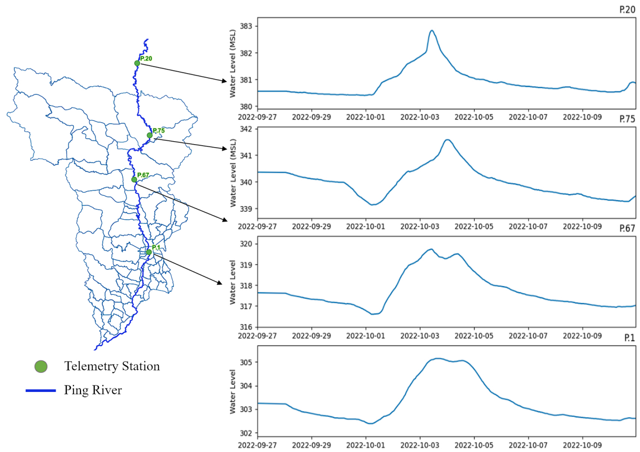


Fig. 2. The Study Area; Ping River, Chiang Mai, Thailand

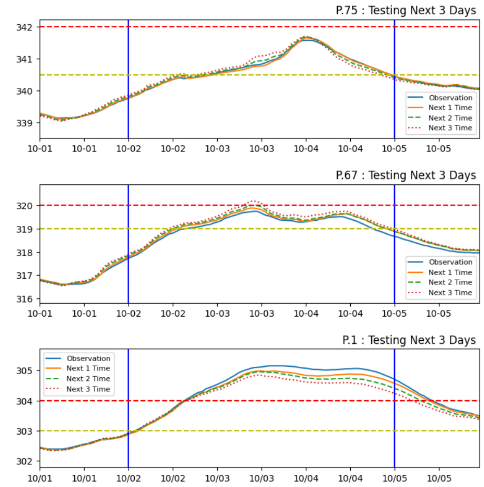


Fig. 3. Testing LSTM Outputs of Three Days of Three telemetry Stations.

is interpolated and estimates unknown values based on its distance from known values points [12]. The unknown values are calculated with the weighted average of the values at the known points. The weights are determined by the inverse of the distance between the unknown and known points. Thus, the river will know the water level values for all points, even if there is no telemetry station installed at that point. The data will be converted into spatial data. The DEM values are compared with the water level values to identify potential flood events. Finally, The flood hazard maps represent the flood events in this area on the map.

IV. The Results

A. Study Area

The study area is the Ping River basin covering six districts in Chiang Mai province, Thailand. The river length is 118 km. The watershed covers an area is 2,146 km^2. The river is the main river of Chiang Mai province, and it flows through economic and community areas. Flooding can have a significant impact on the province. Fig. 2 shows the study area and the position of the telemetry stations. The water level monitoring in this area is conducted by four telemetry stations and one upstream station, named P.1, P.67, P.75, and P.20, respectively. The lag time between P.1 to P.20 takes approximately 24 hours. The telemetry has the hourly observed water level data.

B. Performance Metrics and Parameter Settings

The performance evaluation of the prediction model is the accuracy based on the mean absolute percentage error (MAPE). It calculates the mean of the absolute percentage error difference between predicted values and observed values as depicted Eq. (1). Where O_i represents the observed water level value from the telemetry station, P_i represents the predicted water level value from the telemetry station and n is the number of data.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{O_i - P_i}{O_i}| * 100 \quad (1)$$

The performance evaluation of the spatial data applies the confusion matrix technique. The technique summarizes the performance of a classification model consisting of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of the prediction model. The confusion matrix technique can calculate the metrics including accuracy, precision, recall, and F1-score.

The dataset of the case study covers the period from April 2016 to March 2023 with 61,345 hourly records and 2,557 daily records. The training dataset is 80% of the data with 49,077 hourly and 3,196 daily records. The testing dataset is 20% remaining 20%, with 12,268 hourly records and 640 daily records. The LSTM model parameters include 16 inputs for the hourly model, 9 inputs for the daily model, and 9 outputs for both. The hidden layers consist of 64 neurons for each model, with each model trained between 100 - 500 iterations.

C. Temporal Model Evaluation

The temporal prediction model is evaluated using the MAPE. The evaluation is conducted at iterations of 100, 300, and 500, with results as shown in Table I. The table shows the metrics including the minimum, mean, and maximum of the training and testing process. The MAPE of hourly and daily temporal prediction models for the training processes have accuracy are 3.17% and 4.88%. The testing processes have accuracy are 3.48% and 4.72%, respectively.

Fig. 3 shows a sampling of the water level prediction from the testing process of the hourly model. The graphs represent water levels at the three telemetry stations (P.75, P.67, P1). Each station has four line graphs representing the water level predictions for the next three hours and the observed values. The yellow line represents the warning level, and the red line indicates the critical level. These graphs shows the water level peak in October 2022.

D. Map Evaluation

In Fig. 4, the flood hazard map represents the Ping River in Chiang Mai province from 3 to 5 October 2022. The output exceeded the warning level on 2 October 2022 from the daily temporal prediction model. During this period, Chiang Mai province suffered extensive damage from Typhoon Noru. The observation map is data from the Geo-Informatics and Space

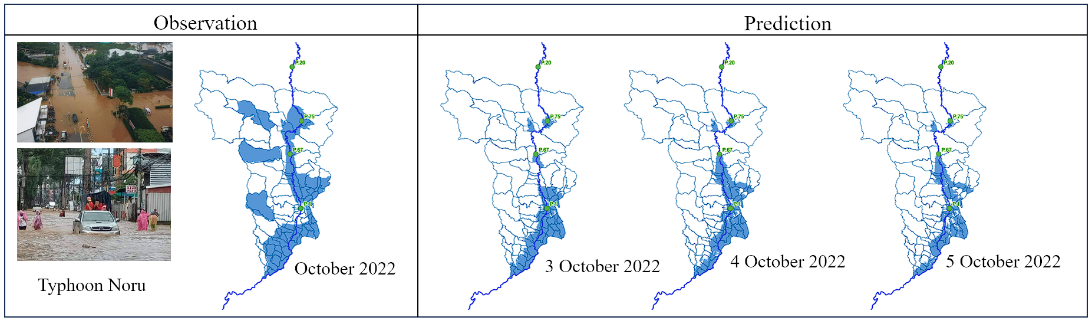
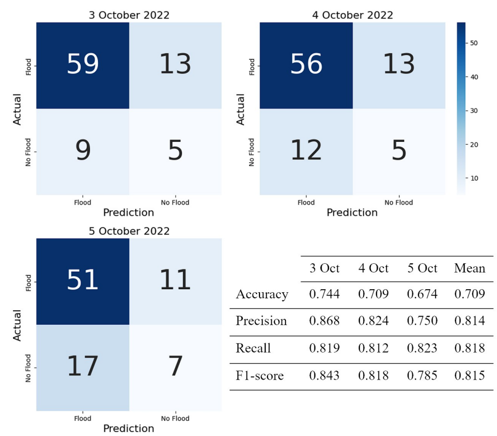


Fig. 4. Flood Map with Typhoon Noru on October 2022

TABLE I
THE MAPE(%) OF TEMPORAL PREDICTION MODULE

	Iteration					
	Hourly Model			Daily Model		
Matrix	100	300	500	100	300	500
Training Process						
Min	3.95	2.96	2.92	4.89	4.86	4.85
Mean	4.36	3.34	3.17	4.76	4.90	4.88
Max	4.80	3.64	3.41	4.79	4.93	4.90
Testing Process						
Min	4.33	3.21	3.15	4.72	4.70	4.69
Mean	4.85	3.65	3.48	4.76	4.76	4.72
Max	5.25	3.92	3.78	4.79	4.76	4.75



	3 Oct	4 Oct	5 Oct	Mean
Accuracy	0.744	0.709	0.674	0.709
Precision	0.868	0.824	0.750	0.814
Recall	0.819	0.812	0.823	0.818
F1-score	0.843	0.818	0.785	0.815

Fig. 5. Confusion Matrix to Flood Map Evaluation on October 2022

Technology Development Agency (GISTDA) in Thailand, which only provides monthly-scale data.

The accuracy of the flood hazard map was evaluated at the sub-district level by predicting flood occurrence covering 86 sub-districts, as follows in Fig 5. The overall prediction accuracy for October is 70.90%. However, the F1-score is 81.50% , as follow Fig 5.

V. Conclusion

This paper proposes to integrate the spatial and temporal data to generate flood hazard mapping using interpolation telemetry station data and a temporal prediction model. The case study is in the Ping River basin, Chiang Mai province, which is an economically significant area in the province. The evaluation results of the proposed model show the MAPE values of the hourly and daily temporal prediction model using LSTM are 3.17% and 4.88% in the training process. Moreover, the accuracy result of the flood hazard map is 70.90% and the F1-score is 81.50% of the model from the confusion matrix technique. The flood hazard map result shows low accuracy because of the monthly cumulative nature of the observation data, whereas the proposed model operates at a daily scale, leading to high bias. Future work will further enhance the correlation parameters with temporal and spatial data, and adjust the prediction model to support variation of parameters.

References

[1] H. Hapuarachchi, Q. Wang, and T. Pagano, "A review of advances in flash flood forecasting," *Hydrological processes*, vol. 25, no. 18, pp. 2771–2784, 2011.

[2] A. Wannachai, S. Aramkul, B. Suntaranont, Y. Somchit, and P. Champrasert, "Hero: Hybrid effortless resilient operation stations for flash flood early warning systems," *Sensors*, vol. 22, no. 11, p. 4108, 2022.

[3] S. Zhong, C. Wang, Z. Yu, Y. Yang, and Q. Huang, "Spatiotemporal exploration and hazard mapping of tropical cyclones along the coastline of china," *Advances in Meteorology*, vol. 2018, 2018.

[4] C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, "Deep learning with a long short-term memory networks approach for rainfall-runoff simulation," *Water*, vol. 10, no. 11, p. 1543, 2018.

[5] Z. Fang, Y. Wang, L. Peng, and H. Hong, "Predicting flood susceptibility using lstm neural networks," *Journal of Hydrology*, vol. 594, p. 125734, 2021.

[6] F.-W. Chen and C.-W. Liu, "Estimation of the spatial rainfall distribution using inverse distance weighting (idw) in the middle of taiwan," *Paddy and Water Environment*, vol. 10, pp. 209–222, 2012.

[7] W. Li, K. Lin, T. Zhao, T. Lan, X. Chen, H. Du, and H. Chen, "Risk assessment and sensitivity analysis of flash floods in ungauged basins using coupled hydrologic and hydrodynamic models," *Journal of Hydrology*, vol. 572, pp. 108–120, 2019.

[8] M. Rahman, C. Ningsheng, M. M. Islam, A. Dewan, J. Iqbal, R. M. A. Washakh, and T. Shufeng, "Flood susceptibility assessment in bangladesh using machine learning and multi-criteria decision analysis," *Earth Systems and Environment*, vol. 3, pp. 585–601, 2019.

[9] M. S. Tehrany, B. Pradhan, and M. N. Jebur, "Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method," *Stochastic environmental research and risk assessment*, vol. 29, pp. 1149–1165, 2015.

[10] A. B. Massada, A. D. Syphard, S. I. Stewart, and V. C. Radeloff, "Wildfire ignition-distribution modelling: a comparative study in the huron–manistee national forest, michigan, usa," *International journal of wildland fire*, vol. 22, no. 2, pp. 174–183, 2012.

[11] Y. Sun, S. Kang, F. Li, and L. Zhang, "Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the minqin oasis of northwest china," *Environmental Modelling & Software*, vol. 24, no. 10, pp. 1163–1170, 2009.

[12] G. Y. Lu and D. W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," *Computers & geosciences*, vol. 34, no. 9, pp. 1044–1055, 2008.