DEVELOPMENT OF A HYDRODYNAMIC-SIMULATION-INFORMED RANDOM FOREST METHOD FOR EFFICIENT SPATIOTEMPORAL FLOOD PREDICTION

効率的な時空間洪水予測に向けた氾濫解析結果を活用したランダムフォレスト手法の開発

March 2026

LIU KEXIN

Graduate School of
Environmental, Life, Natural Science and Technology
(Doctor's Course)
OKAYAMA UNIVERSITY

Table of Contents

Abstract

Over the past decades, flood disasters have become increasingly severe with recent occurrences across entire Japan, leading to severe fatality and property loss. Furthermore, the frequency of flood occurrences is expected to increase due to climate change, which brings challenges to the existing flood managements. The need for flood forecasting, therefore, is increasing in order to reduce flood damage through immediate non-structural responses under emergency situations of excessive floods. Two important criteria for flood forecasting system are accuracy, the details of spatiotemporal flood propagation, and efficiency such that sufficient lead time between prediction and actual occurrence of flood disaster could be provided. Taking advantage of gradually refined database on topography and advancing techniques to provide observations with high quality, the conventional flood analysis approach through hydrodynamic (HD) models is able to provide flood simulation results with high resolution and accuracy. Yet the high computational cost still significantly restrains the real-time applications of HD models over large scales. The gradually developing method of machine leaning (ML) models, in comparison, significantly reduces computational cost by formulating nonlinear relationships between dataset. However, a large amount of high-quality training data is required for the construction of model, and such spatiotemporal flood propagation observations, especially data on excessive floods, are significantly scarce. A method that could provide accurate flood predictions with efficiency is still in need. Therefore, this research proposed an augmentation method of the HD model and ML model, the random forest (RF) model targeting for efficient spatiotemporal flood prediction induced by practical rainfall events.

Targeting for the prediction of detailed spatiotemporal inundation depths at each location, this approach attempts to resolve the limitation on data scarcity of high-quality training data by taking advantage of highly accurate simulation results of HD model and fulfill the efficiency requirements through the usage of efficient ML models with an iterative prediction process. The main procedures of the approach are a) data generation through flood analysis using HD model, b) training phase for RF model using simulation results as the training data, and c) prediction for untrained rainfall pattens through iterative prediction process.

Numerical experiments were carried out as a validation for the proposed augmentation approach. Flood analysis under constant rainfall conditions was performed with of HD models over a simple bowl-shaped hypothetical terrain setting. Using simulation results as the training dataset, RF models were constructed with different feature selection investigating necessary factors for optimal prediction accuracy. Tested against constant rainfall with untrained intensity, the results suggested that prior state water depth information of adjacent meshes significantly improved model performance by providing more accurate classification on inundation areas and predictions of water depths. Evaluated then over

complex terrain settings with various land use conditions, it was concluded that the representation of river boundary conditions as inflow and outflow discharge increased model accuracy since the control volume within computational domain was better reflected. The features for land use indicator as categorical features, on the other hand, do not significantly affect model accuracy. The iterative prediction process investigation suggested that this approach was sensitive to initial error and error accumulation during long-term predictions. Under the settings of the constant rainfall conditions, an iteration interval of 2.0-hr was able to provide high accuracy predictions with overall RMSE within 0.02 m.

The downstream of Karube river basin, where severe pluvial flood inundations occurred multiple times in history, was selected as the target area to evaluate the applicability of the augmentation method over practical terrain conditions. Adopting the optimal hyperparameter settings from numerical experiments, the RF models were trained with simulation results from HD models under the same constant rainfall conditions. Using 2.0-hr iterative prediction interval, the models resulted in a slightly compromised performance in the delineation between land uses. Certain areas of roads and residential areas were falsely predicted as inundated. The prediction for water depths also suggested error accumulation due to iterations. Accuracy of models decreased significantly around the time when overflow occurs for all land uses. Even when provided with accurate prior information, significant underestimation by more than 0.2 m was still observed at 8 hr. In comparison, the model constructed with an interval of 10-min, with all other parameter and conditions being the same, resulted in improved accuracy when predicted non-iteratively, suggesting that shortened interval better reflects immediate hydrodynamic changes. On the other hand, a shortened interval required a greater robustness of RF model due to increased number of iterations, and the model when tested against historical rainfall events of 2021 resulted in a significant degradation in performance. The observations highlighted the importance of incorporating more diverse rainfall patterns in the construction of training data, and refining feature selections in order to improve model generalization capability for practical rainfall conditions.

The selection of hypothetical design rainfalls was investigated with a simplified temporal distribution equation where different shapes, peak arrival time, peak rainfall intensity and calculation time were considered. Refinement of features was first carried out by constructing multiple RF models with the same 6 patterns of training dataset. The predictive accuracy was significantly improved by replacing prior state water depth and elevation of adjacent meshes into prior water level of adjacent meshes. Even though the two sets of features represent the same physical property, RF model interpreted dynamic features more importantly over static features and resulted in better performance. Revising representation of land use from categorical features into dynamic ones as runoff coefficient, a state-based variable shifting from unsaturated to saturate value when cumulative rainfall exceeds the saturation threshold of each land use, also increased model accuracy. The representation of inflow and

outflow conditions at river boundaries was also refined as water levels considering data availability, and the model including such features resulted in the highest consistency in comparison to simulation results in terms of both identifying inundation area and predicting detailed water depths.

How the diversity of training data influence predictive performance was then evaluated using the above-mentioned features. The prediction for water depths in rivers resulted in a notably high initial RMSE regardless of target rainfall pattern compared to other land uses due to small percentage of meshes as well as cold start from initial state of normal water level. Overall, the performance of RF models was compromised if either the shape or temporal distribution of the target event falls outside the patterns in the training data. Including diverse patterns of training data enhanced the generalization capability of the model. On the other hand, such large database also led to the potential for conservative prediction because of the ensembled nature of RF regression algorithms such that predicted results tend to average towards mean values when targeting extreme or rara events.

Finally, assessments were carried out testing the capability of the RF models, trained with simplified hypothetical rainfall patterns, in reproducing spatial and temporal propagation of inundation under actual rainfall events. Two rainfall events that led to inundation at the Karube region in recent years, an extreme scale rainfall event of 2018 and a medium-scaled 2021 rainfall event, were selected as the target event for prediction. For both rainfall events, the model constructed with training data of 5-hr calculations resulted in significant accuracy degradation at the time when the cumulative rainfall of the target event exceeded the maximum range in the training data. In comparison, the model further including training data of extreme scale of 10-hr durations significantly improved the prediction performance with a final overall RMSE of 0.15 m and 0.05 m for the two rainfall events respectively. Such difference in model performance provided evidence for the limited capability of RF models to extrapolation, which highlighted the importance of careful selection of training data in correspondence to the target rainfall event. Comparison of contour maps of flood arrival time of the latter model suggested an overall high agreement of flood extent prediction for both of the event with slight delay observed in certain areas, suggesting the predictions of the current model falls on the danger end and might induce risk-underestimation in practical applications. Overall the model resulted in high performance where scatter plot of density aligned closely to the perfect prediction line. Due to error accumulation over long-term iterations, slight bias was observed for the maximum depth prediction where more than half of the meshes were overestimated by 0.10 m to 0.15 m for 2018 rainfall and approximately 30% of meshes were underestimated by 0.05 m for 2021 rainfall event. This prediction behavior again provided evidence that the inclusion of training data on extreme events improved the generalization ability of the model while also introducing the potential for predictions biasing towards averaging at shallower depths.

In terms of computational cost, the HD simulation through GPGPU required 140 min and 114 min for the 2018 and 2021 rainfall events respectively. In comparison, the prediction time of RF model

using CPU was 8.8 min and 7.3 min, which averages around 1.5 s per iteration. Considering that this cost also included the time consumed in data reconstruction between iterations, the RF model was significantly efficient compared to HD models, and prediction process was expected to be further accelerated if performed on GPU.

In conclusion, this research proposed an iterative hydrodynamic-simulation-informed RF model targeting for efficient spatiotemporal flood predictions under practical rainfall conditions. With detailed investigations regarding rainfall patterns in training data, feature selection and hyperparameter tuning, this method was able to reproduce step by step spatial flood depth predictions highly consistent with simulation results as well as reducing computational cost, enabling for potential real-time application under emergency situations.

Acknowledgement

If I were to tell myself five years ago that you have gone much further than you ever imagined, I probably would not have believed myself. What might not have begun as envisioned unfolded into a journey abundant with rewards.

This doctoral dissertation would not have been completed without the guidance and support of my supervisor, Professor Ryosuke Akoh, who has been offering insights that are both innovative and firmly grounded in practice starting my research student period. As a person not always strong in self-discipline, I benefited strongly from regular discussions either individually or through seminars that helped keep the research on course. The detailed suggestions he provided each time prevented issues to be overlooked and postponed throughout this journey marked by quite some amount of trial and error. He also provided me opportunities to participate in academic conferences, both domestically and internationally, which significantly broadened my horizons. Thanks to his guidance, I acquired not only a wide range of programming skills and knowledge, but also incisive and original ways of thinking.

I would like to express my gratitude to Professor Shiro Maeno for his mentorship. He generously took the time to participate in discussions and offer valuable suggestions despite demanding schedule. He carefully reviewed my manuscripts and daily writing materials for discussion, both in English and Japanese, and gave suggestions on word choices and highlighted noteworthy details. Although Prof. Maeno was not my primary supervisor, he offered great guidance and supervision during the one year when Prof. Akoh was abroad in Germany. Besides academic matters, Prof. Maeno was always ready to offer guidance from broader perspectives including advices for career choices and developments, which I deeply appreciate.

My appreciation also extends to Professor Keisuke Yoshida for the support. The discussion with him during the first year, when I did not fully develop my research stand, helped clarify my direction of work. I would also like to thank my colleagues in the lab. Mr. Hua helped me settle when I first arrived in Japan and tutored me with programming knowledge. Mr. Pan advised me on the choice of algorithms with his knowledge on machine learning. I appreciate Mr. Yamamoto, Mr. Takuno, and all other students for their support and assistance in research and in life.

I would like to thank the Rotary Yoneyama Memorial Foundation for the 2-year scholarship, which not only supported me financially, but also provided the opportunity to encounter people from different cultural and social backgrounds. The members of Club of Okayama Marunouchi, regardless of career, age and gender, offered warm welcome throughout the 2-year period, and invited me to all types of events where I was able to enjoy myself and share my thoughts.

I am deeply grateful to my parents for their constant support of every choice I have made. The last 10 years were full of changes in many ways and admittedly reckless decisions were made, a lot. I could not have achieved this far without their understanding.

Special thanks to my friends and all those who I am lucky to encounter, HT, LXH, AM and DI, for the emotional encouragement, endless patience in listening to my concerns, and always being there when I needed help.

Finally yet not the least, acknowledgement is owned to myself for never stop trying even during the moments when the journey seemed hopeless and no future were to come.

Not everything turns out as you wish; in truth most things do not. Yet all experiences, whether marked by bitterness or joy, are valuable and deserve to be embraced.

To the past that shaped who I am.

To the present that offers moments to share.

To the future that allows every possibilities to be explored.

List of Figures

List of Tables

CHAPTER 1 INTRODUCTION

1.1 Background

Over the past decades, flood disasters have become increasingly severe with recent occurrences across entire Japan, such as the 2015 Kanto and Tohoku heavy rainfall induced flood event, 2018 flood in west Japan, 2019 East Japan Typhoon event, and 2020 flood event in Kyushu. These severe disasters not only lead to fatalities, but also significant property loss with a maximum economic loss over 2157.4 billion yen in 2019[1)]. The growing severity of flood risks is closely related to the increasing frequency of precipitation with high intensities and intensification of discharge in rivers. Illustrated in **Fig. 1-1** is the yearly statistic for the number of occurrences of precipitation with intensity of 50 mm/hr and above over 1300 locations collected by the Automated Meteorological Data Acquisition System (AMeDAS) of Japan Meteorological Agency (JMA)[2)]. Based on the statistics, the long-term trend for 10-year average has increased from 226 times per year (1976-1985) to 333 times per year (2015-2024). These heavy precipitation events further induced the increasing intensification of discharge in rivers. **Fig. 1-2** shows the number of rivers having water levels exceeding the bank full threshold at least once over the years[3)], where target rivers for the statistics are the rivers with flood warning and water level notifications and the number in parenthesis indicates the total number of designated rivers. It can be observed that despite the implementation of structural countermeasures such as construction of dams, floodplains and river channel excavation, the number of rivers at risk of overflowing is gradually increasing for both national and prefecture managed rivers. By 2018, approximately 13% of rivers managed by nation and more than 25%

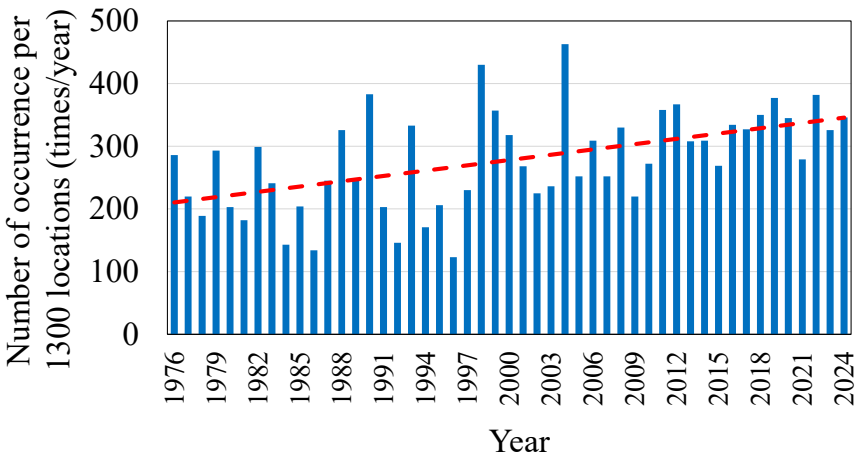


Fig. 1-1 Annual statistics for the number of occurrences of precipitation of 50 mm/hr and above per 1300 locations (national AMeDAS).

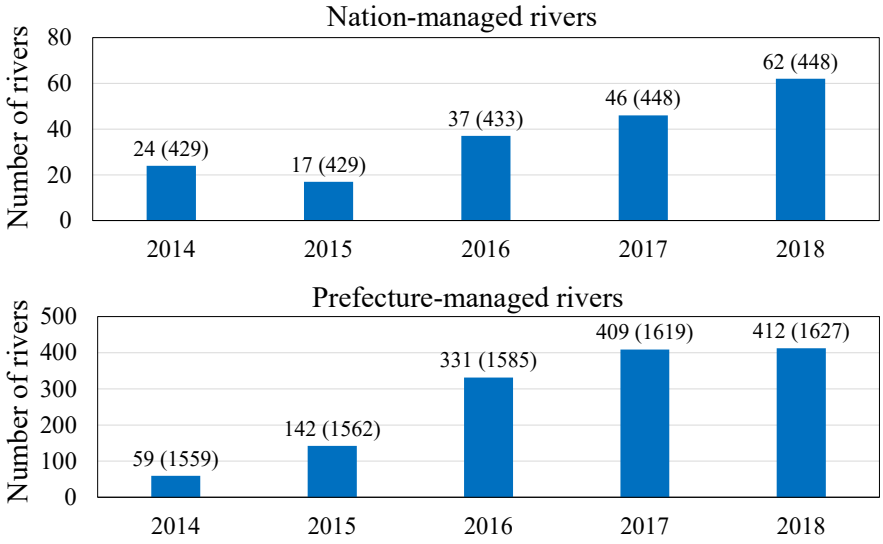


Fig. 1-2 Annual number of rivers exceeding bank full water level.

Table 1-1 Ratios for the intensification of precipitation, river discharge and frequency of flood occurrence due to climate change.

Scenario	Precipitation	River discharge	Frequency of flood occurrence
2-degree rise	1.1 times	1.2 times	2 times
4-degree rise	1.3 times	1.4 times	4 times

of rivers managed by prefecture experienced at least one high volume discharge.

Furthermore, this situation is expected to worsen because of global warning. According to the sixth assessment report of the Intergovernmental Panel on Climate Change (IPCC), the global average temperature will rise by 0.3 ℃ to 4.8 ℃ by the end of 21st century[4]. Two scenarios were carried out by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) to estimate the impact of such globally changing climate on Japan. As summarized in **Table 1-1**, the scenario of 2-degree rise in temperature induces 1.1 times increase in precipitation and 1.2 times increase in the scale of corresponding river discharge, and a doubled frequency for flood and the 4-degree rise scenario further intensifies flood occurrences by 4 times[5].

With such significant changes bringing challenges to the existing flood managements, discussions have been carried out regarding revising the management standards of current structural countermeasure, together with the active implementation of other non-structural flood control measures. For example, the fundamental policy of river management takes into account for the impact of climate change, and the determination of corresponding river peak

discharge of excessive floods incorporates not only current observed extreme conditions but also predictions of future precipitation conditions. The target for management also expands from disaster restraining to further including flood control measures to reduce total damage assuming the occurrence of excessive floods. One of such actions is the promotion of River Basin Disaster Resilience and Sustainability by All, which is an integrated approach consisting of both structural measure and non-structural aspects[6]. Shifting target area to the entire river basin where inundations could occur, this comprehensive concept involves all stakeholders including not only the government but also residents of the relating watershed. In this case, efficient and effective actions could be taken during all stages of a flood event starting from early-stage flood prevention to post-disaster stage actions such as damage reduction, early recovery and reconstruction. Another non-structural measure is the early preparation of information regarding potential disasters, and guidelines of corresponding actions for different authorities. For instance, to enhance recognitions of flood damage of the public, the configuration and distribution of flood hazard maps that cover the impact of inundation due to both pluvial and fluvial floods, high waves and tsunami assuming disasters at different excessive scales have been promoted[7]. Including information of expected inundation area, inundation depth, evacuation locations and suggested evacuation routes, hazard maps aim to advise residents of potential danger throughout daily life such that immediate actions could be taken when disasters occur. On the other hand, these inundation maps provide only static information of potential maximum damage, and the actual spatiotemporal propagation of floods would be complicated, especially under compound disasters involving landslides and earthquakes for example. Information sharing, preparedness and early actions during the occurrence of disasters are, therefore, essential for flood damage reduction.

One example of the dynamic aspect of disaster prevention action carried out in recent years by MLIT is the implementation of the timeline plan[8]. Specifically, this involves anticipating potential situations that may arise during a flood event, sharing the understanding among relevant disaster management agencies, and examining actions to be taken beforehand. On this basis, a timeline is established that clearly delineates the disaster response actions to be undertaken by these agencies, including the dissemination of disaster-related information prior to an event, the establishment of organizational structures, evacuation guidance, the operation or suspension of public transportation services, the securing of emergency evacuation routes, and other essential measures. Having strengthened collaboration, relating agencies implement disaster prevention actions during early stage of the occurrence of a disaster in an efficient and effective manner during actual disaster response based on the pre-established timeline system and in accordance with the actual condition and progression of situation. This suggests that even assuming the occurrence of disasters, the time of when to implement each

countermeasure, from early warning to the action-taking at all levels, is still an event-based decision depending on the specific characteristics of each flood event. Therefore, real-time information on disaster progression, that is real-time flood prediction, is essential to provide timely information not only for authorities who oversee policy-making decisions such as releasing flood warnings, but also for residents to take actions such as evacuation to alleviate damage. And such information is the key prerequisite to the success of the implementation of such countermeasure.

Two important criteria for flood forecasting system include high resolution of flood propagation prediction both spatially and temporally, and sufficient lead time between prediction and actual occurrence of a flood disaster, in other words prediction capability[9]. Hydrodynamic (HD) models have been the conventional approach used to characterize the dynamics of floods for their ability to produce simulation results with high resolution and high accuracy. Depending on the complexity of study area and research needs, flood analysis can be performed in 1D, 2D, or a coupling method of 1D and 2D models. 1D models are widely used in scenarios involving pipe flow, for example drainage sewers in urban areas, or confined river flow conditions at large scales since such models are easily configured and efficient in terms of computational cost. In comparison, 2D models are capable of representing detailed shape of river and microtopography within inundation areas, thus able to generate better simulation results, especially under situations where water level varies over river cross-sections[10),11)], or situations where inundation analysis over large areas of river basin is necessary. The configuration of such 2D models becomes complicated since representations on geometry of rivers and microtopography between different land uses are necessary. The choice of unstructured meshes is superior compared to regular grids for the flexibility in the representation of geometry of narrow channel and sharp bends. The resolution of unstructured meshes is also adjustable and models using such meshes have been found to generate higher accuracy and less computational time in complex domains where representation of irregular terrain features and microtopography are necessary[12)]. Coupling 1D and 2D models, the flux exchange between surface flow and underground drainage systems is also incorporated and methods of such have been widely used in urban flood analysis. Moreover, the resolution and quality of topographical and observable data necessary for the construction and calibration of HD models are also improving with developing technologies. For example, the recent release of digital elevation model (DEM) of 1-m resolution from the Geospatial Information Authority of Japan allows for finer and detailed representation of local elevation variations[13)]. Advancing tools of water level sensors together the facilitation of monitoring cameras and image recognition techniques are able to provide real-time conditions on river water levels. Investigations for precipitation forecasting also indicated promising results in recent years.

With these gradually refining databases, HD models are capable of producing detailed simulations in inundation area, water depth and velocity of flux, therefore satisfying the high accuracy and resolution criteria for flood disaster prevention.

On the other hand, the major limiting factor of the application of HD model in real-time situations is the high computation cost. Since the computational domain cover large areas of floodplain and inundation areas with high-resolution, HD models are still computationally expensive when unstructured meshes are implemented[14] even as computing techniques such as parallel processing are gradually advancing[15]. Therefore, the application of HD models over large scale is still impractical for real-time flood predictions since the substantial computing demands far exceed the time available for operational decision-making during severe flood events when immediate actions are necessary.

To overcome the constraints of such computational cost, the use of machine learning models in the field of hydraulics has been gaining popularity in recent years for efficiency in calculation. Machine learning (ML), as a data-driven approach, belongs to the field of artificial intelligence and the fundamental idea behind this technique is that computers analyze and identify the patterns between data on their own and make final decisions with as few human interventions as possible[16]. Analyzing purely on the training data, ML models numerically formulate the nonlinear relationships between factors without incorporating the underlying physical equations, providing low computational cost. On the other hand, this also suggests that a large amount of high-quality training data is required for the construction of ML models. Regarding the application of ML models on flood predictions over large areas, the data scarcity issue becomes even more complicated because unlike predicting discharge or water levels of rivers, flood propagation changes both spatially and temporally. Observations of historical flood events are limited, and the ones available are mostly qualitative records on the inundation area, with some additional information on maximum depth at most. Even with the assistance of live cameras and advanced monitoring equipment in recent years, details of large areas throughout an entire flood event are still limited. Another limitation of ML models is the generalization problem, which is the ability to be applied to targets of untrained characteristics[17]. The accuracy of ML models may be significantly compromised when exposed to untrained pattern of flood events, for example those extreme floods induced by climate change, if the training data is limited to historical inundation observations due to the inherited property of ML models.

As a result, in response to the increasing frequency of extreme rainfall and flood events, there is still an urgent need for the development of an approach for an efficient and accurate spatiotemporal flood prediction over large areas to reduce economic losses and casualties during emergent situations. To be more specific, the approach should target for real-time flood

predictions of inland area by providing timely and accurate details applicable for decision-making during rapidly developing flood events such as the release of early warning and evacuation route determination.

1.2 Literature review

The use of ML in the discipline of hydrology and hydraulics has been notable in recent years, and the state-of-art studies of such application varies from the field of identifying flood-prone areas and forecasting flood-related factors. For example, the configuration of susceptibility maps of an area has been frequently investigated by multiple studies. Different from flood hazard maps that are developed with respect to a specific scale of rainfall and/or flood and indicate specific severity of flood, susceptibility maps provide information on which area is vulnerable to floods. Therefore, the flood conditioning factors in susceptibility maps investigated in previous research generally involve the inherited topo-geographical properties of the area, such as elevation, slope, aspect, land use, curvature, distance from river, topographical wetness index, sediment transportation index, and soil type. In the study carried out by Khosravi et al.[18], maps of 12 conditional factors over Jiangxi, China was used as the features for the construction of two ML models, Naive Bayes tree and Naive Bayers classifier. When evaluated against 50 locations, 30% of the available data collected from flood inventory maps, both models resulted in high accuracy, and the visualization of flood susceptibility maps was provided. A comparison between different ML models, namely artificial neural networks (ANN), support vector machine (SVM), random forest (RF), random subspace and dagging, using 167 flood points and similar flood influencing factors was performed by Islam et al. in Bangladesh, and the accuracy of models varied significantly between models, both spatially and quantitatively[19]. Another research with a relatively large unbiased datasets (4333 locations) from Berlin, Germany was investigated to generate flood susceptibility maps at different resolutions[20]. Further taking into consideration the impact of roads and stormwater drainage systems, this research targeted urban area with dense population, and the flood susceptibility map generated from the model with best performance coincided highly with the flooded locations collected from inventory maps.

While understanding the intrinsic flooding potential for each area is important, more studies focused on the investigation of dynamic characteristics of event-based floods, in other words, the forecasting aspect of hydraulic properties. Such literatures are further divided into two types, which are the kind of studies using only dynamic observable datasets and those incorporating synthetic data from other sources such as numerical simulation. Since the

configuration of ML models requires a large number of training data, the former type of research typically focuses at one specific location, and the commonly investigated target of prediction is the water level and/or discharge of river. For example, stream flow data of Hunza River basin of Pakistan over a 42-year period (1966-2008) was summarized for the configuration of multilayer perceptron (MLP), RF and Support Vector Regression (SVR)[21]. Using 80% of the available data as the training data and the rest of the data as validation, all three models successfully captured the general shape of how streamflow varies between seasons, but the accuracy on the peak flow at each season was slightly compromised in the testing phase. Similarly, Nguyen and Huu[22] collected continuous daily water level data of the Mekong River from 1994 to 2000 as the training data (1071 samples) to configurate different ML model, namely LASSO, RF and SVR to predict the time-series water level in the year 2001 to 2003 (459 samples). The model with the best performance produced mean absolute error of less than 0.5 m with a lead time of 5 days, which is considered acceptable for a flood predicting model based on the requirement of the Mekong River Commission. Another research with more observed data was carried out by Yaseen et al., in which they developed a neural network model to predict hourly river flow in Australia, using the hourly dataset over a 5-year period (52555 observation partitioned 60%, 20% and 20% for training, validation and testing respectively), and resulted in models with coefficient of determination over 0.9 on test dataset[23].

On the other hand, these studies were carried out under the condition where continuous observations over a long time (over years to decades for example) are available. Yet the reality is that observations are scarce, especially those related to compound hazards and inundation area and/or volume over large areas, which is also the constraining factor for the configuration of a real-time spatiotemporal explicit flood prediction model. Such situations become the motivation for the investigation of integration of HD models with ML models to compensate for data limitation, and the general process consists of the following procedure. HD models are first constructed and validated over the study area. Numerical flood analysis is then carried out either observed historical rainfall or with hypothetical design rainfalls and the simulation results are summarized to be the training data for ML models. Once being trained, ML models are used to perform predictions, and comparisons between predicted values and observations or simulation results from HD models are carried out for the evaluation of prediction accuracy.

For example, targeting for the water level prediction induced by compound flood in a data-scarce delta located in Indonesia, Sampurao et al. first constructed a HD model through SLIM 2D with unstructured triangular meshes[24]. Validated against observations for January 2019 with a high consistency, the HD model was then used for simulations under different atmospheric, tidal river forcing and precipitation conditions. The authors further configured

RF, MLR and SVM models using simulation results from HD models as the training data and evaluated the accuracy of ML models against three actual historical events, where RF model was concluded to have the best performance of all. This augmentation approach of HD model and ML algorithms is especially necessary for the prediction over large areas since observations of such, for example flood inundation over a river basin, generally require large amount of event-based data and labor for field measures. Therefore, previous literatures have investigated different aspects of this approach with objectives varying from prediction water depth at multiple locations to the maximum depth of a river basin. Aiming for the water depth prediction of a downtown area in Tianjin, China, Yan et al.[25] constructed an urban flood simulation model and applied 35 design rainfall scenarios (5 design rainfall patterns and 7 intensities) for the configuration of neural network model. 7 high-risk locations were selected for accuracy evaluation of ML model, with one location having observed data and the rest validated against simulation results, and the average relative error when tested against the rainfall event in May 2019 was approximately 19%. The authors also compared the computational efficiency between HD models and ML models, and for the same calculation period performed using the same computer, HD models required 15 minutes while ML only took 20 seconds. Targeting detailed flood inundation mapping, Bermudez et al.[26] also performed predictions on maximum water depth and velocity using a surrogated model of HD model with least-squares support vector machines. The method was first validated through numerical experiment of idealized topography and then applied to real study area located in the northwest of Spain with the use of unstructured triangular meshes. Trained with simulation results from HD models, the ML model was able to predict the maximum water depth with precisions of approximately 0.9, and 0.7 for velocity prediction when evaluated against the four control points. Similarly, Jang et al.[27] used simulation results of HD models using 99 observed rainfall events with 10-minutes interval as the training data. Two RF models using different features were constructed and predictions on maximum depth were performed, and both models were concluded to have high accuracy of R^2 greater than 0.95, and the model including statistical characteristics of rainfall resulted in better performance.

Besides the investigations on maximum depth, the temporal propagation of floods has also been studied. One attempt of such was carried out by Burrichter et al.[28], who targeted the prediction of pluvial floods in urban areas with 2 m high resolution grids. Using a 1D/2D coupled HD model, the authors investigated a total of 258 rainfall events, both hypothetical and actual, to construct the training data. After data preprocessing simulation results, multiple models including artificial neural networks (ANN), temporal graph convolutional network (T-GCN) and conditional generative adversarial network (cGAN) were configured using different combinations of inputs. The accuracy of models was then evaluated with different water depth

thresholds and model with highest accuracy, T-GCN, was tested over historical rainfall event, where high performance was observed for a lead time of 15 minutes. For predictions with lead time of 30 and 60 minutes, the range of errors expanded to a maximum of 25 cm.

Despite the above-mentioned attempts, the constraints regarding two major factors for early warning, high resolution and sufficient lead time between prediction and occurrence of floods, have still not been fully resolved. Specifically, the flood predictions carried out by majority of studies were grid based. As mentioned in earlier section, the representation of microtopography in terms of shape and delineation between land use, even with 2-m resolution, is limited and the accuracy of corresponding HD simulation results is constrained. Since the accuracy of ML models depends highly on the quality of training data[23], the accuracy of surrogated models, therefore, could only be as best as HD models regardless of the choice of ML algorithms and fine-tuning processes. Such spatial resolution of prediction is not sufficient enough to be applied for potentially practical purposes such as decision-making regarding evacuations route calculations. Furthermore, the target for prediction focused mostly on maximum depth and the recognition of inundation area, and the studies investigating temporal aspect of flood propagation evaluated only a few choices of prediction periods within 1 hr. Such studies fail to reproduce detailed spatiotemporal information of flood propagation and only provide limited lead time for emergent responses before the arrival of floods.

In summary, an integrated method of both spatial and temporal variations of prediction accuracy over a large river basin during an entire rainfall/flood event is still necessary, especially considering the increasing need for accurate real-time flood forecast techniques as a prerequisite for the implementation of future countermeasures to reduce damage under emergency situations.

1.3 Objectives

The objective of this study is to develop an iterative HD-simulation informed ML model for efficient spatiotemporal flood prediction induced by practical rainfall events. The proposed augmentation approach attempts to resolve the above-mentioned research gaps on data scarcity of high-quality training data by taking advantage of highly accurate simulation results of conventional HD model and fulfill the lead time requirements through the usage of efficient ML models with an iterative prediction process.

Using high-resolution unstructured triangular meshes, the delineation of land use boundaries and representation of microtopography are included in the comprehensive flood analysis model. With careful calibrations against historical flood events, this HD model generates

highly accurate simulation results, which are further processed and summarized as the training data for the configuration of ML model, the Random Forest (RF) model. Incorporating different hypothetical design rainfall patterns with various shapes, intensity and peak distribution, the data scarcity issue regarding both limited historical flood observations over large areas and unprecedented excessive floods is resolved. Furthermore, this research attempts to enhance the interpretability of RF model by selecting features that represent hydraulic properties in the same way as HD model both spatially and temporally. Specifically, the features include both static inherited properties of each mesh such as coordinates and elevation, and dynamic properties of input rainfall conditions, land use, prior state information of adjacent meshes, and boundary conditions to reflect exchange of flux at computation domains. In such way, the configuration of RF model is carried out with simulation results from HD model in the training process. In the prediction phase after model construction, this research proposes an iterative prediction process to achieve long-term predictions with efficiency. A non-inundated condition over inland area (for example a time before precipitation occurs) is applied as the initial input for the first prediction of the RF model. The predictions at each step are then reconstructed as the prior state input for the next time step and process loops throughout the entire rainfall period.

First evaluated through numerical experiments, the approach was applied to the Karube river basin. With detailed investigations regarding feature selection and the design of hypothetical design rainfalls, the potential of this augmentation approach to obtain high-accuracy results is evaluated against actual historical rainfall events. In addition to conventional ML evaluation metrics, the assessment of the RF model also examined its consistency with HD simulation results in providing insights to flood propagation behaviors. Time consumption for calculation with HD model and prediction using RF model was also compared to evaluate the efficiency aspect required in the early warning system under during urgent situations.

1.4 Structure overview

The structure of this thesis is organized as follows.

Chapter 1 states the need for a real-time flood prediction method, summarizes the state-of-the-art machine learning methodology and surrogated models in flood predictions and stresses the objectives of this research.

Chapter 2 provides a flowchart of the augmentation approach, and details on the hydrodynamic model and machine learning algorithm used in this research.

Chapter 3 summarizes evaluation metrics, both qualitatively on the performance and

quantitatively on the accuracy of prediction and carries out a methodology validation of the proposed augmentation approach using numerical experiments under rainfall conditions of constant intensity. Investigations on features, hyperparameter tuning and prediction intervals are also detailed in this section. The accuracy of predictions is evaluated with different metrics to obtain the optimal parameter setting of the model.

Chapter 4 evaluates the potential of the augmentation approach to be applied to a complicated practical terrain settings, the Karube river basin, under constant rainfall conditions. With an overview of the study area and detailed settings of both the HD model and ML model, the accuracy of the predictions is evaluated and discussions on the limitations regarding practical applications under actual rainfall conditions with temporal variations of intensities are summarized.

Chapter 5 carries out investigations regarding further fine-tuning of features in model configuration and then evaluates how the selection of hypothetical rainfall patterns in the training data influences prediction performance. Targeting for the flood prediction induced by practical rainfall events, the simplified hypothetical design rainfalls include considerations on the shape of rainfall, peak intensity and peak arrival time. Refinements of features and cross-validation methods are discussed to validate the interpretation of features selections as well as maintaining prediction accuracy. The generalization capabilities of the models constructed with different combinations of training data are then examined by testing the models against multiple hypothetical rainfall events.

Chapter 6 assesses the capability of the models, trained with simplified hypothetical rainfall patterns, in reproducing the spatial and temporal inundation propagation under practical rainfall conditions. Two historical rainfall events that induced floods over the study area are selected as the target events and detailed evaluations regarding the overall predictive performances and reproducibility of hydraulic behaviors are carried out.

Chapter 7 summarizes the main findings and future perspectives of this study.

References

1) Ministry of Land, Infrastructure, Transport and Tourism, Water Management and Land Conservation Bureau, River Planning Division, River Planning Coordination Office (2025). *Flood statistics survey, trends for the amount of flood damage since the Meiji Era*.
2) Japan Meteorological Agency: Changes in extreme events such as heavy rainfall and excessive heat. https://www.data.jma.go.jp/cpdinfo/extreme/extreme_p.html (Accessed August 1, 2025).
3) Ministry of Land, Infrastructure, Transport and Tourism (2019). *Regarding the effects of global warming*.
4) Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Luca, A. D., ... & Allan, R. (2023). Weather and climate extreme events in a changing climate. The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766, doi:10.1017/9781009157896.013.
5) Ministry of Land, Infrastructure, Transport and Tourism (2019). *Proposal for flood control measures in correspondence to climate change.*
6) Ministry of Land, Infrastructure, Transport and Tourism: *Fundamental concept for River Basin Disaster Resilience and Sustainability by All.*
7) Ministry of Land, Infrastructure, Transport and Tourism, Water Management and Land Conservation Bureau, River Environment Division Flood Prevention Planning Office (2023). *Guidelines for the preparation of flood hazard maps.*
8) Ministry of Land, Infrastructure, Transport and Tourism, Water-Related Disaster Prevention and Mitigation Headquarters, Disaster Prevention Action Planning Group (2016). *Guidelines for the formulation and utilization of timeline (Disaster prevention action plans), First edition.*
9) Zhao, G., Balstrøm, T., Mark, O., & Jensen, M. B. (2021). Multi-scale target-specified sub-model approach for fast large-scale high-resolution 2D urban flood modelling. *Water*, 13(3), 259.
10) Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., & Kim, S. J. E. M. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental modelling & software*, 90, 201-216.
11) Mudashiru, R. B., Sabtu, N., Abustan, I., & Balogun, W. (2021). Flood hazard mapping methods: A review. *Journal of hydrology*, 603, 126846.
12) Kim, B., Sanders, B. F., Schubert, J. E., & Famiglietti, J. S. (2014). Mesh type tradeoffs in 2D hydrodynamic modeling of flooding with a Godunov-based flow solver. *Advances in Water Resources*, 68, 42-61.
13) Geospatial Information Authority of Japan: *Basic map information download service*.

14) Hoch, J. M., van Beek, R., Winsemius, H. C., & Bierkens, M. F. (2018). Benchmarking flexible meshes and regular grids for large-scale fluvial inundation modelling. *Advances in water resources*, 121, 350-360.
15) Karim, F., Armin, M. A., Ahmedt-Aristizabal, D., Tychsen-Smith, L., & Petersson, L. (2023). A review of hydrodynamic and machine learning approaches for flood inundation modeling. *Water*, 15 (3): 566.
16) Mitchell, T. (1997). *Machine Learning*. McGraw Hill.
17) Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536.
18) Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., ... & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology*, 573, 311-323.
19) Islam, A. R. M. T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., ... & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3), 101075.
20) Seleem, O., Ayzel, G., de Souza, A. C. T., Bronstert, A., & Heistermann, M. (2022). Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. *Geomatics, Natural Hazards and Risk*, 13(1), 1640-1662.
21) Hussain, D., & Khan, A. A. (2020). Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth Science Informatics*, 13(3), 939-949.
22) Nguyen, T. T., Huu, Q. N., & Li, M. J. (October 2015). Forecasting time series water levels on Mekong river using machine learning models. *In 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)* (pp. 292-297). IEEE.
23) Yaseen, Z. M., Naganna, S. R., Sa'adi, Z., Samui, P., Ghorbani, M. A., Salih, S. Q., & Shahid, S. (2020). Hourly river flow forecasting: application of emotional neural network versus multiple machine learning paradigms. *Water Resources Management*, 34, 1075-1091.
24) Sampurno, J., Vallaeys, V., Ardianto, R., & Hanert, E. (2022). Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta. *Nonlinear Processes in Geophysics Discussions*, 2022, 1-20.
25) Yan, X., Xu, K., Feng, W., & Chen, J. (2021). A rapid prediction model of urban flood inundation in a high-risk area coupling machine learning and numerical simulation approaches. *International Journal of Disaster Risk Science*, 12(6), 903-918.
26) Bermúdez, M., Cea, L., & Puertas, J. (2019). A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. *Journal of Flood Risk Management*, 12, e12522.
27) Jang, S. D., Yoo, J. H., Lee, Y. S., & Kim, B. (2025). Flood prediction in urban areas based on

machine learning considering the statistical characteristics of rainfall. *Progress in Disaster Science*, 26, 100415.

28) Burrichter, B., Hofmann, J., Koltermann da Silva, J., Niemann, A., & Quirmbach, M. (2023). A spatiotemporal deep learning approach for urban pluvial flood forecasting with multi-source data. *Water*, 15(9), 1760.

CHAPTER 2 METHOD AND MODELS

2.1 Overview of iterative hydrodynamic-simulation-informed random forest method

The overall schematic of the augmentation method of the hydrodynamic (HD) model and Random Forest (RF) model proposed in this study is depicted in **Fig. 2-1**. Targeting for the prediction of detailed spatiotemporal inundation depths at each location, this approach consists of three main procedures: a) data generation through flood analysis using HD model, b) training phase for RF model using simulation results as the training data, and c) prediction for untrained rainfall pattens through iterative prediction process. Detailed workflow for each stage of procedure is illustrated in **Fig. 2-2**.

First, flood analysis with HD model is conducted under different rainfall conditions to generate results on inundation area and water depth. Rainfall intensities are applied spatially evenly over the target area for all purposes of this study. As for temporal variation, hypothetically designed rainfalls of both constant rainfall and variations with different shapes, distribution and peak intensities are included in HD calculation. The simulation results for each mesh over time are summarized to be the training database for RF model. The second procedure selects different subsets of the above database for RF model configuration. Feature selection and hyperparameter tuning are performed with numerical experiment to validate for this augmentation approach as well as optimizing model performance. The third procedure then applies the configurated RF model over untrained rainfall events through iterative

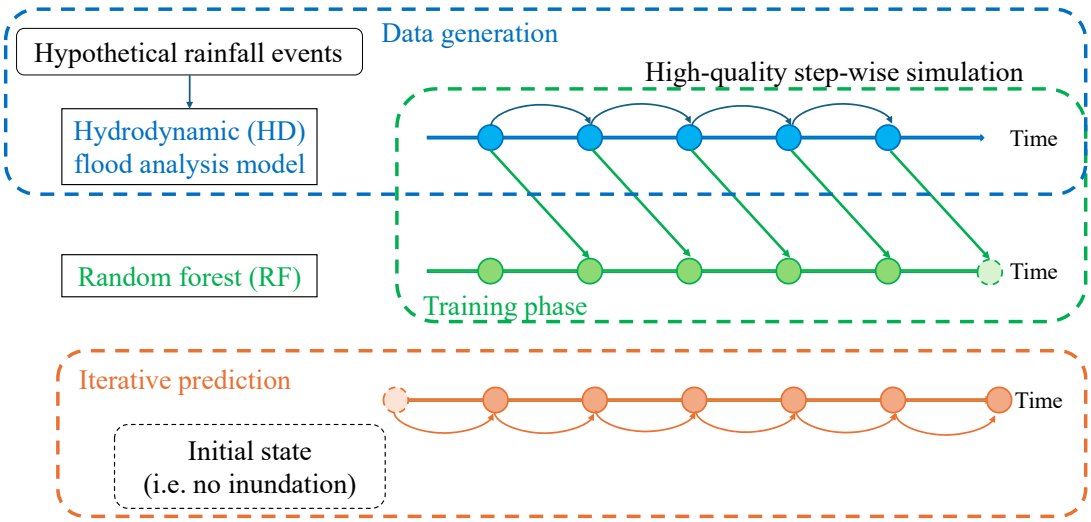


Fig. 2-1 Overall logistic of augmentation method.

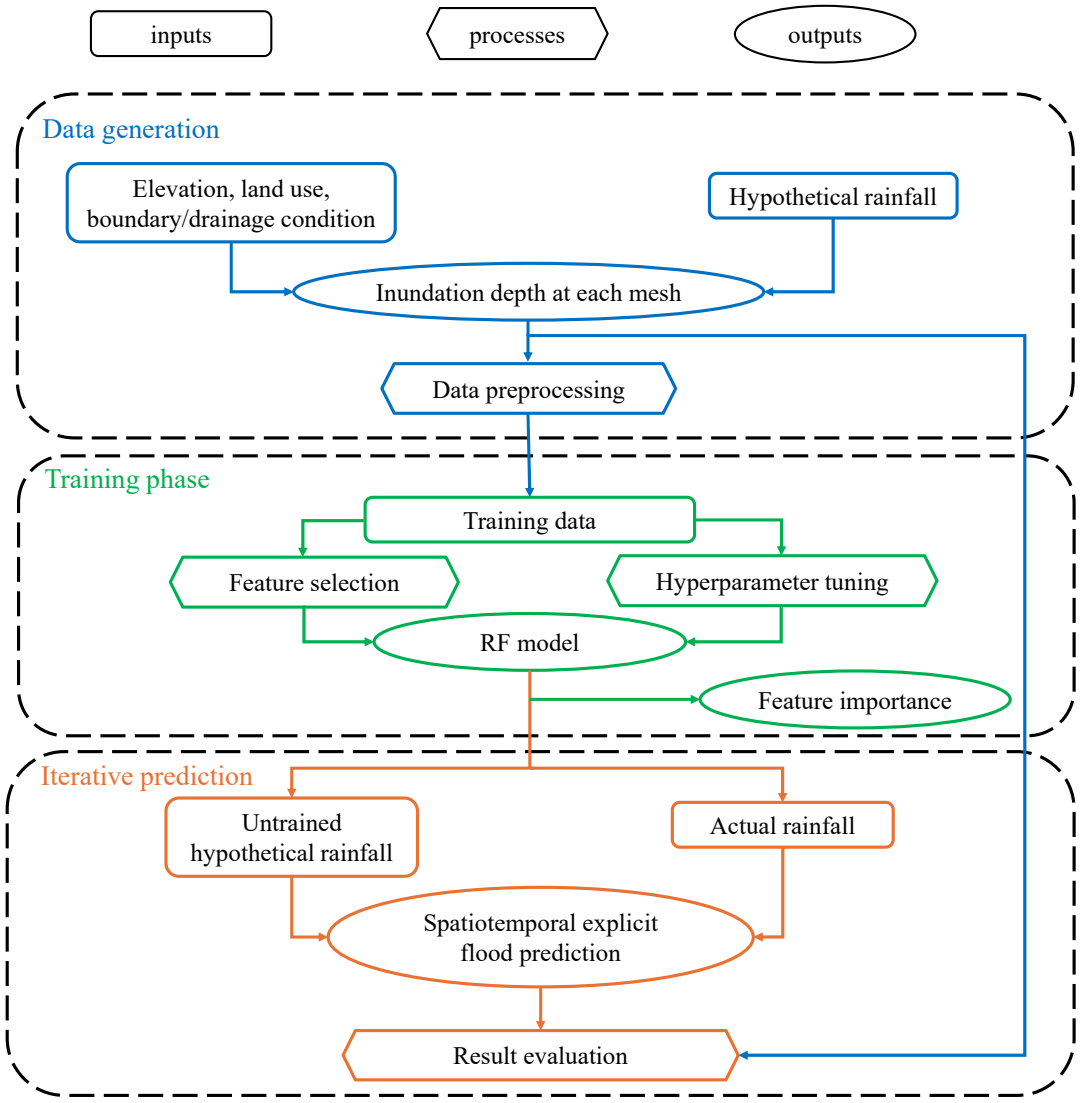


Fig. 2-2 Detailed workflow for each stage of procedure.

prediction process and evaluations of predictive performance are carried out against comparisons to simulation results from HD models.

Investigations are first conducted under constant rainfall conditions with simple synthetic terrain conditions for method validation and then performed over practical topographical settings. With the final goal being the spatiotemporal explicit flood prediction induced by actual rainfall events, evaluations regarding the selection of hypothetical rainfall patterns as the training data are also carried out in detail. The evaluation of the RF mode's predictive capability is accessed with both conventional machine learning metrics and comparisons to simulation results to verify whether the model preserves the physical consistency and spatiotemporal dynamics represented in HD models.

2.2 Model description

2.2.1 Comprehensive flood analysis model

This research adopted the comprehensive flood analysis model previously developed by Maeno et al.[1)] for flood simulation over river basins. With the use of triangular unstructured meshes over the entire target area, microtopography of not only river and roads but also drainage channels and buildings can be represented in detail without compromising the original shape. The resolution for unstructured meshes are also adjustable with flexibility compared to structured grids. On the other hand, the complex representation of geometry requires more memory to store the connectivity between meshes, which leads to higher computational cost. As a result, this flood analysis model takes advantage of general-purpose computing on graphics processing units (GPGPU) that allows parallel processing technique to increase calculation efficiency[2)].

Governed by two-dimensional shallow water equation, this model is capable of handling river flow, overflow from embankments, and flood inundation over inland areas at the same time. Detailed equations are summarized below.

$$\frac{\partial \mathbf{U}}{\partial t}+\frac{\partial \mathbf{F}}{\partial x}+\frac{\partial \mathbf{G}}{\partial y}=\mathbf{S_b}+\mathbf{S_f}+\mathbf{Q} \tag{3.1}$$

where t is time, (x, y) are coordinates in Cartesian system, $\mathbf{U}$ is conservation vector, $\mathbf{F}, \mathbf{G}$ are flux vector in the x, y direction respectively, $\mathbf{S_b}$ is the momentum source term due to gradient of bottom surface, $\mathbf{S_f}$ is momentum source term due to friction of bottom surface and $\mathbf{Q}$ is the source term for rainfall. Source terms are represented by the following equations.

$$\mathbf{U}=\begin{bmatrix} h \\ hu \\ hv \end{bmatrix}, \mathbf{F}=\begin{bmatrix} hu \\ hu^2+gh^2/2 \\ huv \end{bmatrix}, \mathbf{G}=\begin{bmatrix} hv \\ huv \\ hv^2+gh^2/2 \end{bmatrix} \tag{3.2}$$

$$\mathbf{S_b}=\begin{bmatrix} 0 \\ -gh\dfrac{\partial z_b}{\partial x} \\ -gh\dfrac{\partial z_b}{\partial y} \end{bmatrix}, \mathbf{S_f}=\begin{bmatrix} 0 \\ \dfrac{-gn^2u\sqrt{u^2+v^2}}{h^{1/3}} \\ \dfrac{-gn^2v\sqrt{u^2+v^2}}{h^{1/3}} \end{bmatrix} \tag{3.3}$$

$$\mathbf{Q} = \begin{bmatrix} Q + r_a \\ 0 \\ 0 \end{bmatrix} \tag{3.4}$$

where h (m) is water depth, (u, v) (m/s) is the average velocity of water depth in the (x, y) direction respectively, z_b (m) is elevation, g (m/s^2) is gravitational acceleration, n (s/m$^{1/3}$) is Manning's roughness coefficient, and r_a (mm) is effective rainfall. The effective rainfall is calculated through $r_a = r \cdot Fr$ where r (mm) is observed rainfall and Fr is runoff coefficient. Runoff coefficient, a representation of land use in this calculation, is calculated through the following equation.

$$Fr = \begin{cases} f_{usa} (0 \le r_{sum} < R_{sa}) \\ f_{sa} (R_{sa} \le r_{sum}) \end{cases} \tag{3.5}$$

where R_{sum} (mm) is the cumulative rainfall at each mesh, R_{sa} (mm) is saturated rainfall, f_{usa} and f_{sa} are the runoff coefficient under unsaturation and saturation respectively. Parameters settings with respect to each land use will be detailed in the following chapter. Depending on the saturation rainfall threshold and cumulative rainfall of the actual rainfall event, the runoff coefficients switch from unsaturation to saturation, thus acting as a simplified representation for infiltration of different land uses. The boundary lines between different land uses where overflow would occur, for example river levees and ridges of paddies, are defined by carefully overlapping the edge of triangular meshes to the delineated shaped polygons of each land use. And exchange of flow is calculated through Honma's overflow equation shown below.

$$Q = \begin{cases} 0.35 h_1 B \sqrt{2 g h_1} & if \quad h_2 / h_1 < 2/3 \\ 0.91 h_2 B \sqrt{2 g (h_1 - h_2)} & otherwise \end{cases} \tag{3.6}$$

where h_1 (m) and h_2 (m) refer to the water depths above the crest of the levee and/or ridge of the two meshes respectively, and B (m) is the length of the mesh. The numerical analysis employs the Godunov-type finite volume method, the conservative numerical scheme for solving partial differential equations that are effective in hyperbolic equations, in this case the shallow water equation[3].

This comprehensive inundation analysis model has been validated against historical rainfall and flood events over multiple areas with various applications. For example, flood analysis over Mabi District of Kurashiki city that incorporated detailed embankment breach during the 2018 rainfall event was performed and validated against observations of flood traces from field investigations[1]. Further investigations with an extended simulation period over the drainage period after rainfall ceased were carried out to study the behavior of floating litter and the flood analysis coincided with the discharges and water levels of observation stations[4]. Furthermore,

construction of the HD model over the Karube region, the target area of this research, had also been performed previously by Akoh et al.[5] with evaluations against 2018 rainfall events. With refinement on micro-topographical representation of drainage channels, the model was then applied to investigate the effect of paddy field dams on flood mitigation[6]. With the above-mentioned applications, it is concluded that this comprehensive flood analysis model is capable of generating reliable simulation results.

2.2.2 Random forest model

Random forest (RF) is a supervised learning algorithm for both classification and regression with the use of decision trees[7]. The training datasets are sampled randomly with replacements (bootstrap sampling), which are then partitioned binarily in each tree (**Fig. 2-3**). For regression purposes, the predictions from each tree are ensembled to generate the ultimate output[8]. By averaging results from uncorrelated decision trees, RF model is capable of suppressing prediction error caused by outliers, thus reducing the risk of overfitting compared to using

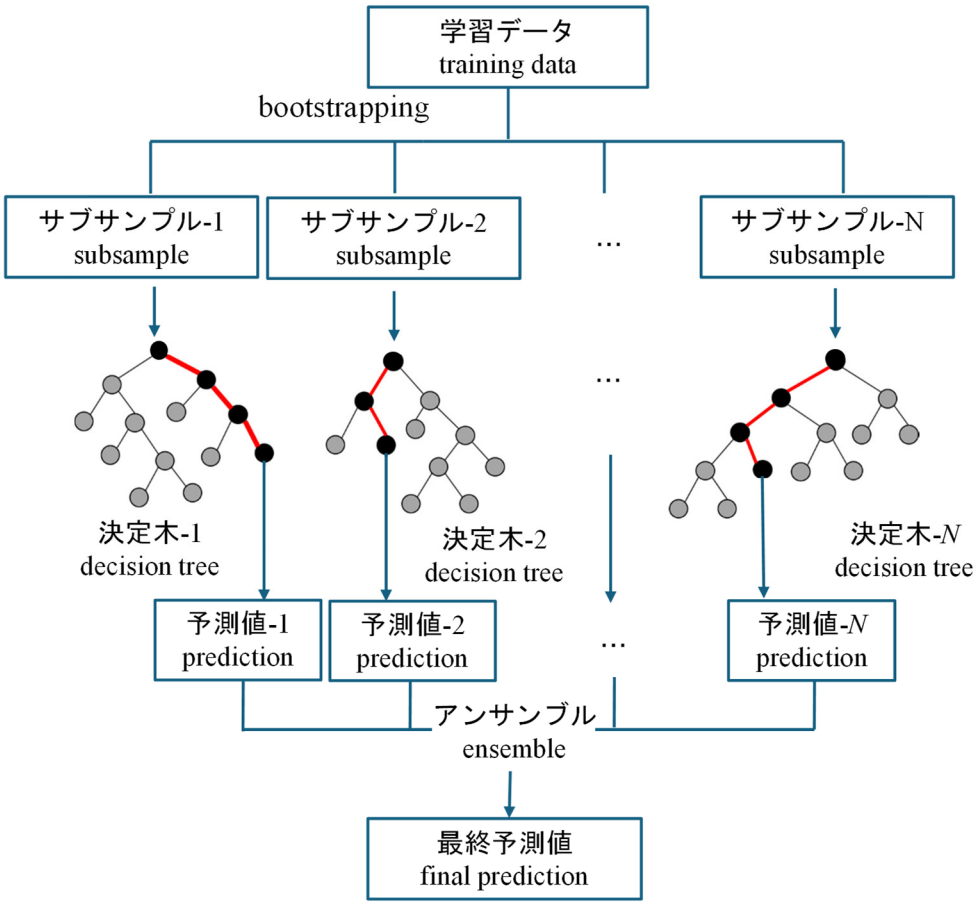


Fig. 2-3 Schematic of RF model.

single decision trees. On the other hand, the interpretation of each tree becomes challenging, and the risk of overfitting cannot be completely eliminated when applied on small numbers of training data. The accuracy of RF model is optimized through hyperparameters, the parameters that need to be adjusted in the training process of the algorithm. The most important hyperparameters in RF model are number of trees in the forest (*n_estimator*), maximum depth of tree (*max_depth*), and number of features at each split (*mtry*). Increasing the number of trees in the forest typically results in higher performance, but the corresponding computational cost would also increase. Similarly, deeper trees can capture more information between data, but would also increase the risk for overfitting, the phenomenon where the model captures too many characteristics of the training data including noise and outliers such that the model's performance is significantly compromised when predicting on unseen data.

The construction of RF model is performed with the *RandomForestRegressor* from the machine learning library scikit-learn (ver. 1.3.2) published by Python[9]. This algorithm was selected for its simplicity of use, efficiency of training, and ability to handle large datasets[10], considering that the training data obtained from HD models consists mesh-based temporal inundation information covering large areas of a river basin. On the other hand, this model is considered slightly subjective to extrapolations. As a result, how the RF models perform when the target event for prediction approaches or exceeds the range covered in the training data was also investigated and discussed regarding the importance of careful selections of training data in order to increase the generalization capability of the RF model for practical application purposes. The default values of the above-mentioned hyperparameters for this version of RF model are 100, None, and 1.0 respectively. In order to balance model accuracy and computational cost, hyperparameter tuning processes are performed and described in the next chapters.

2.3 Model interpretability and feature selection

The definition of the interpretability of an ML model is the ability to provide the meaning or explanations that humans could understand[11]. Especially for ML models that are intended for practical applications related to decision-making, for example as a simplified representation of a system, it is commonly suggested that the model should be somehow interpretable such that the application of the model is reassuring. The construction of the RF model in this research, therefore, attempts to retain the interpretability of the model by transferring the logistics of HD model into the features.

As the objective of this research is an efficient spatiotemporal explicit flood inundation

prediction, the target variable is selected to be the inundation depth at each mesh. Features used in the construction of RF models, while not directly incorporated as equations, are designed to reflect the relationships between and predicting features and the target variable[12]. As a result, it is reasonable to include features representing actual hydrodynamics characteristics both spatially and temporally. On the other hand, RF model does not inherently account for time series nor spatial dependencies (i.e. assuming independently and identically distributed data points in the bootstrap process). As a result, features selection becomes important to adapt the RF model for spatiotemporal explicit flood predictions over time. The predictive features investigated in this study are divided into static features and dynamic features. The static features included x and y coordinates and elevation at each mesh. The representation of land use is first carried out using binary indicators (0 and 1) since it is an inherited property that does not change over time. Yet this property is incorporated in the HD model as runoff coefficient, the value of which differs for unsaturated and saturated conditions as described in the previous section. Therefore, the feature representation of land use is further refined in the RF model as a dynamic feature.

Transferring the logistics of HD model in terms of considering spatiotemporal exchange of flux between meshes, the dynamic features proposed in this study takes into consideration spatial factors as information of adjacent meshes, and temporal factors of prior state (t-1). By including the inundation information from past time steps of each target mesh and adjacent meshes, the hydraulic information of previous states is represented. As for locations at calculation boundaries where connected meshes are partially missing, the information of the mesh itself is used. Other properties reflecting the incoming/outcoming volume of water from computational domains are selected to be precipitation (current and prior), which is assumed to be available from rainfall forecast, and the inflow and outflow boundary conditions from rivers. Details of the feature selections and corresponding notations will be introduced in the next chapters.

Feature importance, an evaluation on how each of the features affect the accuracy of model, is another essential factor for model interpretation. The built-in attribute (*feature_importances*) from the scikit-learn library of RF model evaluates the Gini importance metric, which is a relative measure on how significantly the mean accuracy decreases when each of the feature is omitted in model configuration. In other words, feature importance reflects how strong the RF model uses each feature to make predictions. This research takes advantage of this metric as supporting evidence to verify whether the RF models capture relationships that are consistent with the underlying dynamics of HD models.

2.4 Iterative prediction process

The prediction process of the RF model requires the input of all features used in the training process. While rainfall information and river state can be obtained from precipitation forecast and water level observatories and real-time measurements from sensors to some extent, the prior state of inundation information over large areas is in general not available during an actual rainfall event. To compensate for this data scarcity for long-time predictions, this study proposed an iterative prediction method. An initial state of no inundation over the region (i.e. right before precipitation occurs), that is $depth_{t=0} = 0$ m for inland inundation area, is assumed for all predictions in this research. The initial condition for rivers is assumed to be normal water level conditions. Using this prior state and observable features regarding rainfall and river state as input, the first prediction could be carried out. The predicted outputs are then reconstructed into the input data for the next prediction based on mesh connectivity, and the iteration progresses until the end of the rainfall event. The iteration interval for prior state is also investigated in the following sessions to balance computational cost and accuracy.

References

1) Maeno, S., Akoh, R., Nihei, Y., Akamatsu, Y., & Yoshida, K. (2019). Flood disaster characteristics in Mabi district of Kurashiki city due to heavy rain in 2018 and investigation of flooding scenario using river flood flow combined inundation flow model. *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, 75(1), 387-402.
2) Zeng, X., Akoh, R., Ishikawa, T., & Nakamura, T. (2013). On the Efficiency of Gpgpu Acceleration of Tsunami Simulations Using AN Unstructured Triangular Mesh System. *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, 69(4), I_619-I_624.
3) Godounov, S. (1959). A difference method for numerical calculation of discontinuous solutions of the equation of hydrodynamics. *Matematicheskii Sbornik*, 47(89-3), 271-306.
4) Hua, W., Akoh, R., Liu, K., & Maeno, S. (2024). Evaluation of behavior of floating litter during flooding with levee breach by using inundation analysis. *Journal of JSCE*, 12(2), 23-16012.
5) Akoh, R., Ikejiri, Y., & Maeno, S. (2022). Investigation of Flood Control Measures Using Rice Fields in the Karube River Basin of the Takahashi River System. *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, 78(2), I_253-I_258.
6) Akoh, R., Takuno, T., T., Matusi, D., & Maeno, S. (2024). Proposal of Efficient Analysis Method for Paddy Field Dam for River Basin Flood Control and Field Application. *Japanese Journal of JSCE*, 80(16), 23-16073.
7) Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
8) Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. *Morgan kaufmann*.
9) Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
10) Genuer, R., Poggi, J. M., Tuleau-Malot, C., and Villa-Vialaneix, N. Random forests for big data. *Big Data Research*, Vol. 9, pp. 28-46, 2017.
11) Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
12) Grant, T. D., Wischik, D. J., Grant, T. D., & Wischik, D. J. (2020). Finding Patterns as the Path from Input to Output. *On the path to AI: Law's prophecies and the conceptual foundations of the machine learning age*, 41-48. Cham: Springer International Publishing.

CHAPTER 3 METHOD VALIDATION WITH NUMERICAL EXPERIMENTS

3.1 Overview of numerical experiments and evaluation metrics

Numerical experiments were performed as method validation to examine the feasibility of using the augmentation approach, features and iterated prediction process described in the previous chapter in the configuration of RF model under constant rainfall conditions. First, a rainfall-only scenario with a bowl-shaped hypothetical terrain setting was constructed for the investigation of cases with and without the consideration of adjacent meshes to confirm the validity of the proposed feature selection. The portion of data splitting for training and validation, together with hyperparameter tuning were also performed aiming to optimize model accuracy and feature selection. Once the optimal features and parameter settings were determined, discussions on prediction interval using iterated prediction approach were then performed to balance model accuracy and computational cost. The parameters for the model with highest accuracy were then applied to a more complicated terrain scenario where river and inundation areas are included. Features further indicating boundary conditions and land uses were evaluated in order for the RF model to be applicable to future practical real-life scenarios.

Evaluation metrics for predictive machine learning (ML) models are typically divided into two major categories. Performance metrics for the effectiveness, efficiency and reliability of the model, which are mostly standardized evaluations in terms of the model's ability to classify instances, and error metrics for the quantification of discrepancy between predicted and true values[1)]. Performance metrics adopted in this study include confusion matrix, accuracy, precision, recall, F1-score, and coefficient of determinations (R^2), and error metrics for regression as root mean squared error (RMSE). The performance of RF model in the recognition of inundation areas is first evaluated with confusion matrix (**Table 3-1**), where

Table 3-1 Confusion matrix.

		Prediction	
		Inundated	Not Inundated
Simulation	Inundated	True Positive (TP)	False Negative (FN)
	Not Inundated	False Positive (FP)	True Negative (TN)

simulation is the result from HD model (i.e. the true value) and prediction is the output from RF model. Meshes with inundation depth above a certain threshold are considered inundated and vice versa. Based on the confusion matrix, accuracy, precision, recall and F1-score are further calculated with the following equations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3.4}$$

Here accuracy is the quantification of the proportion that is correctly classified, whether positive or negative, out of the total number of predictions. Precision refers to the ratio of correctively predicted as positives out of the total number of true positives. Recall, also known as true positive rate, is defined to be the proportion of all actual positives that are correctly classified as positives. And F1-score is interpreted as the harmonic mean between precision and recall. All four metrics range from 0 to 1 with 1 being the optimal value. Accuracy serves as a comprehensive evaluation incorporating all outcomes of a confusion matrix, while F1-score is especially useful when the datasets are imbalanced.

Furthermore, evaluations of variances using R^2 and RMSE in terms of regression using the following equations were performed.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{3.5}$$

$$RMSE(m) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{3.6}$$

where y_i (m) and $\hat{y}_i$ (m) are the simulated and predicted water depth at each mesh *i*, $\overline{y}$ (m) is the average simulated water depth, and *n* is the total number of meshes. The significant number for RMSE evaluation in the method evaluation sections is preserved to the scale below millimeter aiming to observe subtle differences between evaluated cases. And the practical field applications further consider the physical meaning and interpretation of the metrics, thus evaluated at centimeter scale. In order to investigate the difference in hydrodynamics characteristics of land uses, these metrics are also carried out with respect to each land use.

In addition, spatial agreement between prediction and simulation is examined in terms of water depth contours and difference maps to capture the ability to reproduce the spatial flood propagation process as well as detailed inundation depth at each specific location. Assessments regarding inundation area and inundation volume are also performed to evaluate the capability of RF model to quantify flood characteristics.

3.2 Rainfall-only scenario with bowl-shaped terrain setting

A bowl-shaped computational domain was constructed with the dimensions shown in **Fig. 3-1(a)** with elevation given by the following equation.

$$z = a(x^2 + y^2)(a = 5.0 \times 10^{-5}) \tag{3.7}$$

The size of meshes was 50 m, resulting in a total of 916 meshes. Here the land use was assumed to be miscellaneous for all meshes and the effective rainfall was carried out using 0.6, 0.9, 55 mm as the unsaturated runoff coefficient, saturated coefficient, and saturated rainfall threshold, respectively. The generation of training data of the augmentation approach was first carried out through flood analysis using HD model. Constant rainfall was applied uniformly to the computational domain. Five patterns of rainfall intensity were performed for the construction of training dataset (10 mm/hr, 15 mm/hr, 20 mm/hr, 30 mm/hr, 50 mm/hr). The calculation was carried out over a 20-hour period with an output interval of 10 minutes. Results obtained from HD simulations for all meshes at each time step were summarized into the training dataset for the RF model and the total number of training dataset was 554,180. The total training dataset was divided into subsets for model configuration process, within which feature selection and hyperparameter tuning were also performed, and subset for performance validation. Confirming the accuracy of prediction, the configurated RF models were further

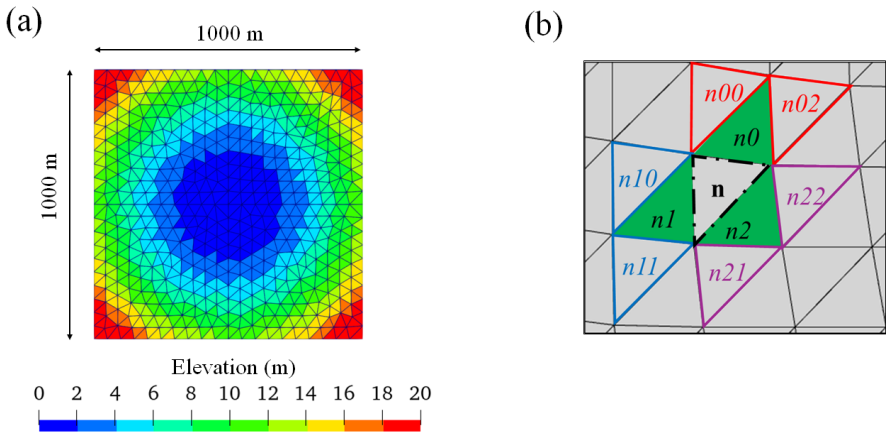


Fig. 3-1 Schematic of (a) bowl-shaped computational domain and (b) adjacent meshes.

tested for untrained rainfall intensities, and accuracy of predicted results were evaluated against calculated results from HD models.

3.2.1 Feature selection and hyperparameter tuning

The inundation depth of all meshes at each time step (h_t) were used as the target variable. RF models with different features were constructed to investigate the necessity of including features that represent hydrodynamic properties in the same logistics as HD model (**Table 3-2**). M1 included mostly the static properties of each mesh in terms of mesh number (n), coordinates (x, y), elevation (z) and temporal indications as of step number ($step$), rainfall at each step ($rain_t$) and cumulative rainfall ($rain_{cum}$). The temporal change of water depth at each mesh is determined by the flux exchange between adjacent meshes, that is, the prior state of adjacent meshes directly affect the current water depth at each target mesh. Therefore, how to represent this spatiotemporal factor as features is essential in the configuration of RF model. In this study, the spatial aspect was proposed by the selection of adjacent meshes, which include the three meshes directly connected to each target mesh (noted as n_i where i = 0, 1, 2) and secondarily connected adjacent meshes (noted as n_{ij}, where i = 0, 1, 2, and j = 0, 1, 2) as shown in **Fig. 3-1(b)**. As for the temporal aspect, the prior state water depth of the total 12 of adjacent meshes ($h_{t-1,\,i}$, $h_{t-1,\,ij}$) were selected as features to be further included in M2. As for the meshes located at the boundary with adjacent meshes falling outside of target area, the prior state information of the mesh itself was applied as the values for adjacent meshes. The assumption for this situation is that the boundary of RF model was carefully aligned with locally elevated topography that approximates wall conditions. While this approach introduces slight redundancy between local and adjacent features, it compensates for the missing neighbor values as well as acting as a boundary flag. On the other hand, the performance of the RF model is expected to degrade when applied over river basins under situations where flux exchanges occur across boundaries. Details for this phenomenon is discussed in later chapters when targeting flood predictions induced by practical rainfall events. The interval for prior state of features (i.e. t-1) was first arbitrarily selected to be 10 minutes, which is the same as the output interval for HD model.

Table 3-2 Features for model configuration.

Models	Features	Hyperparameter tuning
M1	n, x, y, z, $step$, $rain_t$, $rain_{cum}$	$n\_estimator$ =50, 100, 200 max_depth = 10, 20, 50 $mtry$ = 0.33, 0.5, 1.0, $sqrt$
M2	M1 + $h_{t-1,\,i}$, $h_{t-1,\,ij}$	

RF models involve several types of hyperparameters for the determination of the structure of each individual tree, the overall structure and size of the forest, and randomness. While several previous literatures have investigated the impact of hyperparameter setting on model accuracy, clear guidance is still missing and the choice for optimal values are dependent on the dataset[2]. Therefore, hyperparameter tuning was performed to confirm whether the arbitrarily selected hyperparameter contributes to optimal accuracy. The hyperparameter tuning procedure in this research adopted three most commonly used parameters: the number of trees, maximum depth of each tree, and the number of features to be considered to generate best split. The number of trees and maximum depth of each tree, denoted by *n_estimator* and *max_depth* in scikit-learn library (hereafter referred to as sklearn) of python respectively, are parameters controlling the size of the forest and higher values are generally preferable to obtain better performance[2]. On other hand, high values of these parameters also induce the risk of overfitting and result to large models, which could lead to constraints in memory of hardware and efficiency. The number of features at each split, denoted by *max_features* in sklearn, determines the correlation between individual trees. Lower values construct different, less correlated trees, thus returning better aggregated results. However, low values may also lead to average performance as trees are dependent on the random selection of features. A

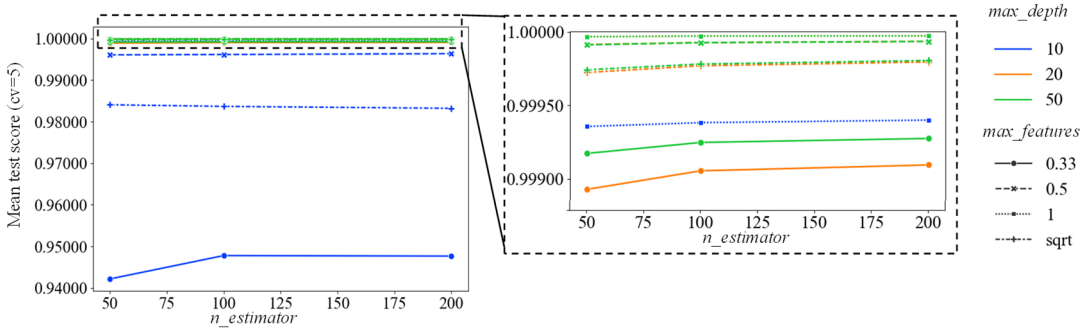


Fig. 3-2 Mean test score for hyperparameter tuning of M1.

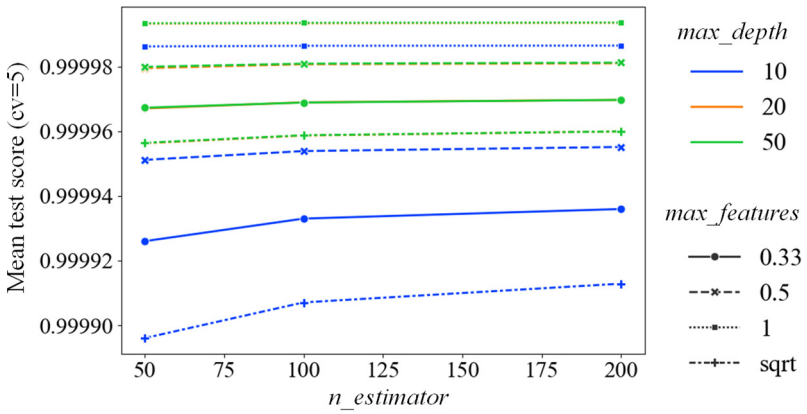


Fig. 3-3 Mean test score for hyperparameter tuning of M2.

commonly recommended value for number of features in regression is *sqrt(p)*, with *p* being the total number of features used in model configuration[3), 4)], yet higher values are also found to contribute to low error in other researches dealing with high dimensional data[5)]. The value is essentially dependent on the characteristics of the training data and guidance regarding the selection of this parameter is still missing. Therefore, hyperparameter optimization process is necessary to obtain RF model with high performance.

The optimization of hyperparameter in this study was performed with grid search method, denoted by the *GridSearchCV* library in sklearn, which considers an exhaustive evaluation of all combinations of hyperparameters given[6),7)]. The range of hyperparameter search is summarized in **Table 3-2**, and performance validation was carried out with 5-fold cross-validation using R^2 as the scoring metric. Depicted in **Fig. 3-2** and **Fig. 3-3** are the mean test score of the cross validated results for M1 and M2 respectively. For both of the models, performance increases as the depth of tree grows, with high accuracy R^2 of almost 1.0 for depth greater than 20. The difference between depth of 20 and 50 is rather negligible, especially for M2. An improvement in accuracy is also observed as the number of trees grows from 50 to 100, yet the difference between 100 and 200 trees, with other parameters being the same, is less significant for both M1 and M2. As for the selection of *max_features*, the general trend observed in this study is that among the selections of 0.33, 0.5, and 1.0, higher values result in better outcomes. The performance using *sqrt(p)* concludes to be the model of lowest and second lowest accuracy for M1 and M2, respectively. On the other hand, the feature importance of M2 for prior depth suggests that using all features at each split may introduce bias, even though having the highest accuracy among all parameters (**Fig. 3-4**). Shown on the horizontal axis represents the prior state water depths of directly connected meshes, h_i for corresponding meshes of n_i, and those of secondarily connected meshes, h_{ij} for meshes denoted as n_{ij} in **Fig. 3-1(b)**, where i and j are indices for mesh numbering. Following the local mass balance of hydrodynamic model, the water depth for each target mesh is a weighted sum of

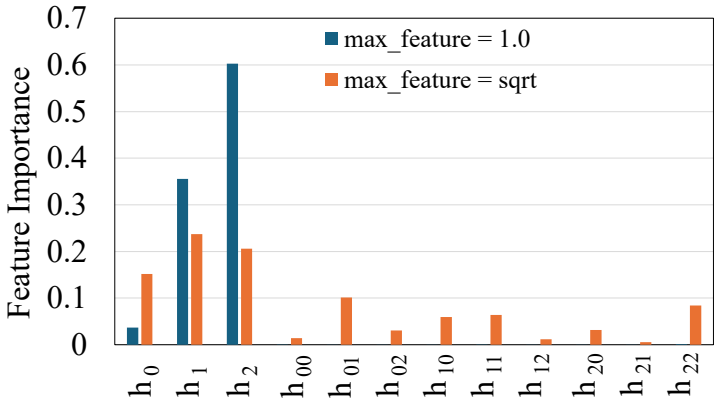


Fig. 3-4 Feature importance of prior depth for M2.

Table 3-3 Evaluation results for validation set.

	M1		M2	
	Case 1	Case 2	Case 1	Case 2
R^2	0.999	0.999	0.999	0.999
RMSE (m)	0.0027	0.0031	0.0042	0.0044

flux contributions from adjacent meshes. Since the geometry of computational domain was designed to be roughly symmetric and the area of meshes roughly the same, the contribution from each adjacent mesh is expected to be similar. And similar trend should be observed for secondary connected meshes. That is, the distribution of feature importance using *max_feature = sqrt* better coincides with the logistics of HD model that prior state of the directly connected meshes contributes approximately equally to current water depth, with slightly less contribution for further adjacent meshes. The results are also consistent with the findings from other studies where higher values are sometimes over-credit features with strong contribution while masking fewer influential variables[5]. The small stochastic differences between adjacent meshes are considered to be induced by the process of bootstrapping used in RF models and the not perfectly symmetric distributions of meshes, both inducing statistical subtle correlations in the actual field. Therefore, taking into consideration the comprehensive influence of the hyperparameters, the number of features for model configuration in the following investigations was selected to be *sqrt(p)* such that the model provides relevant performance gains of less influential yet important features. In terms of the other two hyperparameters, since the difference between depth of 20 and 50, and number of trees of 100 and 200 was insignificant, values of 20 and 100 were selected in order to reduce the size of configurated model.

Using the optimized hyperparameters, the process of subset splitting into training and validation dataset was performed with the built-in library *train_test_split* of sklearn, with the consideration that data division may affect the accuracy of model if the training data consists of significantly imbalanced data or the size of dataset is small. Two cases were investigated where Case 1 used 80% of the training data randomly selected for model construction and the rest 20% of data used for model validation, and Case 2 used 70% for training and 30 % for validation. Data shuffling was applied using the same random seed number to maintain reproducibility of results for both M1 and M2 and summarized in **Table 3-3** is the evaluation of performance against validation set. The R^2 for both models regardless of partition ratio indicates high accuracy, and RMSE for all cases also results in error of less than 0.01 m. Therefore, it is concluded that the impact of the ratio in splitting dataset is rather insignificant

when the size of training data is large enough, and the results for the following discussion uses model configured with 70% of total training data.

3.2.2 Model performance on test dataset

M1 and M2, constructed using different features, were then tested against two untrained rainfall intensities: 25 mm/hr and 40 mm/hr. Considering the features of prior state used in the configuration of M2, assumptions were first made that this information is correctly known for the purpose of prediction accuracy investigation in this section. That is, the input for all prior state ($h_{t-1,i}$, $h_{t-1,ij}$) at each step of prediction used the simulation results from HD models such that the error evaluations can be performed for each prediction rather than accumulating overtime. Regarding the concerns of data scarcity and computational cost of RF models in practical applications, discussions of iterated prediction process and investigations of other intervals for prior state and corresponding accuracy evaluation were carried out and summarized in the following subsection.

(a) Performance evaluation on rainfall intensity of 25 mm/hr.

Illustrated in **Fig. 3-5** are the prediction evaluation for M1 with respect to time using both performance and regression metrics. The accuracy and precision scores are high over the entire period. The recall and F1-score, in comparison, varies more significantly between predictions.

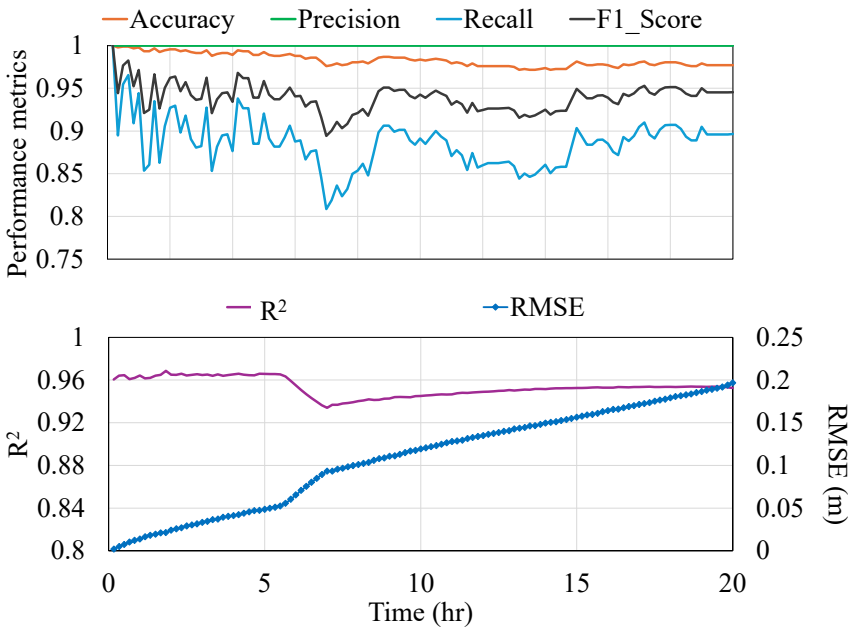


Fig. 3-5 Performance metrics and RMSE evaluation for M1 (25 mm/hr).

Table 3-4 Confusion matrix evaluation for M1 at 7 hr.

		Prediction		Sum
		Inundated	Not Inundated	
Simulation	Inundated	93	22	115
	Not Inundated	0	801	801
Sum		93	823	916

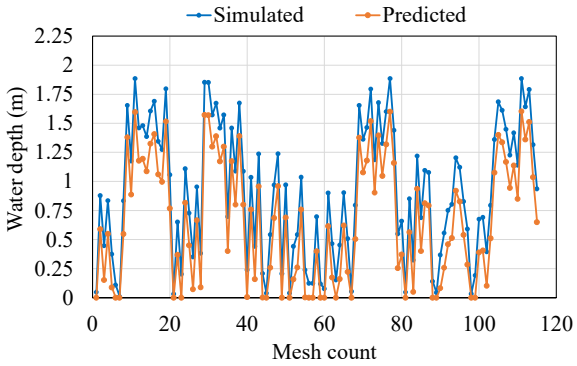


Fig. 3-6 Comparison of predicted water depth to calculated water depth.

It is also notable that while the R^2 for M1 is greater than 0.9 over time, which is also commonly considered as high performance, the RMSE gradually increases to a maximum error of 0.2 m. To evaluate what contributed to inconsistency between evaluation metrics, a detailed analysis was performed taking the results at 7 hr, where the recall, F1-score and R^2 values are the lowest, as an example. **Table 3-4** summarizes the count of meshes in terms of confusion matrix with a threshold that meshes with water depth greater than 0.01 m are classified as inundated. The intentionally designed computational domain concentrates flow at the center of the area such that only approximate 13% of meshes are classified as inundated at this time. Because of this heavily imbalanced dataset, the high accuracy and precision score of 0.976 and 1.0 only evaluates to have low recall rate and comprehensive F1-scores of less than 0.9. Furthermore, based on the proportion of meshes predicted to be false negative, it is indicated that M1 has a tendency of underestimation. To understand how this error is distributed between meshes, a comparison of water depth (**Fig. 3-6**) and corresponding error scatterplot (**Fig. 3-7**) of prediction against calculation for the inundated meshes are plotted. It can be observed from both figures that M1, while capturing the general trend of water depth, fails to provide inundation depth by constantly underestimating depth by approximately 0.27 m. This result further confirms that other features as a representation of hydraulic exchange of water are necessary for better prediction of water depth.

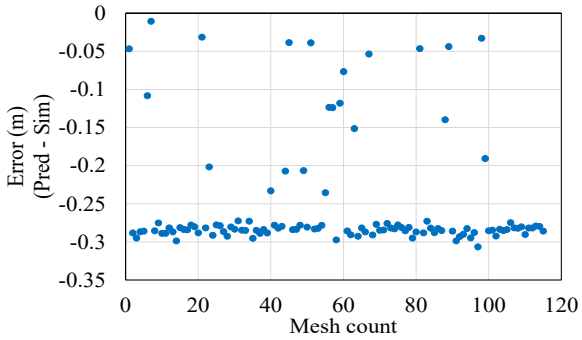


Fig. 3-7 Scatter plot for prediction errors compared to simulation results.

The performance evaluation for M2, configured by further including prior state water depth of each target mesh and adjacent meshes, using the same metrics is depicted in **Fig. 3-8**. The average temporal variation of depth obtained from HD simulation (*avg_delta_h*) is also illustrated in the same figure and the two abrupt increases of incremental depth variation correspond to the wet/dry transient adjustment phase and the shift from unsaturation condition to saturation condition respectively. Despite certain variations between predictions, the accuracy and recall for M2 during the entire period are close to the optimal value. Having slightly low values in the first few predictions in terms of R^2, precision and F1-score, these metrics also gradually approach 1.0, which suggests that M2 overall outperforms M1 in classifying inundated meshes. As for predictions regarding detailed water depth, the RMSE of

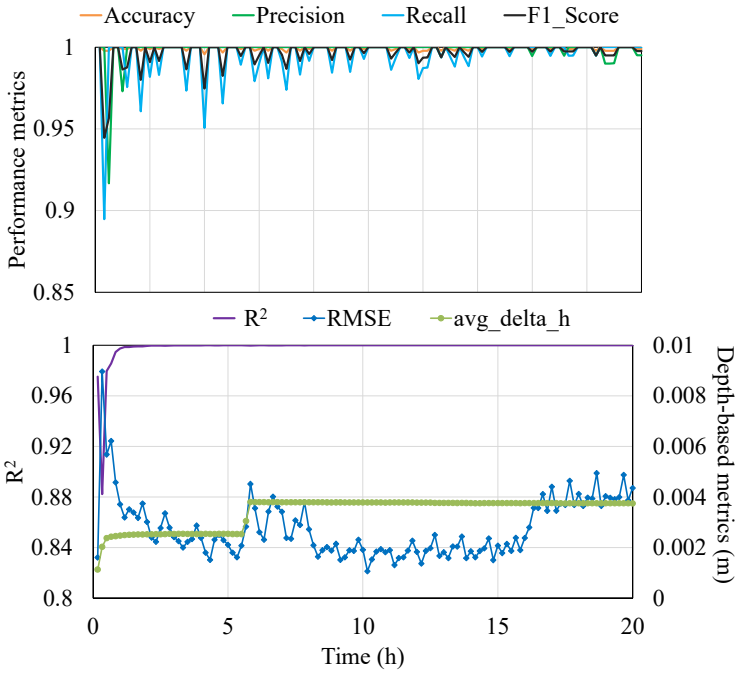


Fig. 3-8 Performance metrics and depth-based metrics evaluation for M2 (25 mm/hr).

M2 is significantly improved to less than 0.01 m, and the issue of increasing trend over time in M1 is also resolved. Illustrated in **Fig. 3-9** is the distribution of error, defined to be the difference between predicted water depth and calculated value, with respect to calculated water depth for the entire calculation period. It can be observed that the errors cluster tightly around perfect prediction, with some outliers shown as transparent points at shallow water depths. The spread of error across different water depths also clusters closely within the range of -0.025 m and 0.025 m, suggesting that the prediction overall has low bias. Furthermore, a water depth contour comparison was carried out to better visualize the difference between models. Illustrated in **Fig. 3-10** and **Fig. 3-11** are the depth contour for HD models and predictions at 7 hr and 20 hr, respectively, and a difference contour of predicted values and calculated values for M2 is also depicted to observe the minor errors. The results are consistent with the previous findings that M1 has a tendency for underestimation. Comparing the results between different steps, it is also clear that this bias is more significant in terms of underestimating both inundation areas and water depth as flow spreads out over time. In comparison, the performance for M2 is significantly improved. Although having a slight overall

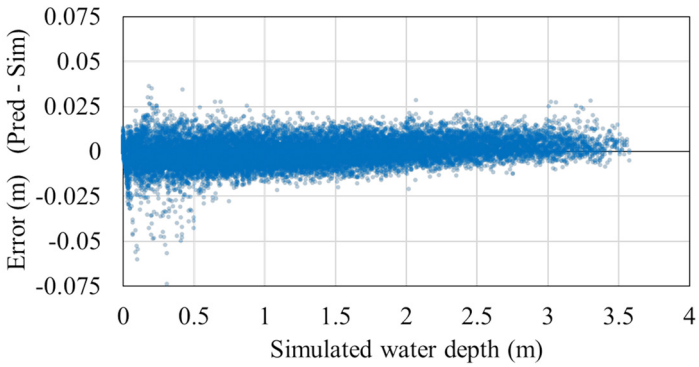


Fig. 3-9 Error distribution with respect to depth for M2 over entire period (25 mm/hr).

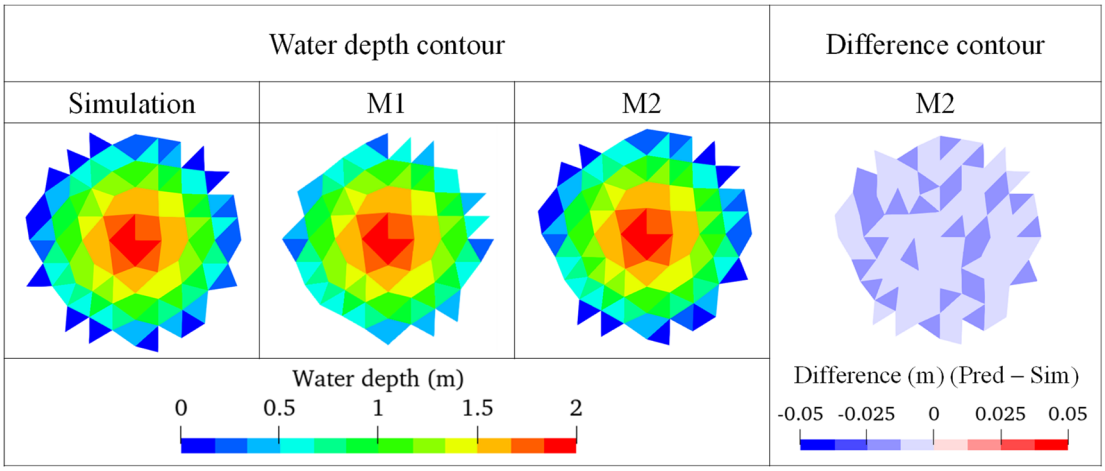


Fig. 3-10 Contour comparison at 7 hr (25 mm/hr).

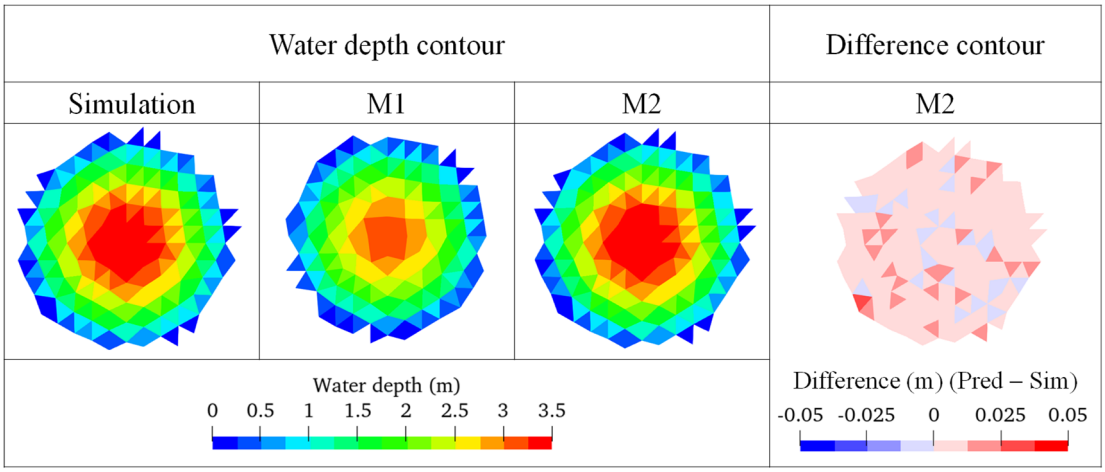


Fig. 3-11 Contour comparison at 20 hr (25 mm/hr).

underestimation at 7 hr, the error is mostly suppressed within 0.01m. And the performance in predicting both inundation area and depth is also high for 20 hr with minor errors of approximately 0.01m overestimations, which is consistent with the observations in **Fig. 3-9** that M2 slightly biases towards higher prediction when water depth exceeds 3.0 m. Based on the performance metrics and regression evaluation results of M1 and M2, together with visualizations, it is concluded that including information on past water depths as features significantly improves the performance of RF model. The finding is also consistent with the same logistics in HD models that current water depth is determined by characteristics of prior state in terms of spatiotemporal flux exchange between meshes.

(b) Performance evaluation on rainfall intensity of 40 mm/hr.

M2 was further tested for untrained rainfall intensity of 40 mm/hr to confirm the generalization capability of the model and depicted in **Fig. 3-12** are the overall performance evaluations using the same metrics, together with the temporal variation of average depth simulated by HD model. Compared to the results evaluated against 25 mm/hr of rainfall, the overall performance for all classification metrics decreases. While a slightly lower accuracy is also observed during the first few steps of 25 mm/hr rainfall, the accuracy degradation in the classification of inundated meshes is more obvious for 40 mm/hr case. The reason for this trend in both cases is due to the small number of inundated meshes over the early stage of simulation such that even one misclassification could lead to large variation in metrics. The accuracy for the prediction of detailed water depth also decreases compared to the previous test cases. Even though the RMSE values are still below 0.01 m, the error variation with respect to each mesh, especially for the inundated meshes, is considered high. As illustrated in **Fig. 3-13** the range of error becomes greater with some dense distributions exceeding absolute value of 0.025 m. Similar underestimation at low water depths is observed, but the errors values of

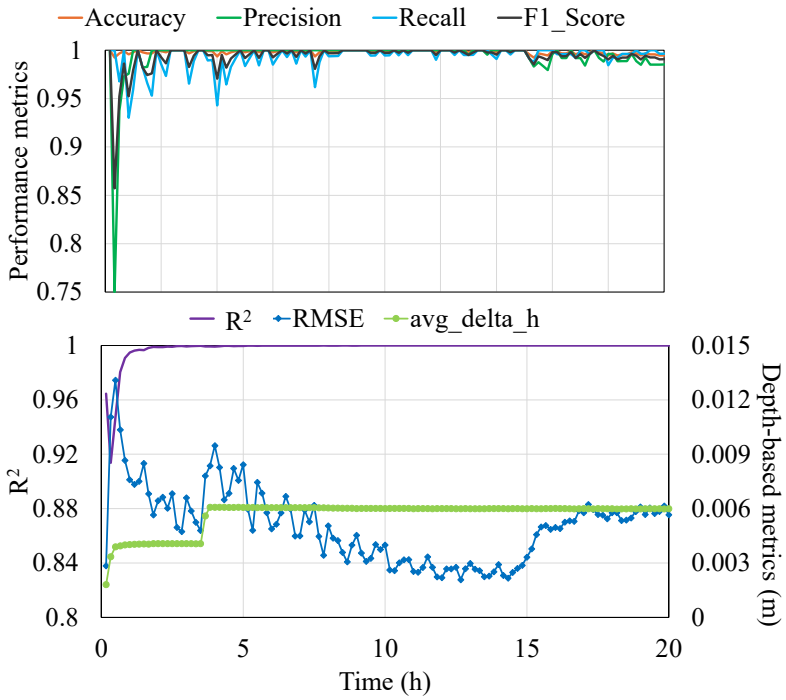


Fig. 3-12 Performance metrics and RMSE evaluation for M2 (40 mm/hr).

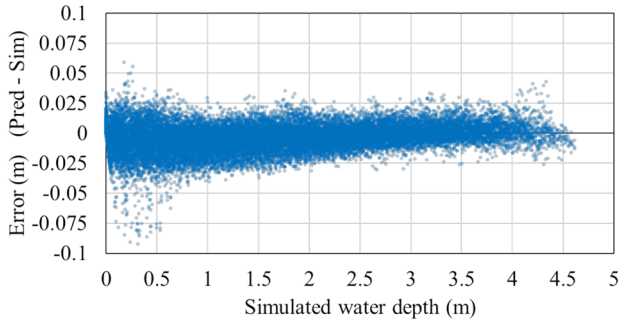


Fig. 3-13 Error distribution with respect to depth for M2 over entire period (40 mm/hr).

-0.075 m or greater indicate a relative error of more than 15%, which is considered rather significant. The higher density as indicated by darker color also suggests that the model struggles at predicting lower depths for rainfall of 40 mm/hr.

Comparing the results over two test rainfall, one reason that may contribute to the overall low performance in the prediction of rainfall of 40 mm/hr is considered to be the limited number of training data exceeding the target event. In the 5 patterns of rainfall intensities used in the configuration of M2, only one pattern exceeded the rainfall intensity of 40 mm/h. Even though the target still falls in the range of interpolation, the lack of data on extreme cases with higher water depth could lead to bias. Therefore, another model was configured for comparison using the same parameter setup for M2, except for extra training data of rainfall intensity of 45 mm/hr. **Fig. 3-14** depicts the error distribution for the model with extra training data and

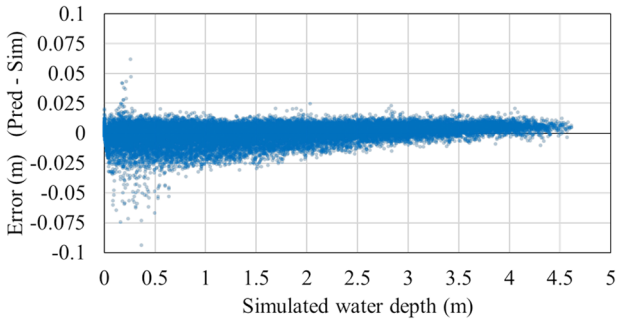


Fig. 3-14 Error distribution with respect to depth for M2 with extra training data (40 mm/hr).

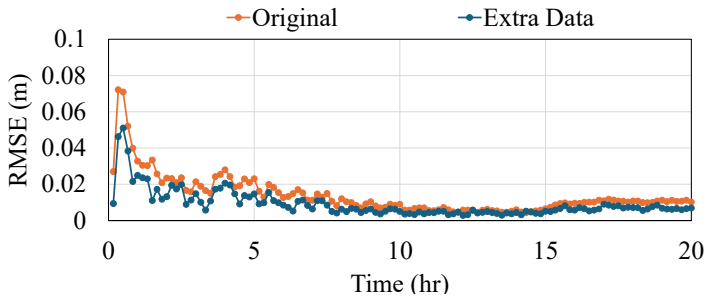


Fig. 3-15 RMSE of inundated meshes for M2 using different training data (40 mm/hr).

the RMSE for the inundated meshes are extracted for easier comparison of the two models (**Fig. 3-15**). With the inclusion of the extra training data exceeding the intensity of the target rainfall event, the range of error further concentrates towards perfect predictions. The number of outliers with high errors at lower water depths in the original model decreases, and the prediction for meshes with higher water depths greater than 3.0 m is also more accurate as shown by the narrower cluster. The model with extra data also outperforms the original by further improving RMSE of inundated meshes within 0.01 m for the majority of calculation period. When plotting the depth contour and difference of prediction and calculation at 20 h, it can be observed that the variation of error is smaller for the one with extra data (**Fig. 3-16**). Although the results still bias slightly towards overestimation, the improved performance over the entire calculation period suggests that the range of training data significantly influences the accuracy of prediction. Therefore, the selection of training data should be investigated carefully taking into consideration in accordance with target event.

3.2.3 Iterated prediction evaluation

This section further investigates the process of obtaining the prior state information of M2 and the intervals for predictions aiming for practical applications of the augmentation approach. The assumption in the performance evaluation of previous session used the results

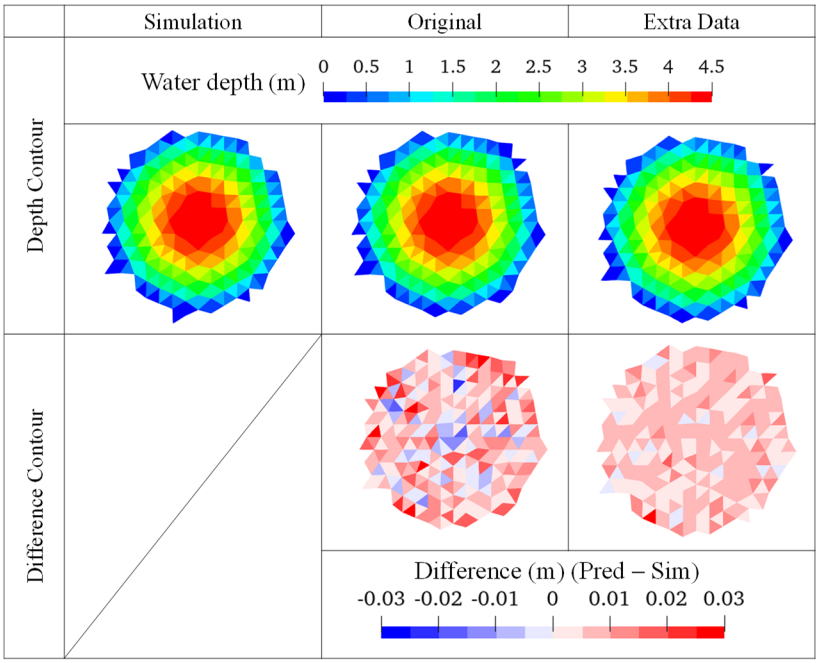


Fig. 3-16 Contour comparison at 20 hr of M2 using different training data (40 mm/hr).

from HD model (i.e. the correct previous information on water depth) as the input for each prediction. Yet the temporal variation of water depth of inundation areas over large river basins is generally not available in real-life situations. To overcome such data scarcity of previous water depth information, an iterative prediction process is proposed in this research. Specifically, the prediction stage starts from an initial stage right before precipitation occurs when no water depth exists, that is prior state of meshes equal 0 m. The first prediction is made by inputting precipitation conditions. From this prediction forward, each predicted output, in terms of water depth at each mesh, is reconstructed as prior state information for the input of the next prediction based on mesh number and connectively such that the prediction process iterates purely with RF models over time without the use of calculation results from HD models. Another factor to consider is the interval for prior state of water depth. Shorter intervals are desired for previous hydraulic states to be correctly captured. Yet computational cost would increase for shorter intervals provided that the total period for prediction is fixed. Another concern is the accumulation of error overtime as he iterated prediction process uses only predicted results for input without calibrations. Therefore, it is important to determine the optimal interval such that prediction accuracy and computational cost could be balanced.

This study investigated three intervals of prior state: 10-minute, 1.0-hr and 2.0-hr intervals and carried out accuracy evaluation comparison with and without iteration processes. For the simplicity of method validation, this section investigates only test prediction on rainfall of 25 mm/hr using the original M2 model, which was constructed with 5 patterns of constant rainfall as training data. Summarized in **Table 3-5** and **Table 3-6** are the R^2 and RMSE values for

Table 3-5 R^2 summary with different intervals.

Interval	Time (hr)			
	2	4	6	8
10-min	0.9995	0.9999	0.9999	1.0000
1.0-hr	0.9942	0.9991	0.9994	1.0000
2.0-hr	0.9658	0.9931	0.9967	0.9976

Table 3-6 RMSE(m) summary with different intervals.

Interval	Time (hr)			
	2	4	6	8
10-min	0.003	0.002	0.004	0.003
1.0-hr	0.010	0.007	0.007	0.007
2.0-hr	0.024	0.018	0.018	0.020

different intervals during the first 8 hr of prediction without iteration. The overall accuracy of the three intervals is high with R^2 greater than 0.95. As the interval increases, a gradually decreasing performance is observed with an average error increased to approximately 0.02 m for 2.0-hr interval. This result suggests that prior state with long intervals might be insufficient to reflect current status since the temporal change of water depth in this research is dominated by the event-based rainfall characteristics, which is different from those applications targeting long-term seasonal predictions of water level. Another observation is the slightly high error at 2 hr for all cases. This error is induced by the setting of the elevation where only a limited number of meshes are inundated during early stage of calculation, which is also observed from the test evaluations in the previous subsection that the model struggles at predicting shallow water depth by biasing towards underestimation, While not having a significant impact overtime when evaluated without iteration, this high initial error could lead to compromised performance in the iteration prediction process since the input of prior state at each prediction is not calibrated. Shown in **Fig. 3-17** is the comparison of predicted water depth against

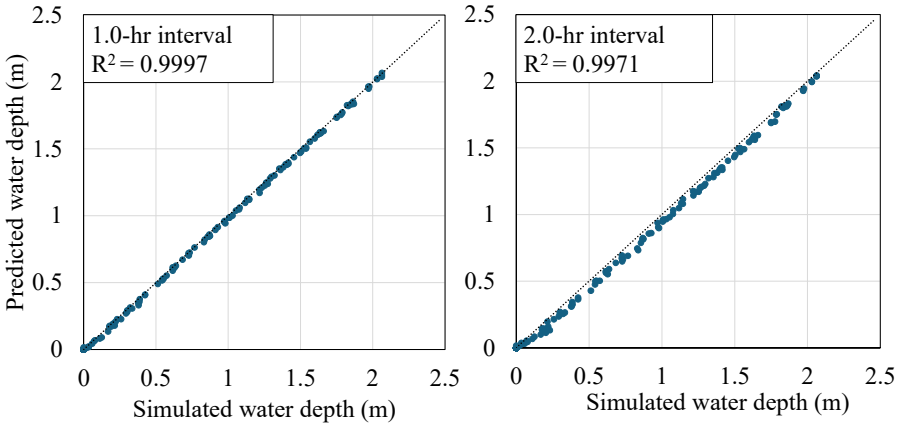


Fig. 3-17 Comparison of iterated prediction for 8 hr with different intervals.

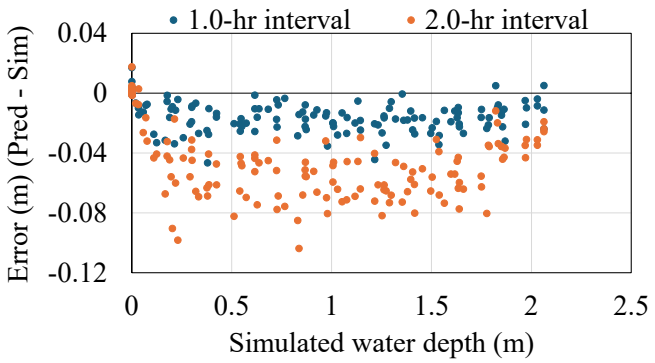


Fig. 3-18 Error scatterplot for iterated prediction at 8 hr with different intervals.

calculated values at 8 hr using iterated prediction with intervals of 1.0 hr and 2.0 hr respectively, and corresponding error is depicted in **Fig. 3-18**. The performance of prediction at 8 hr for both intervals is rather compromised with an overall underestimation, and longer interval for prior states resulted in lower accuracy. In summary, the results suggest that models with shorter interval outperforms the ones with long intervals, with or without iteration. The iterated prediction process, while solving the lack of prior state information of water depth to be used as input for prediction, is vulnerable to initial error. Considering that the prediction without iteration resulted in high accuracy and the fact that the elevation setting for this scenario is rather simple, the potential of applying this augmentation approach using iterated prediction process is further investigated under complex terrain conditions.

3.3 Complex terrain scenario with river and inundation areas

Aiming for the application of the augmentation approach under practical terrain fields, a complex terrain scenario was designed to further investigate features related to representations of land use and river boundary conditions, as well as the feasibility of iterated prediction. The elevation and land use distribution of the terrain is depicted in **Fig. 3-19**. The mesh size was 10 m for the boundary close to the mountain side, and 5 m for the rest of the area, which resulted in a total number of 19682 meshes. A 5-meter river run straight with a slope of 1/1250. The mountain on the right side started with an elevation of 0.7 m and sloped upward with a 0.3 % gradient. 5 square paddies with length 100 m lined along the river with bottom elevation of 0.7 m. The ridges between paddies and other land uses were represented in 1D with a height of 0.3 m, and the slope of the area classified as miscellaneous land use was 0.1%. To take account for difference in land use, parameters of Manning's coefficient, runoff coefficient and

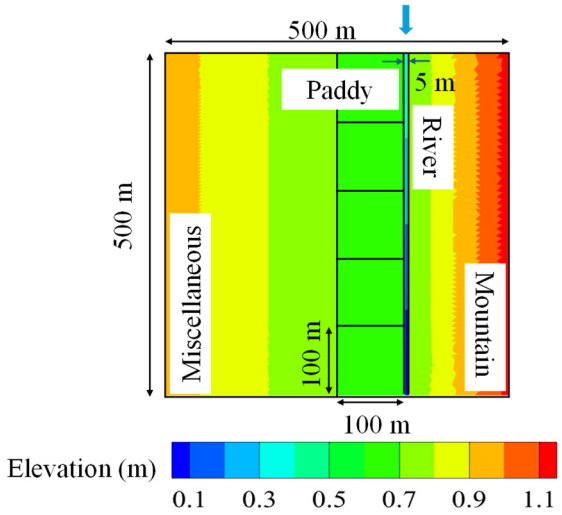


Fig. 3-19 Elevation and land use of the complex terrain.

Table 3-7 Parameter setting with respect to land use in HD model.

	Manning's coefficient	Runoff coefficient		Saturated rainfall (mm)
		Unsaturated	Saturated	
Mountain	0.6	0.3	0.9	150
Paddy	0.02	0.5	0.9	55
River	0.015	1.0	1.0	-
Miscellaneous	0.025	0.6	0.9	55

saturated rainfall were set up separately with the values in **Table 3-7** in the calculation of HD model. The initial water depth in paddy was 0.15 m, and saturated condition was assumed as the initial state for paddy. River levee was also configured along the edge of triangular meshes as 1D representations and the height was 0.3 m higher than the maximum elevation of adjacent land uses. The inflow discharge of river was set to be a constant of 1.5 m^3/s, and the downstream water level was calculated assuming normal flow conditions. Run-up calculation was first performed under no rainfall condition using the constant discharge until the longitudinal water level in river stables. The calculated water depth and velocity obtained from run-up calculation was then used as the initial condition. The rainfall condition used for the construction of training data involved the same constant rainfall intensities of 10 mm/hr, 15 mm/hr, 20 mm/hr, 30 mm/hr, 50 mm/hr and target event for prediction was 25 mm/hr. **Fig. 3-20** illustrates the water depth contour of simulated results over time for the target event. During the first few hours, water depth accumulates within paddies and flow gradually spreads over inundation areas with shallower depth for mountains due to the difference in runoff properties between land uses. Paddies approach full storage capacity around 6 hr and overflow from paddies occurs starting 7 hr. The increase of water depth in paddies, therefore, becomes less significant afterwards and the volume of overflow starts to contribute to the depth variation of river and other land uses.

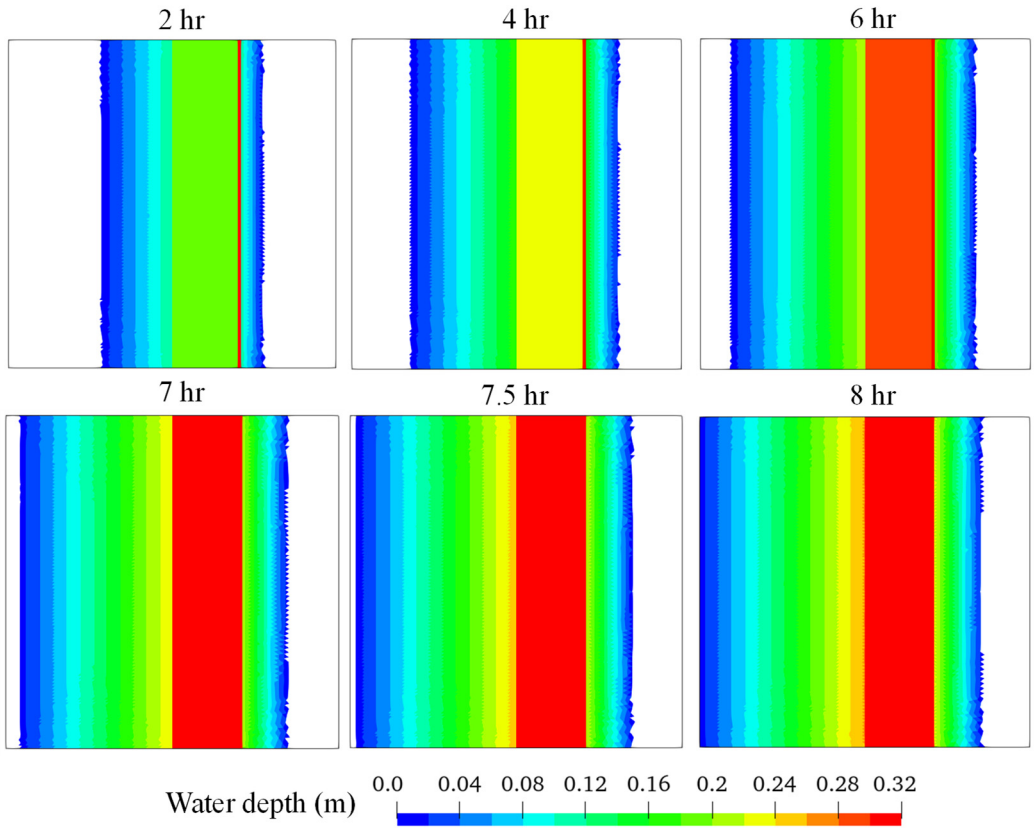


Fig. 3-20 Water depth contour under 25 mm/hr rainfall (HD model).

3.3.1 RF model configuration

As observed from the simulations of HD model, the hydraulic characteristic in this section is more complicated compared to the previous terrain setting. Because of the inherited difference in saturation thresholds, the accumulation of water due to precipitation differs between land uses. With complexity in elevation and flow exchanging at the boundary of computational domain, the direction of hydraulic gradient in inundation areas changes significantly both spatial and temporally, especially when overflows occur. Therefore, more features should be considered in the configuration of RF model to reflect the extra trigging factors of the variation in water depth. **Table 3-8** displays the features investigated in this section. M1 adopted the same selection from the previous investigation and M2 included the elevation of adjacent meshes (z_i, z_{ij}). Since the driver for flux between meshes is essentially the difference in hydraulic gradient (i.e. water surface elevation), the additional elevation information, together with water depth of adjacent meshes is expected to represent the water surface characteristics of neighbors to some extent. With rivers flowing across the computational domain, the inflow and outflow discharge is another important factor

Table 3-8 Features investigated in model configuration.

Models	Features
M1	n, x, y, z, $step$, $rain_t$, $rain_{cum}$, $h_{t\text{-}1,\,i}$, $h_{t\text{-}1,\,ij}$
M2	M1 + z_i, z_{ij}
M3	M2 + Q_{in}, Q_{out}
M4	M3 + land use (binary)

dominating the total control volume of water besides precipitation. Therefore, M3 included the discharges at river boundaries (Q_{in}, Q_{out}) as features and applied the same values for all meshes, with the assumption that the values are observable. As for the consideration of land use, the one hot encoding method was used to convert categorical variables into binary format to eliminate ordinality[8]. Since the difference in categorical land use does not have inherent order, assigning numerical values may introduce ranking of the features and lead to biased prediction. Instead, categorical features of paddy, river and mountain were incorporated in M4, and values of 1 were assigned to corresponding meshes where the land use is present, and 0 was assigned for the rest of categorical features. In this way, the land use factor was incorporated within the regression algorithms while eliminating bias.

Shown in **Fig. 3-21** is the feature importance of the configurated models where footnotes of the features denote corresponding index of adjacent meshes. It can be observed that the inherited static properties of coordinates and land use in general result in low importance values since these properties do not change over time. The elevation of each mesh, together

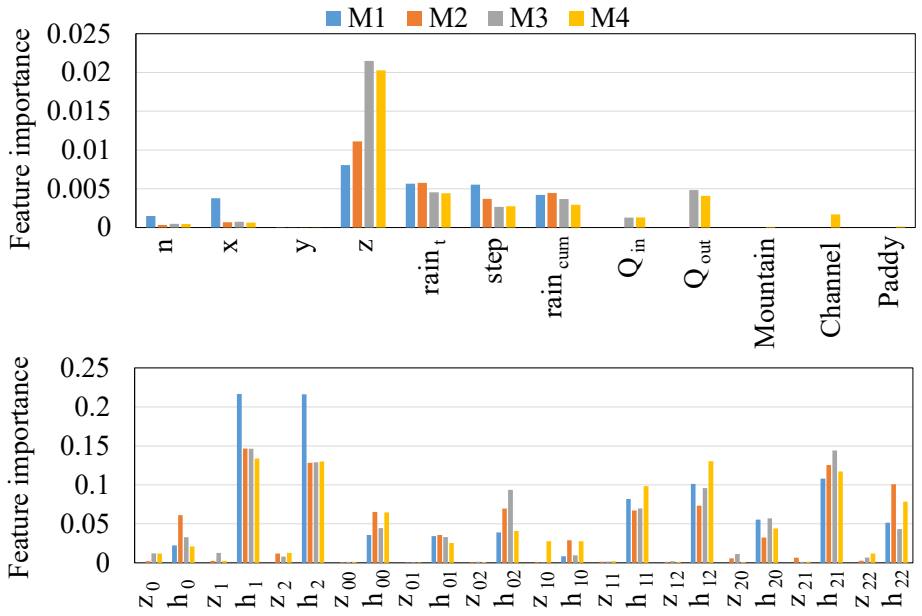


Fig. 3-21 Feature importance of configurated model.

with adjacent meshes are considered important to some extent. For M3 and M4 considering river boundary conditions, the importance of outflow discharge is slightly higher than inflow. The reasoning for this phenomenon is the difference of constant inflow versus changing outflow discharge under normal flow condition. Among the three categorical land use features for M4, the importance values decrease in the order of channel, paddy and mountain, which is consistent with the complexity of hydraulic change that most significant depth variation occurs in rivers during the entire period while that for paddy and mountain starts with simple accumulation of precipitation and varies significantly only under overflow conditions. Since static properties do not differentiate between cases, the contribution of such features in impurity reduction of feature importance calculation is rather insignificant, which results in an averaged lower value. The temporal varying features of prior state of water depth, in contrast, hold high feature importance. Especially observed for M1, the importance of directly connected meshes with indices of 1 and 2 overweighs that with index 0, which could be contributed from how the meshes are numbered under regular shapes of this specific terrain setting. The variations of prior water depth between M2, M3 and M4 are rather insignificant, with a total importance value greater 0.9. This results again confirm that prior state information of adjacent meshes contributes significantly in the determination of current water depth in RF model configuration.

3.3.2 Model performance evaluation

The model performance evaluation was carried out testing against 25 mm/hr of rainfall using 2-hr iterated prediction process and summarized in **Table 3-9** is the R^2 comparison during the first 8 hr. All having high R^2 values, M3 and M4 slightly outperform the other two models. Even though the iteration error accumulates over time, as indicated by the decreasing accuracy of all models, the overall performance of the RF models is promising. To further observe how the error is distributed, the RMSE with respect to each land use was evaluated and detailed in **Fig. 3-22**. Despite the differences in performance, similar trends are observed for all models. First, the RMSE gradually increases up until 6 hr regardless of land use, which is also consistent with the results of R^2. The reason contributing to this observation is the error accumulation of the iteration process since the input for each prediction directly takes the previous predicted results without calibration. Further proving this reasoning is the decrease of RMSE in paddies at 8 hr while all other land uses having increasing errors. As observed from the result of HD mode, the water depth of paddy stables after overflow occurs after 6 hr, that is the incremental rise of water depth is shallow. On the other hand, the water depth in paddy during earlier steps has been underestimated and the error is increasing by an almost

Table 3-9 R^2 comparison of iterated prediction with 2-hour interval.

Models	Time (hr)			
	2	4	6	8
M1	0.9942	0.9846	0.9785	0.9707
M2	0.9941	0.9868	0.9788	0.9773
M3	0.9967	0.9860	0.9791	0.9861
M4	0.9971	0.9893	0.9801	0.9829

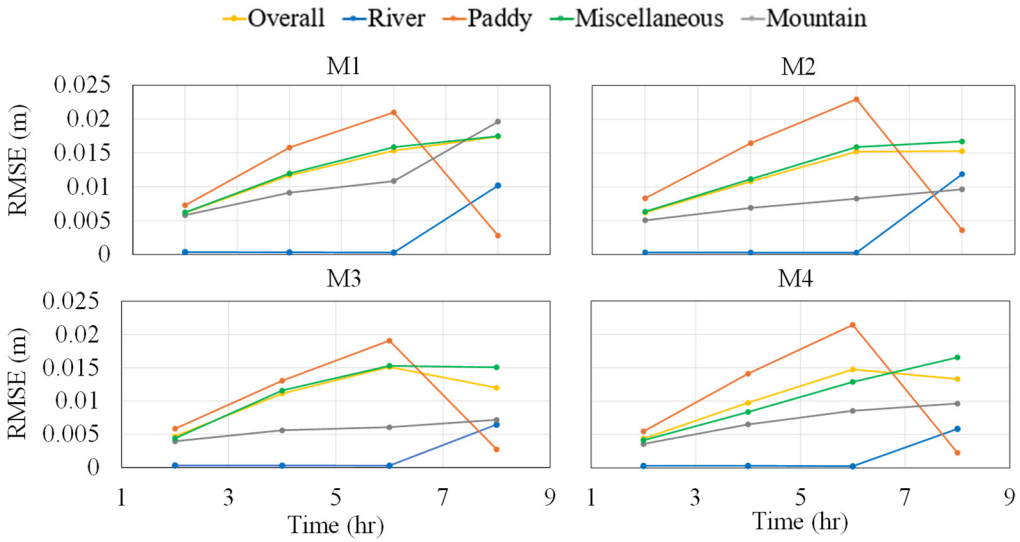


Fig. 3-22 RMSE comparison with respect of land use of iterated prediction with 2-hour interval.

constant rate, which is as shown by the contour maps of iterated prediction on water depth and difference to simulated results illustrated in **Fig. 3-23** and **Fig. 3-24**. This difference of water level rise characteristic of the simulated and predicted values leads to a smaller RMSE at 8 hr when compared relatively. Another common observation for all four models is the extremely low RMSE for rivers. Different from other land uses, the flow velocity in river is high and the dominating factor of water depth, therefore, is the discharge balance of river boundaries rather than direct precipitation. Since the flow rate is constant during the first 6 hr, the water depth does not change significantly and resulting accuracy is high, even for M1 and M2 without features representing river condition. However, the variation in hydraulic characteristics becomes complicated compared to previous steps after overflow from paddies occur. The corresponding accuracy decreases as observed by the sudden increase in RMSE and M3 and M4 with features taking into account for river discharge resulted in lower RMSE in rivers.

All having overall RMSE within 0.02 m, the overall performance of all models is considered promising and the difference between models is not significant based on this evaluation metrics. On the other hand, the contour maps, illustrating the spatial distribution of errors,

further conclude the importance of the additional features discussed in the previous section. Comparing the predicted water depth contour to simulated results shown in **Fig. 3-20**, it is noted that M1 resulted in high errors in mountain areas. In contrast to other areas with rather flat gradient, the steep slope over mountains leads to rapid runoff instead of ponding and smooth-spreading waterfront. That is, without information of the elevation of adjacent meshes, M1 fails to represent water surface gradient, and the corresponding direction of flow is not correctly captured. The prediction of inundation area using the other three models aligns highly with the simulated results, with M2 having an overall tendency for underestimation while M3 and M4 having a mix of overestimation and underestimation at different land uses. It is implied that the significant overestimation of M3 over the left side of paddy is due to the failure in

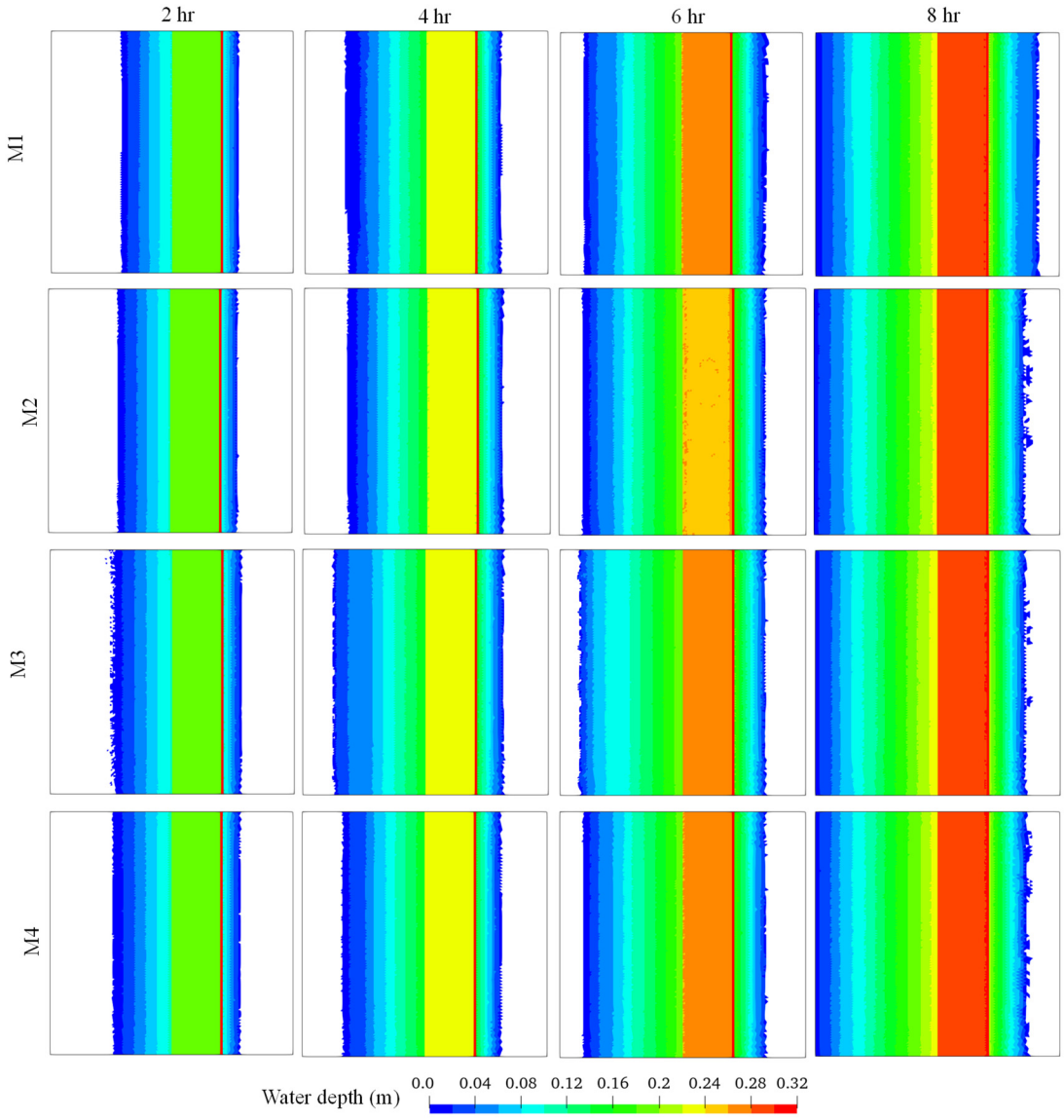


Fig. 3-23 Predicted water depth contour comparison of iterated prediction.

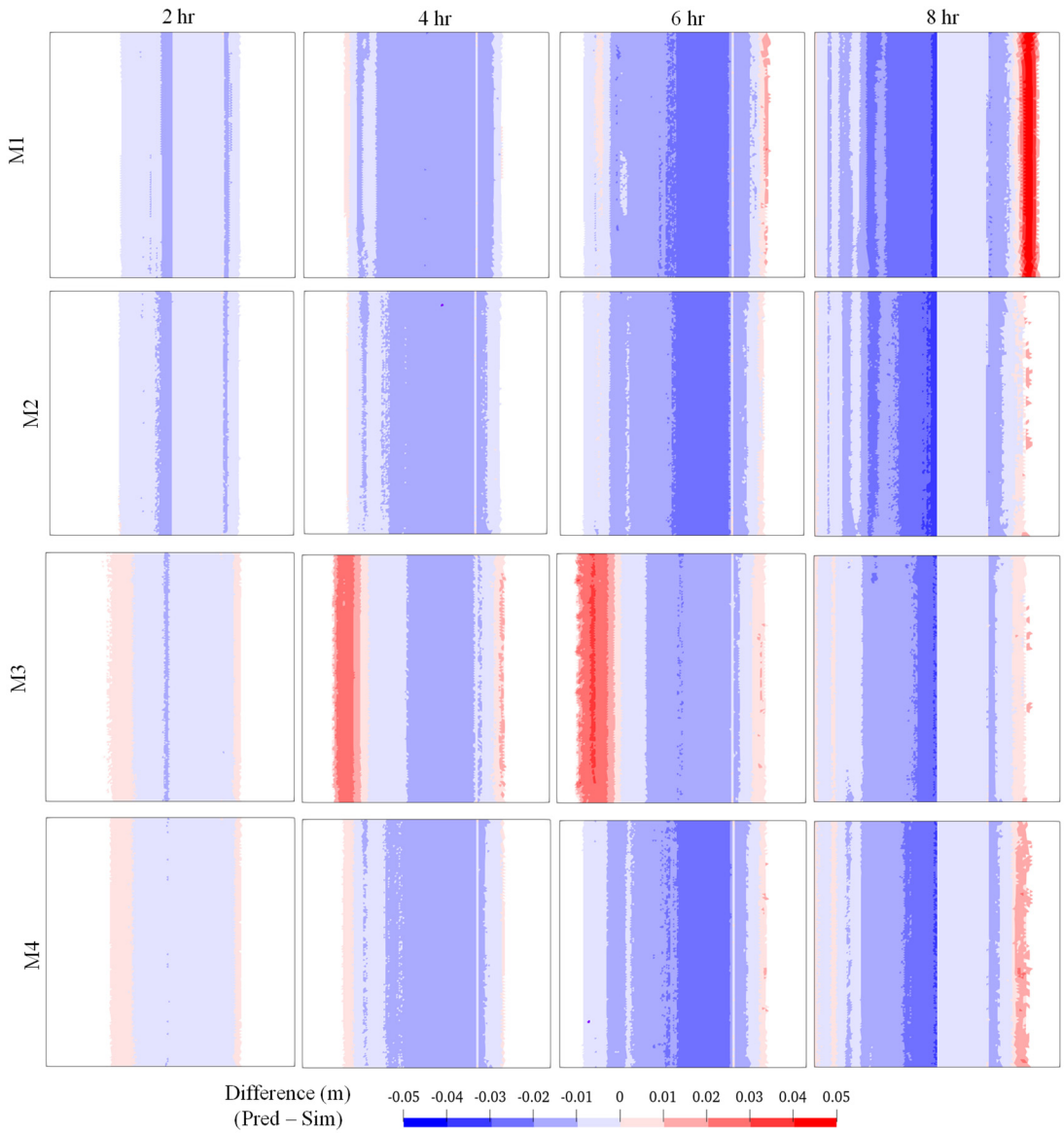


Fig. 3-24 Difference contour comparison of iterated prediction.

representing land use. Although both having the same saturation runoff coefficient, paddy was assumed to be saturated at the beginning of calculation while the other land use takes time before reaching saturation. Therefore, the rate of rise in water depth shifts after reaching saturated rainfall for the other land use while condition for paddy is constantly saturated and mountains unsaturated. The significant bias for M4 is resolved by including land use features. Yet the overall performance does not improve significantly.

To quantitatively evaluate the predicted results, the inundation area and inundation volume of predicted results compared with simulations are illustrated in **Fig. 3-25**. As observed in the contour maps, M1 gradually overestimates inundation area by 8 hr while M3 and M4 show overestimations during the first few hours. M2 resulted in the highest consistency in predicting

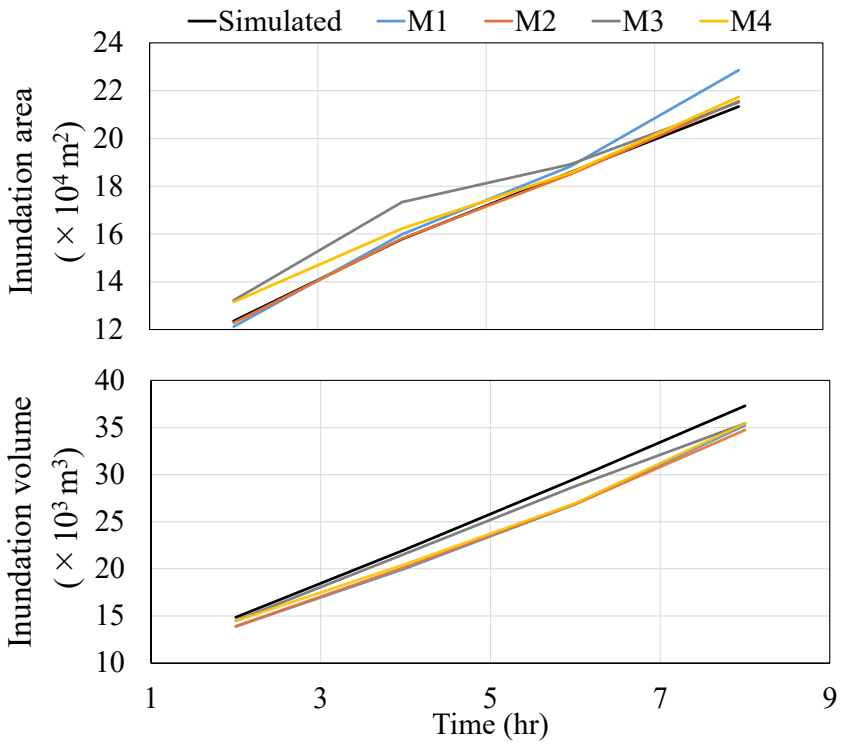


Fig. 3-25 Comparison of calculated and predicted inundation area and inundation volume.

inundation area. On the other hand, the inundation volume for all models gradually biases towards underestimation. Especially considering the fact that most models predicted greater inundation area, the missing inundation volume is rather significant, and it is important to investigate whether this bias is contributed from iteration (i.e. error accumulation over time) or the inherited limitation of model configuration. Therefore, predictions without iteration were further performed and the inundation area and volume comparison is shown in **Fig. 3-26**. It can be observed that the results for inundation area of M2 and M4 are similar to those with iteration, while the significant overestimation in M3 has been refined. When predicted without iteration, M3 and M4 still results in slight overestimation of area during the first 4 hr and the difference between the two models is not negligible. That is, M3 and M4 naturally hold the tendency of overestimating inundation area and M3 is rather sensitive to initial error. In terms of inundation volume, M2 resulted in almost constant underestimation regardless of time even without iteration, which could be inferred as a missing feature in representing other boundary conditions that contribute to the control volume of water other than precipitation. The accuracy of M3 and M4 shows high degree of similarity, but still both slightly underestimating the volume. Based on the contour maps and above evaluation, it can be concluded that M4 resulted in better prediction of the differences in the rise of water depth characteristics induced by land use with the use of one hot encoding binary representations.

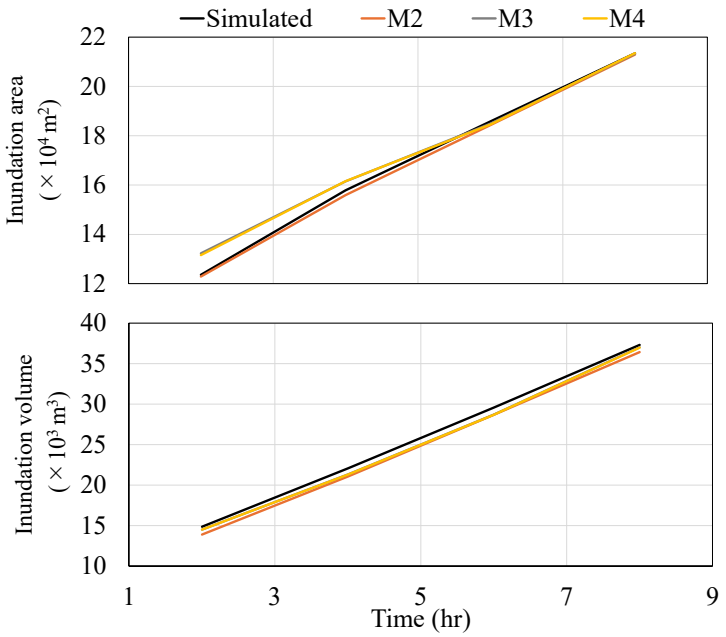


Fig. 3-26 Comparison of calculated and predicted inundation area and volume without iteration.

Yet the tendency of overall underestimation issue of the configurated models is yet to be resolved.

3.4 Chapter summary

This chapter carried out numerical experiments to evaluate the feasibility of using the results from HD model simulations as training data to construct RF models for water depth prediction. With detailed investigations in hyperparameter tuning and feature selections under different terrain conditions, the main conclusions are summarized below.

1. While hyperparameter setting is typically considered as an important factor influencing the accuracy of RF models, the influence of different hyperparameter combinations and split partitioning in this study was rather negligible since the number of training data, as generated from simulation results from HD models, was large enough.
2. The selection of features significantly dominates the accuracy of prediction when targeting untrained test events. Consistent with the physical logistics of HD models, including prior state information of adjacent meshes significantly improved model performance in terms of both classifying inundation area and predicting detailed water depth. Under complex terrain conditions with multiple land use and large variation in gradient between meshes, including features of the elevation of neighbor meshes further improved prediction accuracy since factors indicating water depth together with elevation represent the actual water surface gradient. Features indicating river boundary conditions were also important

in the prediction of water depth in rivers and reflecting the overall control volume within computational domain besides input precipitation. The representation of land use with one hot encoding method reduced the significant error in prediction water depth at land uses where saturation conditions change over time. On the other hand, the overall accuracy between models with and without features representing land use was not significant.

3. The iterated prediction process proposed in this research, where predicted results are reconstructed as the input for the next prediction, solved for spatiotemporal data scarcity over large inundation areas. While reducing computational cost, the accuracy of the model tended to decrease due to the complexity of variations in hydraulic characteristics as the iteration interval becomes larger. With a 2.0-hr iteration interval, the configuration RF model resulted in high performance with overall RMSE restrained within 0.02 m. Having high consistency in predicting water depth in rivers, the accuracy of other land uses varies over time and between models, all having a tendency for underestimation in inundation volume. Since the iterated prediction process does not involve calibrations between iterations, the model was also subjective to error accumulation over time and sensitive to initial error.

Based on the overall high predictive performance demonstrated in this section, the feasibility of the augmentation approach of HD models and RF models and iterated prediction has been confirmed. Therefore, further investigation of the application of the approach under practical terrain conditions was carried out in the next chapter to evaluate the performance of configurated RF models when interaction and exchange of flow between meshes become more complicated.

References

1) Akinsoji, A. H., Adelodun, B., Adeyi, Q., Salau, R. A., Odey, G., & Choi, K. S. (2024). Integrating Machine Learning Models with Comprehensive Data Strategies and Optimization Techniques to Enhance Flood Prediction Accuracy: A Review. *Water Resources Management, 38*(12), 4735–4761. https://doi.org/10.1007/s11269-024-03885-x
2) Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery, 9*(3), e1301. https://doi.org/10.1002/widm.1301
3) Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters*, *31*(14), 2225-2236.
4) Bernard, S., Heutte, L., & Adam, S. (2009, June). Influence of hyperparameters on random forest accuracy. *In International workshop on multiple classifier systems* (pp. 171-180). Berlin, Heidelberg: Springer Berlin Heidelberg.
5) Goldstein, B. A., Polley, E. C., & Briggs, F. B. (2011). Random forests for genetic association studies. *Statistical applications in genetics and molecular biology*, *10*(1).
6) Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing, 415*, 295-316.
7) Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.
8) Seger, C. (2018). An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing.

CHAPTER 4 MODEL APPLICATION OVER PRACTICAL RIVER BASIN

4.1 Study area and background

This chapter investigates the feasibility of applying the proposed augmentation approach for spatiotemporal flood prediction over large practical river basins where land use and terrain conditions are much complicated. The study area for this research is the Karube river basin of Soja City, Okayama Prefecture, and depicted in **Fig. 4-1** are the overview of this region and elevation contour. With an average gradient of approximately 1/1300, the Karube river diverges from the left bank of the Takahashi river, which is a Class A river in Okayama, through Tatai weir, branches at Imoji diversion, flows through Soja City between large areas of paddy and eventually converges back to the Takahashi river. As shown in **Fig. 4-1(b)**, the elevation of downstream areas south of the Kiyone station is low laying with paddies covering about 11% of the area. Since the area is enclosed by the left embankment of the Takahashi river and Karube mountain, pluvial floods have occurred frequently due to difficulties of drainage,

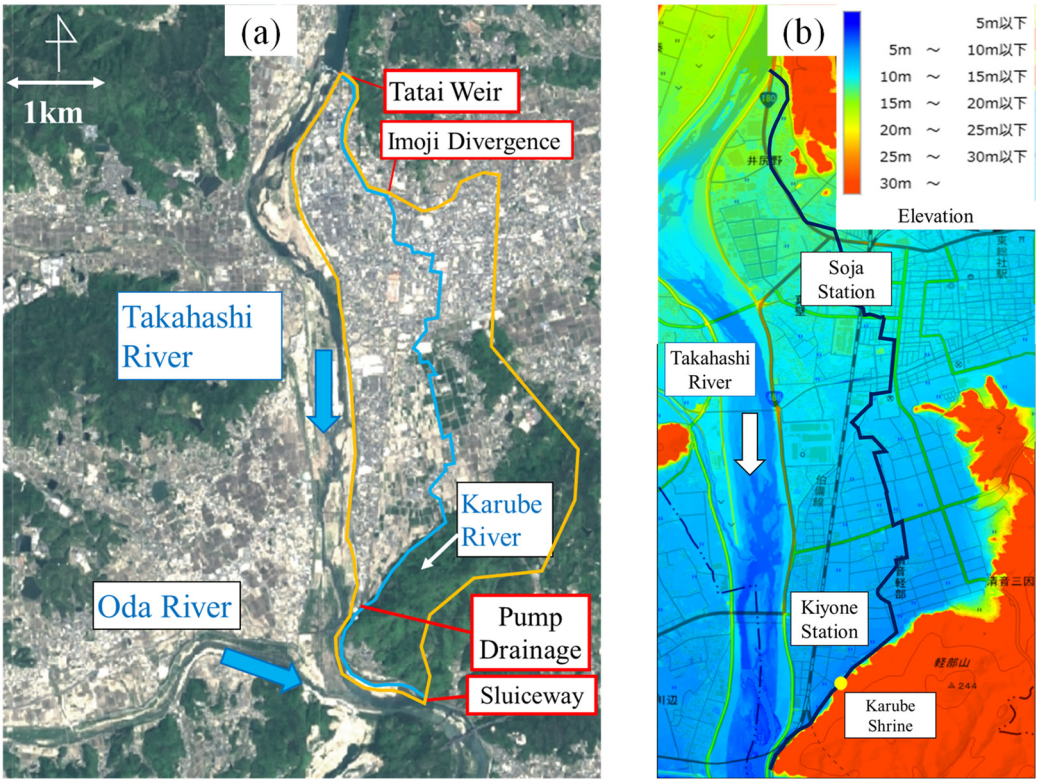


Fig. 4-1 Study area as of (a) overview and (b) elevation contour.

especially over the downstream area. Summarized in **Table 4-1** are the rainfall magnitudes of peak intensity and cumulative rainfall observed at the Soja City observatory that induced significant flood damages and three representative precipitation events in recent years[1),2),3)]. The detailed temporal distribution of the recent rainfall events is illustrated in **Fig. 4-2**[3)]. And illustrated in **Fig. 4-3** are the inundation records for rainfall events of (a) 1972, 1976 and 1985[1)] and (b) 2018 and 2020. From the inundation record maps, it can be observed that even with the facility of pumps drainage, this area is still vulnerable to floods.

To better prepare residents from flood disasters, Soja City has issued hazard maps under different rainfall scales in recent years and illustrated in **Fig. 4-4(a)** is the hazard map under the L1 scale, the planned scale[5)]. The L1 scale refers to heavy rainfalls with return period of approximately 100 years to 150 years, with a maximum two-day cumulative rainfall of 248 mm at this region. Under rainfall events of such scales, the entire downstream area is expected to experience inundation depth of 2.0 m to 5.0 m. **Fig. 4-4(b)** is an aerial photograph of the area enclosed with the red dashed line on the left during the heavy rainfall event of 2018, which was approximately a 150-year event[5)]. It can be confirmed from the photo that a large area was flooded, and a majority of paddies were submerged, which is consistent with the hazard map. On the other hand, the hazard map fails to provide information on how the flood propagates over space and time, and such information is considered important in situations when evacuation or flood damage mitigation countermeasures are necessary. Furthermore, since the hazard maps are designed to depict the potential damage of one single event, inundation information under other rainfall events, even those with small scales, is not provided. For example, **Fig. 4-5** illustrates a comparison of approximately the same location near the Karube Shrine (location shown in **Fig. 4-1(b)**) under normal condition and inundated condition due to the July 2020 rainfall event[6)]. It can be observed overflow from the Karube river occurred and the floods propagated through paddies towards the residential area. Although such floods with low levels of inundation, which are also referred to as nuisance flooding, do not directly lead to significant property loss nor mass casualties, they could disrupt routine activities, interrupt

Table 4-1 Summaries of rainfall information inducing major floods in the Soja city.

Dates	Maximum magnitude observed	
	Peak intensity (mm/hr)	Cumulative rainfall (mm)
1972/07/09 – 07/13[1)]	25.0	291.5
1976/09/08 – 08/13[2)]	23.0	452.0
1985/06/22 – 06/30[2)]	20.0	422.5
2018/07/04 – 07/07[3)]	27.0	287.0
2020/07/06 – 07/08[3)]	15.0	123.0
2021/08/13 – 08/15[3)]	16.0	188.0

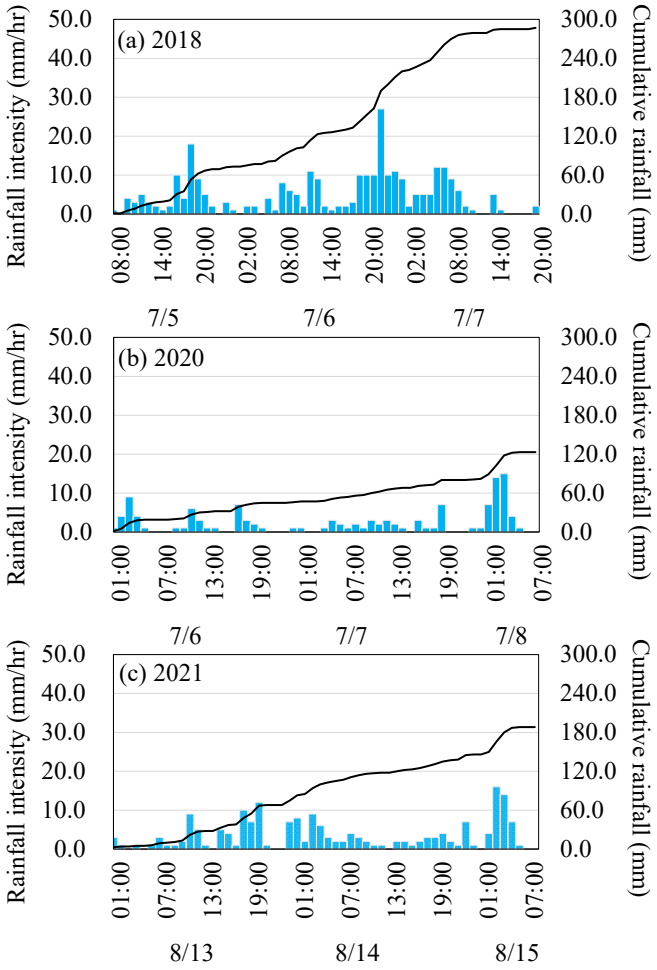


Fig. 4-2 Intensity and cumulative rainfall information of historical events of (a) 2018, (b) 2020, and (c) 2021[3)].

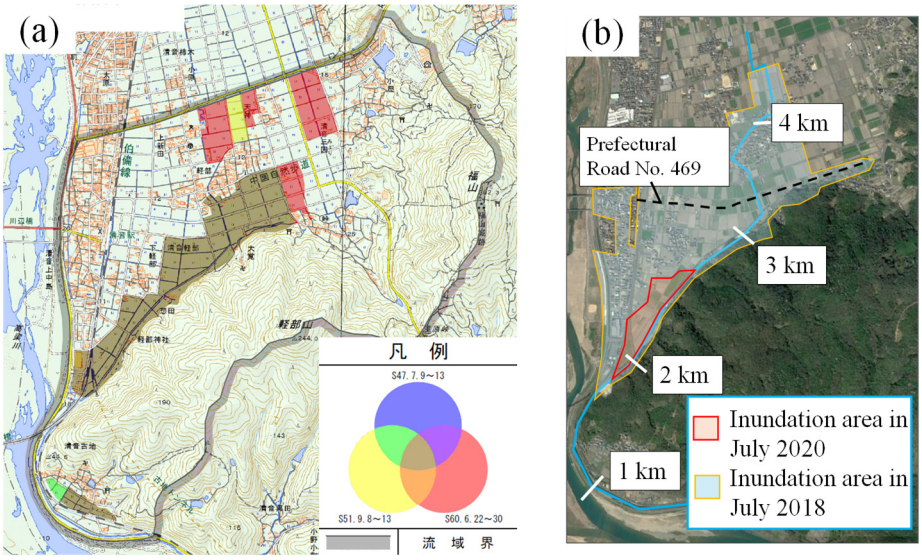


Fig. 4-3 Historical inundation records.

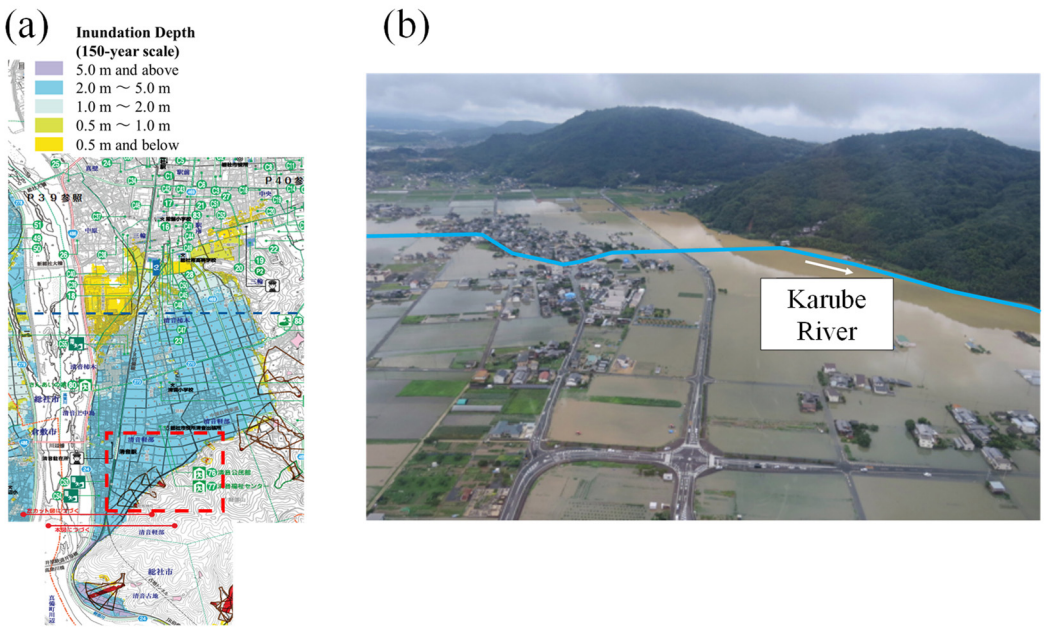


Fig. 4-4 (a) Hazard map for L1-scale (150-yr return period) and (b) aerial photo of 2018 inundation (downstream region)[5].

Fig. 4-5 Comparison of (a) normal condition and (b) inundation situation under July 2020 rainfall at the same location (around Karube Shine)[6].

transportation and may induce health risks and safety threats[7]. Furthermore, at this specific location where the downstream consists large area of paddies, long-lasting inundation over paddies, especially those over 24 hr, could lead to damage on agriculture[8]. As a result, a fast and accurate flood prediction method is still absent for future policy-making applications in correspondence to the upcoming frequent floods.

4.2 Simulation conditions and flood analysis

4.2.1 Construction of HD model and boundary conditions

In this study, the configuration of HD model covers the area enclosed in the yellow line shown in **Fig. 4-6(a)**, where the entire region including inundation area and rivers were divided into unstructured triangular meshes. As illustrated in **Fig. 4-6(b)**, the delineation of land use was represented by one-dimensional grid lines and other areas such as the Karube river and inside of paddies were reflected two-dimensionally such that detailed microtopography and the shape of paddy could be incorporated. The size of meshes is 25 m for mountain area, 2.5 m and 5.0 m for the upstream and downstream of the Karube river, respectively, and 10 m for the rest of the region, resulting in a total of 249,266 meshes. The elevation was interpolated using the 0.5 m resolution ALB survey data conducted by the Forestry Agency in 2018 for the downstream area, and the rest of the area using 5 m DEM data obtained from Geospatial Information Authority of Japan[9]. The riverbed elevation of the downstream Karube river was

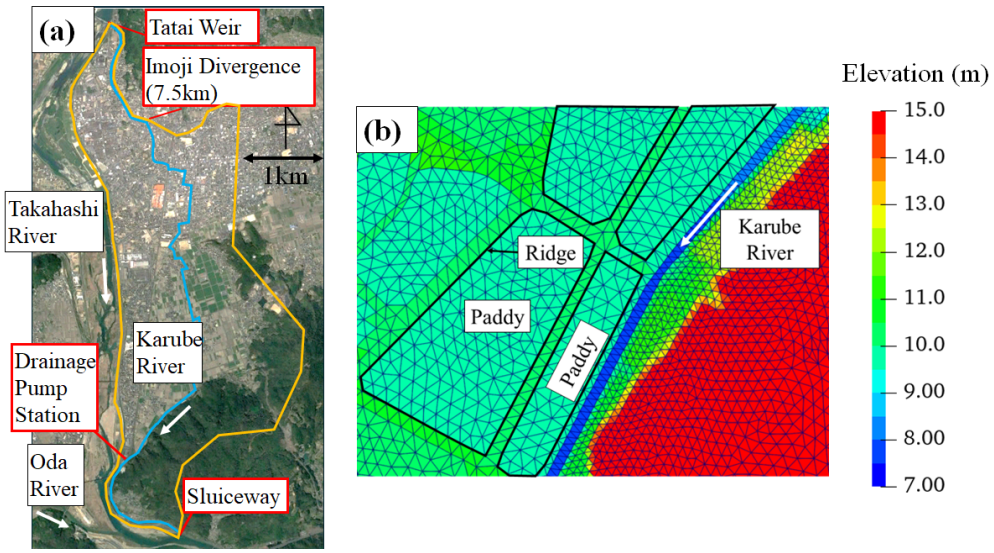
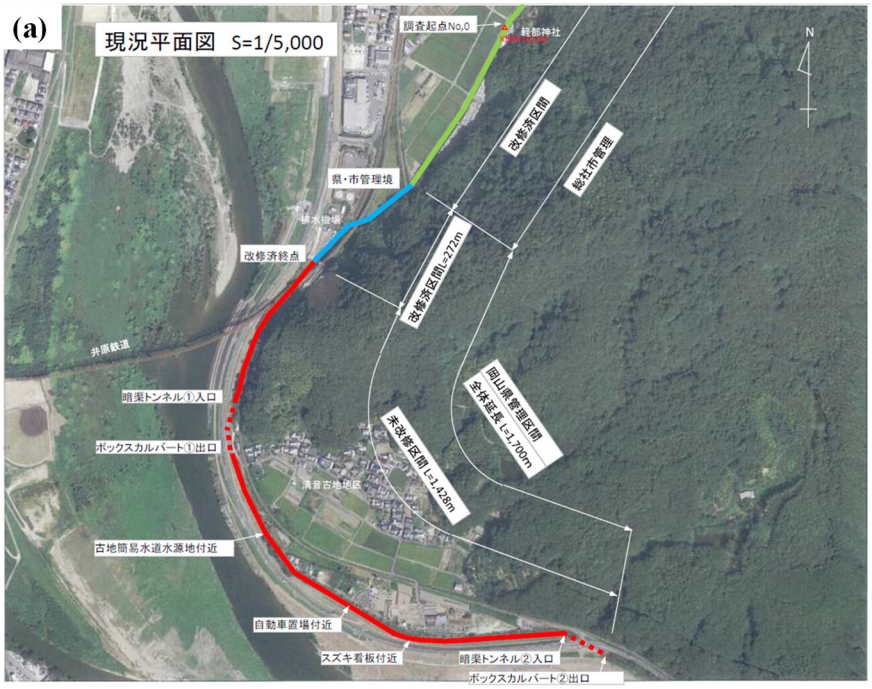


Fig. 4-6 Illustration of (a) computational domain for HD model and (b) unstructured triangular meshes and the delineation between land uses.

interpolated linearly using the longitudinal distance from the cross-section profile information provided by the Soja City[10]. **Fig. 4-7** illustrates (a) the area being investigated and managed and (b) detailed riverbed elevation. For the upstream area where information of riverbed elevation was not available, the topographical gradient of the inundation area was first



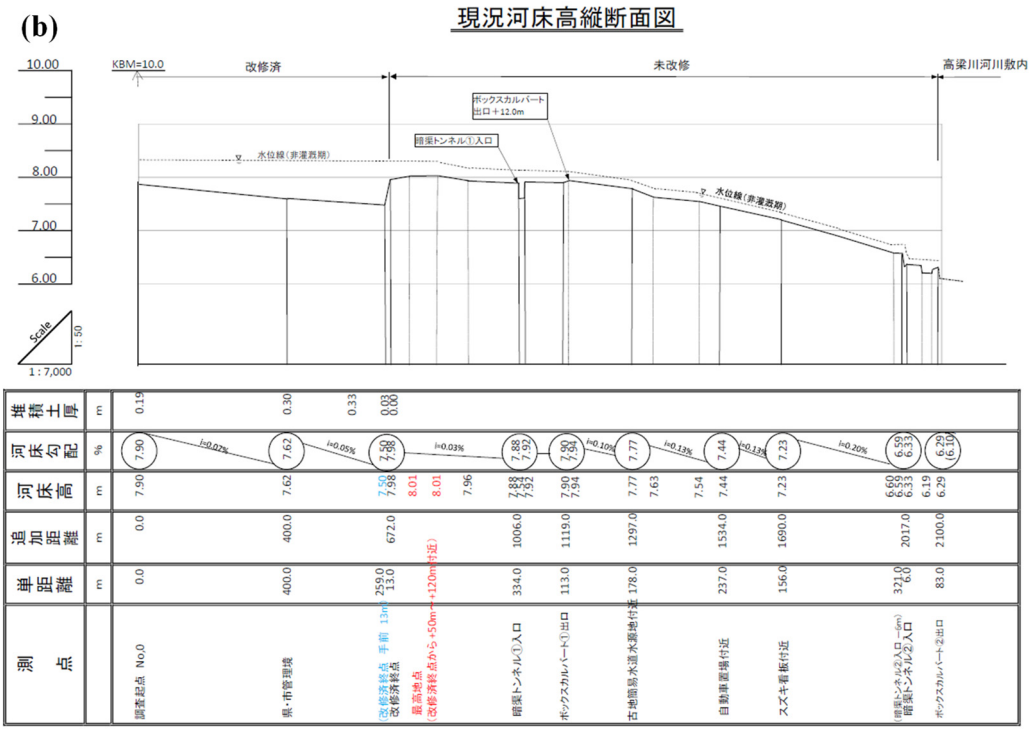


Fig. 4-7 (a) Investigation area of the Soja city and (b) longitudinal riverbed elevation.

calculated based on the ALB data and DEM data, and the riverbed elevation was obtained with the assumption that the longitudinal gradient of the Karube river is approximately the same as the topographical gradient. Due to the limitation of DEM data resolution, the elevation within each paddy polygon was refined to be the average value to avoid unnecessary disturbance of flux (shown in **Fig. 4-6(b)**). The interpolated elevation contour is illustrated in **Fig. 4-8**.

As stated in Chapter 2, the representation of land use in HD models was incorporated as effective rainfall, and the parameters adopted in this study are summarized in **Table 4-2**. Detailed model configuration and validation over historical rainfall events were performed by previous literature[11], and this study adopted the original model with the following changes for simplicity. The drainage from paddies to the Karube river was simulated through a simplified model without the representation of narrow drainage ditches as shown in **Fig. 4-9(a)**. The inflow mesh, which replaced the location of actual weir, was determined to be the centroid of each delineated paddy polygon with a paired outflow mesh set in the Karube river channel. The water depth of each inflow mesh was then used to calculate drainage discharge, which

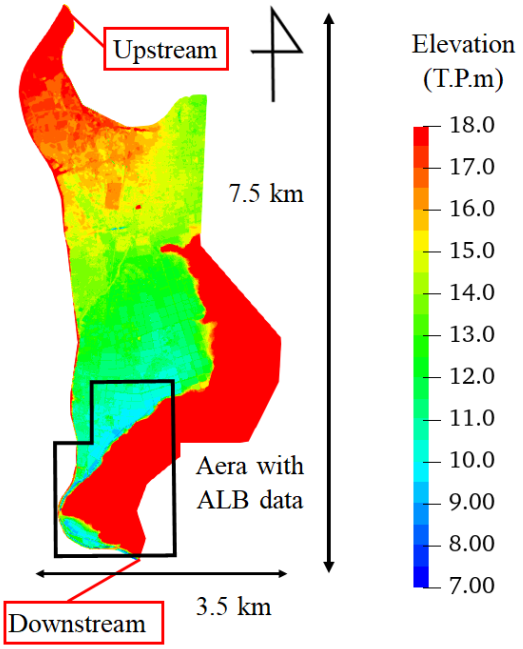


Fig. 4-8 Interpolated elevation contour.

Table 4-2 Parameter setting with respect to land use.

Land use	Manning's coefficient	Runoff coefficient		Saturated rainfall (mm)
		Unsaturated	Saturated	
Mountain	0.6	0.3	0.9	150
Paddy	0.02	0.5	0.9	55
Karube	0.015	1.0	1.0	-
Miscellaneous	0.025	0.6	0.9	55

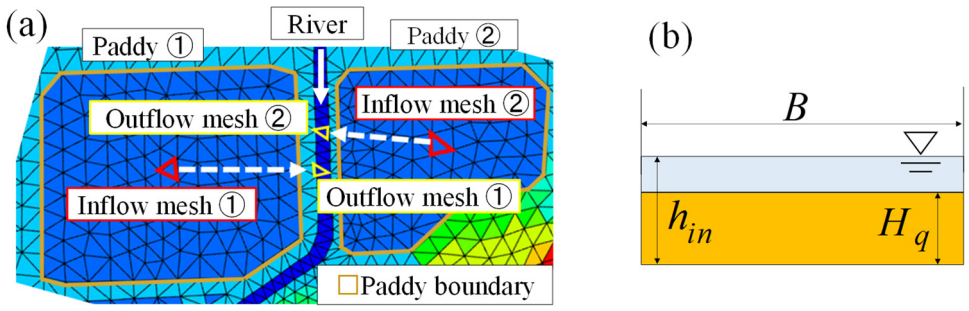


Fig. 4-9 Schematic of (a) drainage model from paddy and (b) weir flow condition.

was then applied as a source term in the continuity equation for each inflow and outflow mesh. The temporal variation of water depth of the inflow and outflow meshes are given below:

$$\text{Inflow mesh:} \quad h_{in}{}^{n+1} = h_{in}{}^{n} - \frac{Q_{in}{}^{n}}{\Delta S_{in}} \cdot \Delta t \tag{4.1}$$

$$\text{Outflow mesh:} \quad h_{out}{}^{n+1} = h_{out}{}^{n} - \frac{Q_{out}{}^{n}}{\Delta S_{out}} \cdot \Delta t \tag{4.2}$$

where h_{in} (m) and h_{out} (m) represent the water depth of the inflow and outflow meshes respectively, Qin (m^3/s) and $Qout$ (m^3/s) are the discharge rate for inflow and outflow meshes, Δt (s) is the time interval for each step of calculation, and ΔS_{in} (m^2) and ΔS_{in} (m^2) represent the mesh area for inflow and outflow meshes respectively. The inflow discharge was assumed to be normal weir drainage over the entire period, and the corresponding discharge rate was calculated based on weir flow equation[12].

$$Q_{in}^{n} = C_1 B (h_{in}^{n} - Hp)^{3/2} \cdot R_{in} \tag{4.3}$$

where B (m) is the width of weir, Hp (m) is the height of weir, C_1 ($m^{1/2}/s$) is the discharge coefficient, which is calculated based on the following equation.

$$C_1 = 1.785 + \frac{0.00295}{h_{in}{}^{n}} + 0.237 \frac{h_{in}{}^{n}}{H_p} \tag{4.4}$$

where Rin is the number of weirs for a unit paddy area. Based on field observations over the Karube region, parameters were set to be $B = 0.3$ (m), $Hp = 0.15$ (m), $Rin = Ain/1250$ with Ain being the area of each paddy. Since the hydrodynamic calculations are not performed with detailed representation of drainage ditches, it is necessary to consider the time difference between discharging from inflow meshes and reaching outflow meshes. The simplified modeling of discharge rate is represented by the following equations.

$$Q_{in}{}^{n} = \frac{V^n}{\Delta T} \tag{4.5}$$

$$V^{n+1} = V^n + (Q_{in}{}^{n} - Q_{out}{}^{n}) \cdot \Delta t \tag{4.6}$$

$$\Delta T = \frac{L}{V_{ch}} \tag{4.7}$$

Here, ΔT (s) is the arrival time draining from each paddy to river channel, V^n (m^3/s) indicates the amount of water discharged from the inflow mesh between time $t = t^n - \Delta T$ (s) and $t = t^n$ (s). The determination of arrival time for each paddy is specified by L (m), the distance between the centroid of inflow mesh and outflow mesh, and V_{ch} (m/s), the averaged velocity obtained from uniform flow condition in rectangular cross sections by assuming a water depth within drainage ditches.

The input of rainfall was applied without spatial variation unless specified otherwise, and intensities were obtained either from hypothetically designed distributions, or from observation records at the Soja city. Runup calculation was first performed under constant normal discharge and normal flow boundary conditions without rainfall until flow stables. The obtained water depth and velocity of river flow was then used as the initial condition, and boundary conditions for both the upstream (Tatai Weir) and the downstream sluiceway were assumed to be closed throughout the calculation period and were therefore set to be wall condition. The initial depth setting for all other land uses was 0.0 m, and paddy was assumed to be already in saturation at the beginning of calculation. Pump drainage from the drainage station to the Takahashi River was also omitted in the calculation, contributing to a more significant pooling condition. Simulation of HD model was performed for 20 hr with 0.1 second calculation interval per step and output interval of 10 minutes.

4.2.2 Inundation characteristics based on flood analysis

Due to the differences in the settings of simulation conditions, especially the omission of drainage pumps at the downstream, comprehensive flood analysis was first performed over the Karube region to capture the overall characteristics of spatiotemporal variation of flood propagation. Two actual rainfall events selected here are the 2018 rainfall event, corresponding to a 60-hr rainfall from 8:00 of July 5, through 20:00 of July 7, and the 2021 rainfall event, a 55-hr precipitation from 01:00 of August 13, to 07:00 August 15. The hourly intensity and cumulative rainfall records obtained from the Soja observation station are previously illustrated in **Fig. 4-2**. The simulated results of HD model in terms of water depth contour maps for both events are illustrated in **Fig. 4-10** and **Fig. 4-11** respectively, with actual timestamps shifted such that the beginning of each rainfall event corresponds to 0 hr, and depth shallower than 0.01 m was cutoff for simplicity and clarity of illustration. For both rainfall events, precipitation accumulates evenly over paddy areas during the first 10 hr. The water level in the Karube river increases significantly due to closed downstream conditions, and

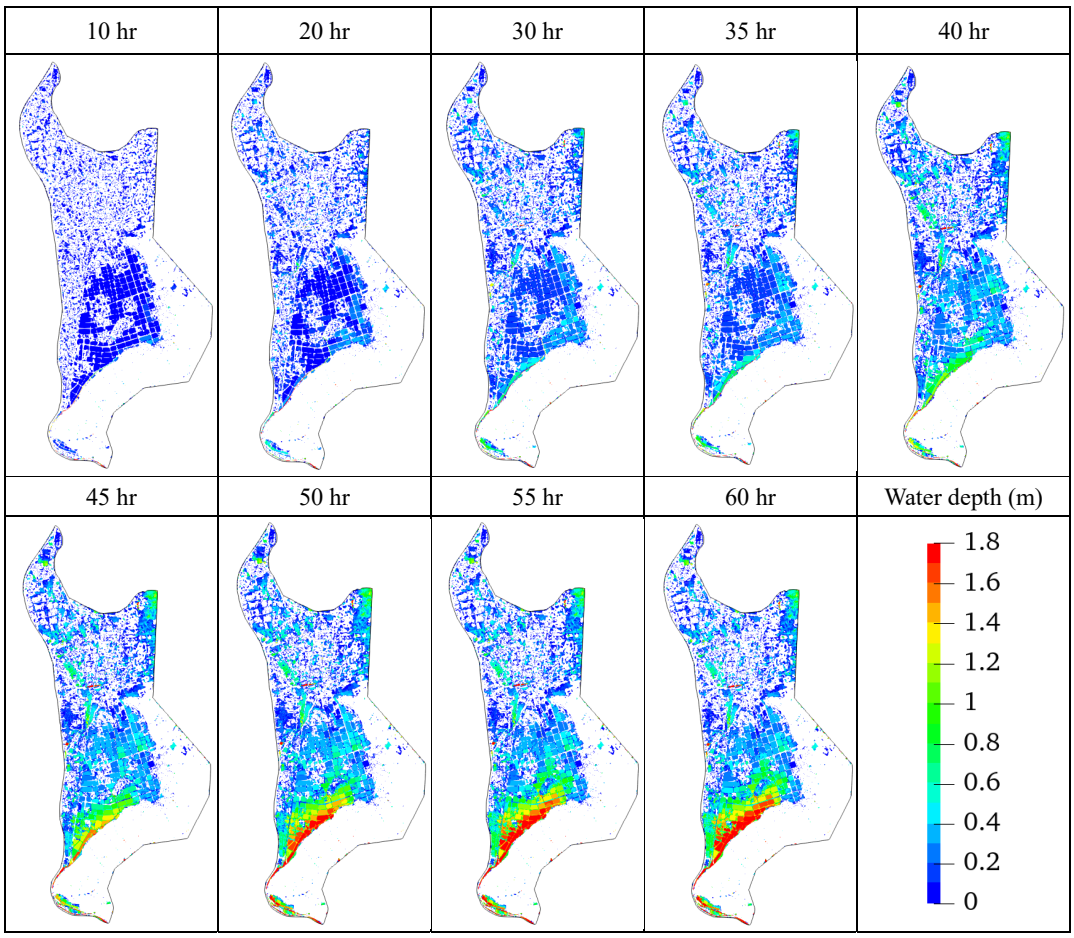


Fig. 4-10 Simulated water depth contour for 2018 rainfall event.

overflow from the Karube river is observed starting approximately 15 hr and 22 hr for the two rainfall events respectively. Influenced by the overflow, the water depth in paddies located along the Karube river also accumulates significantly, and full capacity was gradually reached. The paddies located near mountain areas also suffer severe inundation compared to the ones located upstream considering the surface flow accumulated from mountains. When the depth in paddies exceeds 0.3 m, which is the height of paddy ridges, overflow from paddies also occurs and the entire downstream suffers sever inundation. **Fig. 4-12(a)** and **Fig. 4-12(b)** depict the simulated water depth at the final step of both event focusing specifically at the downstream. Due to large magnitude of cumulative rainfall, the entire downstream regardless of land use is significantly inundated with depths greater than 1.0 m for 2018 rainfall. In comparison, the inundation area for 2021 rainfall is rather constrained by the railway and prefecture road because of locally elevated topography, and the resulting water depth is also shallower. The aerial photo of the downstream area with historical inundation records is shown in **Fig. 4-12(c)**. The simulated flood of 2018 rainfall propagates wider compared to observations since the function for pump drainage to the Takahashi river was omitted in

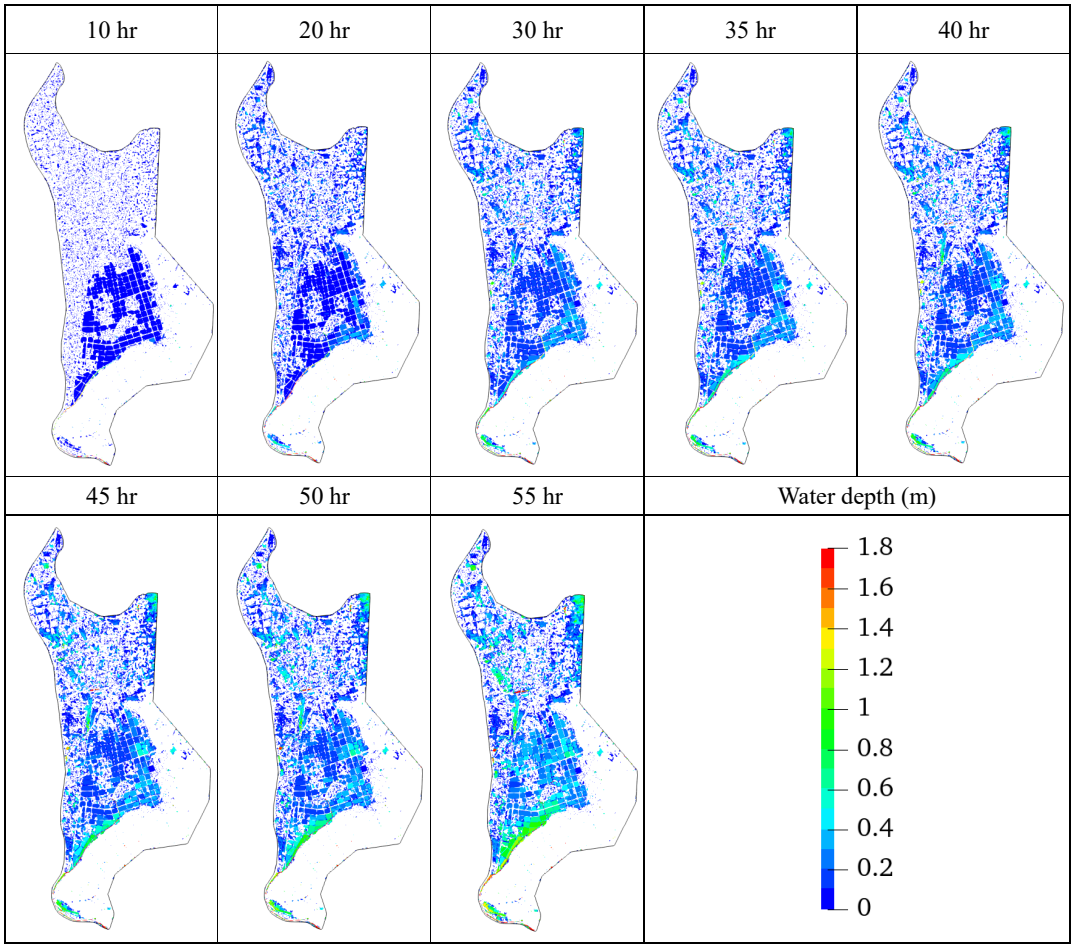


Fig. 4-11 Simulated water depth contour for 2021 rainfall event.

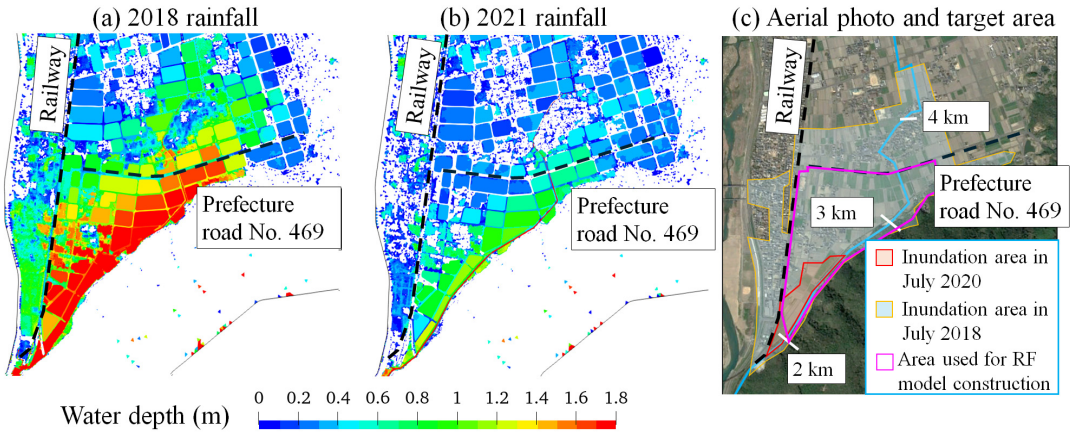


Fig. 4-12 Simulated water depth contour at final step of (a) 2018 rainfall, (b) 2021 rainfall and (c) aerial photo of the inundation records and target area for RF model.

simulation. Another record shown here is the rainfall of 2020, which has a magnitude of cumulative rainfall of 123 mm, and it can be observed that the inundation area also concentrates along the downstream of Karube river.

4.3 Configuration and performance evaluation of RF model

4.3.1 Details of RF model settings

While RF models are generally suitable for large dataset, it is rather impossible to configurate an RF model to predict individual water depths at each mesh spatially and temporarily due to the large number of meshes and long prediction periods. Therefore the target area of RF models must be carefully selected to reduce the computational cost while taking into account the potential for practical applications. Based on the inundation characteristics observed from both flood analysis and historical records, the target area for RF configuration focused on focused only on the downstream area as outlined in pink of **Fig. 4-12(c)**. The outline delineates the railway, prefecture road and mountain such that the most frequently inundated areas are covered as the target for prediction. The feasibility of the model to be applied under extreme floods such as the one induced by 2018 rainfall where inundation areas exceed the outline boundary is also investigated as an objective for this research. The corresponding total number of meshes for the target area of RF model was 18,687, with approximately 3% meshes being the Karube river, 15% being mountain areas, 32.5 % being paddies. The rest of the areas, denoted by miscellaneous land use, include roads and residential areas that are most vulnerable to flood disasters. The spatial distribution for the land use delineation at the target area is depicted in **Fig. 4-13.**

The features and hyperparameter settings for model configuration are summarized in **Table 4-3**. The models configured in this section adopted M2, which included the elevation of

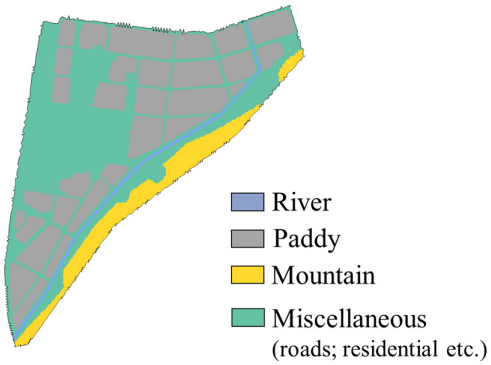


Fig. 4-13 Land use delineation of the target area.

Table 4-3 Features and hyperparameters for RF models.

Models	Features	Hyperparameters
M2	n, x, y, z, $step$, $rain_t$, $rain_{cum}$, $h_{t-1,\,i}$, $h_{t-1,\,ij}$, z_i, z_{ij}	*n_estimator* = 100,
M3	M2 + Q_{in}, Q_{out}	*max_depth* = 20,
M4	M3 + land use (binary)	*max_features* = 'sqrt'

adjacent meshes, since discussions in the previous chapter proved that including such features increased the prediction accuracy especially at locations where terrain gradient varies significantly. M3 and M4 were also constructed to further investigate the influence of river boundary conditions and factors indicating land use. As for the hyperparameter settings, the number of trees was set at 100, maximum depth at 20 and "*sqrt*" for the maximum features at each split for all three models. The training dataset consisted of the same 5 patterns of constant rainfall, 10 mm/hr, 15 mm/hr, 20 mm/hr, 30 mm/hr, and 50 mm/hr and constant rainfall with 25 mm/hr was selected as the target event for prediction. Iterated predictions were carried out with 2.0-hr interval, and evaluations were performed against simulation results from HD models.

4.3.2 Feature importance comparison of RF models

A comparison of feature importance for the RF model models is illustrated in **Fig. 4-14**. Similar to the observations from previous chapter, the static inherited properties resulted in a lower importance value compared to dynamically varying influence of prior depth. Within static properties, the elevation of each mesh results in the highest importance for all three models. The feature importance for inflow weights higher than outflow. The selection of target area focused on the downstream end of the river basin, and the variation of inflow conditions is more complicated as dominated by the inundation condition of upstream paddies, where the outflow discharges, in comparison, is dominated mostly by the gradual rise in water level due to be closed boundary condition. This complexity of temporal water level variation in river also lead to a relatively high importance in channel compared to the other features of land use. As the terrain becomes complicated, the bias of feature importance due to the numbering of mesh orientation also diminishes with similar importance values among adjacent meshes. On the other hand, the importance for the elevation of adjacent meshes is rather negligible. As concluded in the previous chapter, features of neighbor elevation contribute significantly to the accuracy when predicting areas with high gradients. The gradient of mountains in practical terrain settings is much steeper than the settings in the numerical experiment and almost no inundation occurred over mountain areas as shown previously by simulated contour maps in

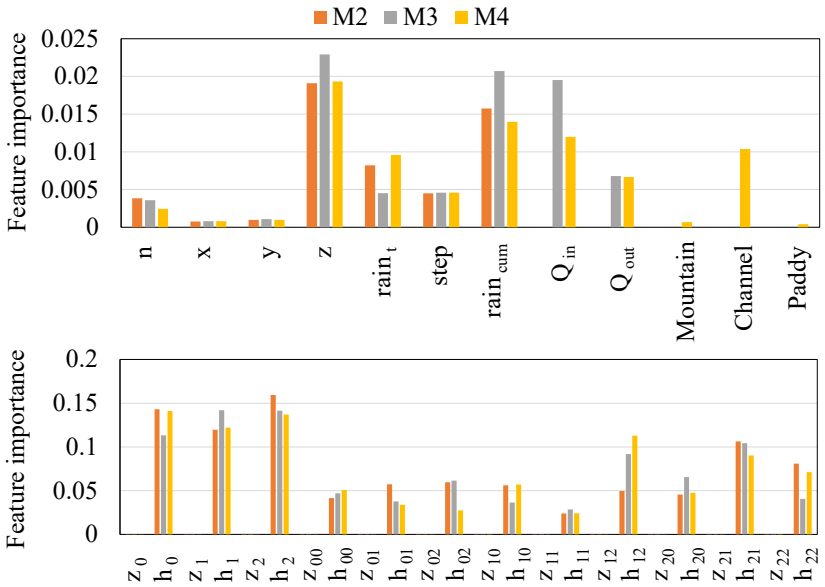


Fig. 4-14 Feature importance of configurated RF models.

Fig. 4-10. Because of the limited mountain meshes classified as mountain, the resulting importance of this feature is low. However, including this feature better improves the interpretability of the RF model since elevation together with water depth represents the actual water surface gradient between meshes, which is consistent with the logistics of HD models. Since one of the objectives for this study is to balance the interpretability and prediction capability of the RF model, the features representing the elevation of adjacent meshes are reserved as an important factor in model configuration.

4.3.3 Simulation results for target rainfall event

The simulated water depth contour of the entire Karube river basin under 25 mm/hr of rainfall is depicted in **Fig. 4-15** with a 0.01 m threshold for inundation. During the first 6 hr, water depth gradually accumulates within paddies with some ponding area observed where the terrain is locally depressed. Since the downstream sluiceway of the Karube river was closed, the water level in the river channel increase as precipitation progresses. As the height of paddy ridge was set to be 0.3 m, water depth exceeding this threshold would lead to overflow from paddy. The storage of certain paddies reaches full capacity at approximately 6.5 hr and 7 hr, and overflow from the Karube river also occurred at the downstream regions as shown by the enlarged figure in **Fig. 4-16** with the area for RF model construction outlined in black. The double overflow contributed to the significant spread of flood over roads, with large ponding observed at downstream areas. The paddies located close to mountain areas also resulted in higher water depths compared to the others located more distantly due to the natural flow

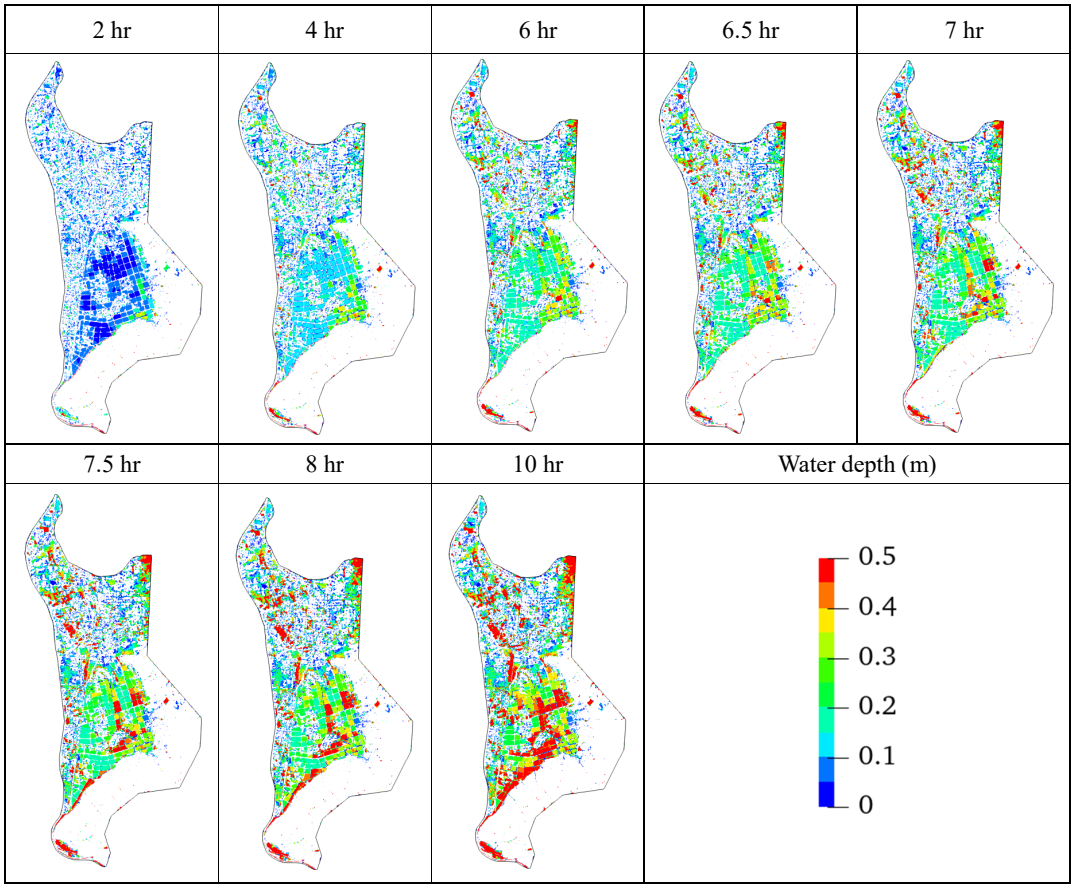


Fig. 4-15 Temporal variation of simulated water depth contour (25 mm/hr).

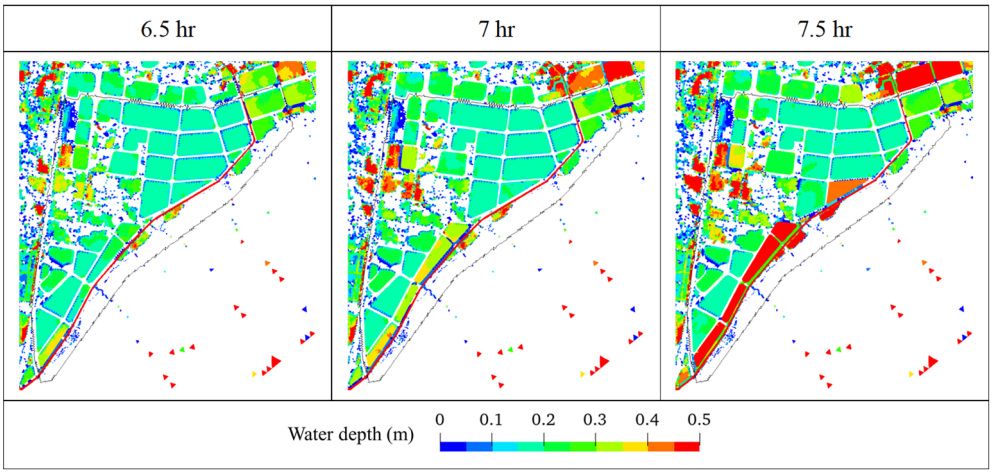


Fig. 4-16 Detailed water depth contour at the downstream target area (25 mm/hr).

accumulation from mountains. By the end of 8 hr, the majority of paddy reached full storage capacity and overflow from both paddies and the Karube river led to significant flooding with depth greater than 0.5 m at 10 hr. Overall, the simulation results are consistent with the observed historical records of this region that inundation occurs mostly at the downstream area and main reasons contributing to inundation are overflows from paddy and the Karube river.

4.3.4 Evaluation of prediction performance

The prediction performance for the RF models was tested against rainfall of 25 mm/hr. Evaluations were first conducted without iteration (i.e. using simulated results as the input of prior state) to observe the temporal variation of accuracy, and evaluations of R^2 and overall RMSE averaged for all meshes are depicted in **Fig. 4-17**. With a 2.0-hr interval for prior state of neighbor information, the first prediction using no initial water depth over the river basin corresponds a prediction at 2 hr. The general trend of temporal variation of the three models is similar with the lowest performance is observed between 7 hr and 10 hr. M3 and M4 slightly outperforming M2 during 9 hr to 15 hr and the performance for M3 and M4 is almost consistent over the entire period. All having R^2 exceeding 0.92 for all predictions, the three models are considered to have high performances. Yet the RMSE indicates certain error in making predictions of detailed water depths. The results during the first 8 hr for the three models highly coincide with each other, and all three models result in a significant increase in overall error up to 0.1 m from 7 hr to 8 hr. This observation could be explained by the sudden change in the inundation characteristics induced by overflow. The main contributing factor of water depth accumulation before 7 hr is mostly rainfall. The characteristics of such depth rise are easy to capture and previous state of 2 hr ago is relatively sufficient to reflect the change. Yet as overflow occurs, the dominating factor for the rise of water becomes the flux exchange between meshes and the incoming inflow from upstream region contributes significantly to the control volume of the target area. Under complicated conditions where boundary conditions and flux exchange vary significantly by the count of seconds, prior information of 2.0-hr interval fails to represent the complexity, leading to a sudden increase in RMSE even provided that input values of prior information are accurate. This also suggests that considering future practical conditions where the temporal change of both rainfall and inundation conditions is significant, the iteration period of the proposed iteration prediction approach

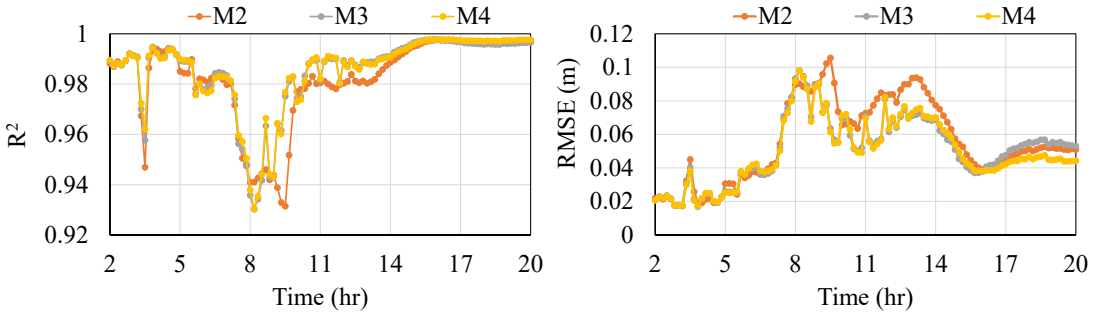


Fig. 4-17 Temporal variation of model performance evaluation without iteration.

should be shortened to increase model accuracy even though computation cost increases. Consistent with the increase in R^2, the error of prediction stables and gradually decreases as time progresses and the performance for M3 and M4 slightly is better compared to M2. In summary, while the overall R^2 values are higher than 0.9 for the entire period, the corresponding RMSE increases to a maximum exceeding 0.1 m, again suggesting that multiple evaluation metrics should be considered to sufficiently determine the performance of model prediction. Overall, M3 and M4 outperform M2, yet all models resulted in similar temporal trend in error variation. Even provided with accurate information of prior state of neighbor meshes, the accuracy of models is significantly compromised around the period when overflow occurs. This error is expected to further accumulate in the practical application of iterated prediction since the results for each prediction are not calibrated. As a result, the following performance evaluation regarding the iterated prediction process only investigated the first few iterations and carried out contour comparisons to observe the differences between land use and between models..

First illustrated in **Fig. 4-18** are the temporal variation of the overall performance metrics for the iterated predictions using a threshold of 0.01 m. The R^2 gradually decreases over time, which is expected to be a result of error accumulation due to iteration. Having values greater than 0.9, M3 and M4 slightly outperforming M2. On the other hand, all three models resulted in the same trend of holding high recall yet low precision and accuracy, which then lead to a

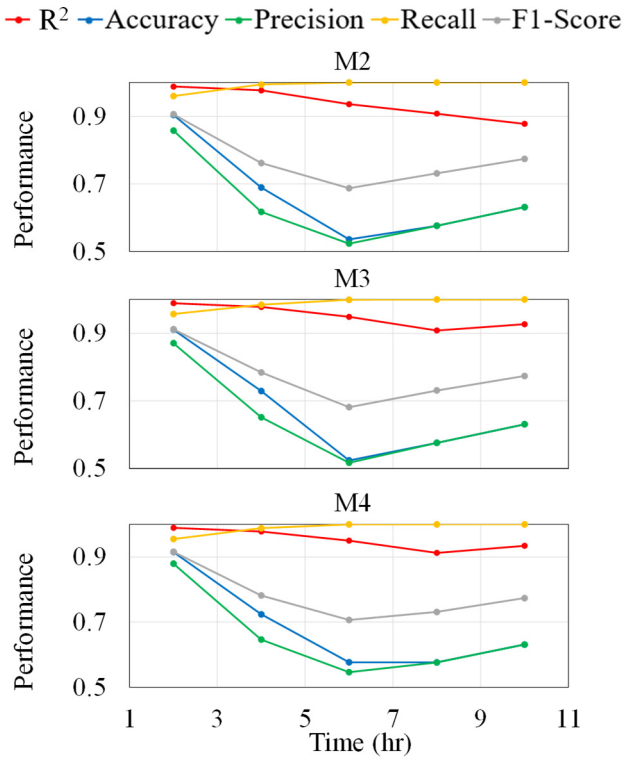


Fig. 4-18 Temporal variation of overall performance metrics of iterated predictions.

relatively compromised F1-score, which comprehensively suggests that the models have a tendency for false positive. That is, predicting larger areas of inundation than actual situation. **Fig. 4-19** depicts the predicted water depth contour of the three RF models compared with simulation results over time, again shown with the same threshold. Consistent with the low precision evaluation, it can be observed that the delineation between land uses is blurred out

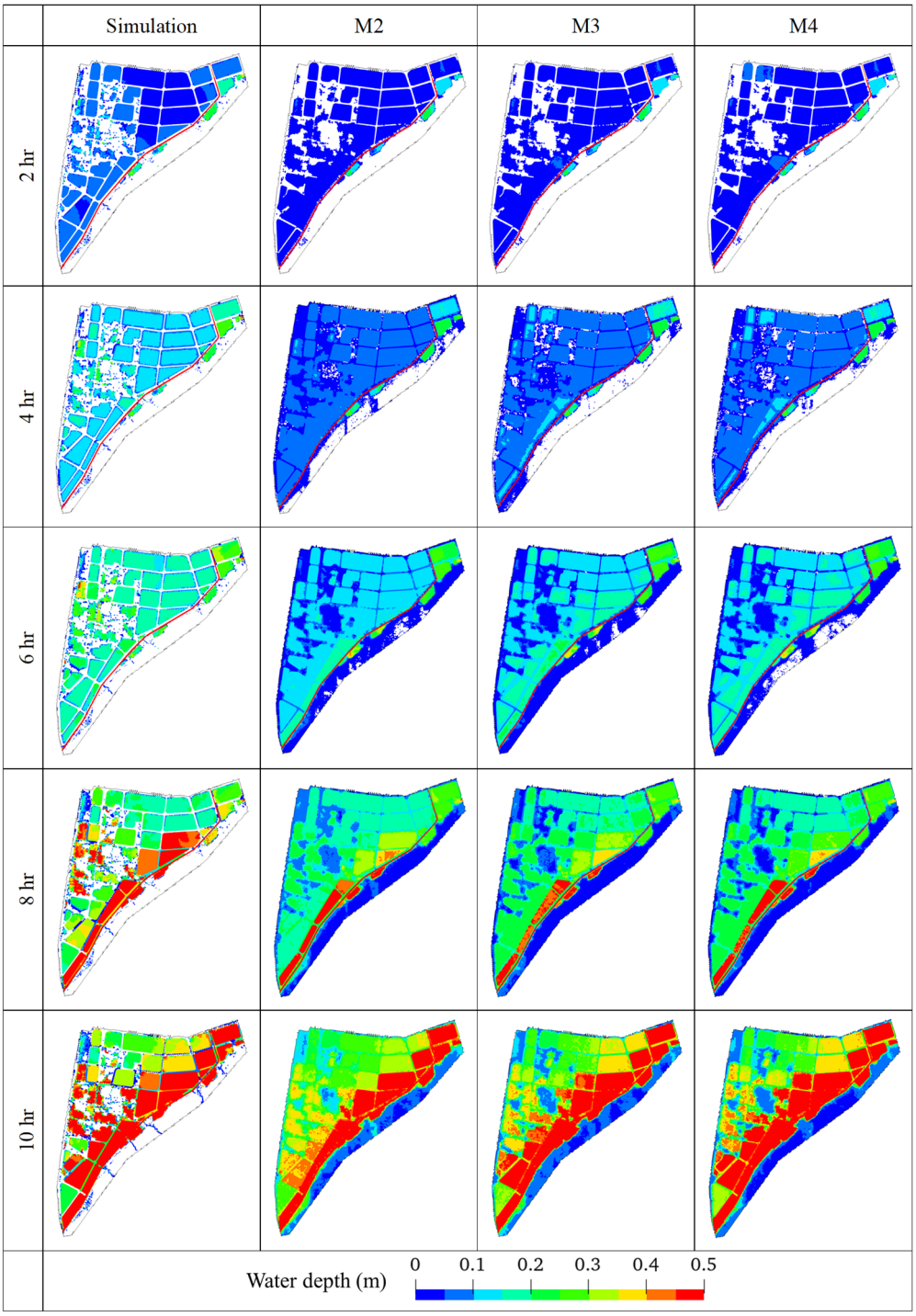


Fig. 4-19 Water depth contour comparison over time.

for the RF models starting with the first prediction and some areas of roads are falsely predicted as inundated. As time progresses, the area of false misclassification gradually expands, and large overestimation of inundated areas are observed in mountains starting 6 hr compared to simulations. Despite overestimating inundation areas, the overall predictions of water depths are shallower than simulation. At 4 hr for example, the predicted water depths range below 0.1 m over most of paddies while the actual depths exceed 0.11 m. This underestimation is more

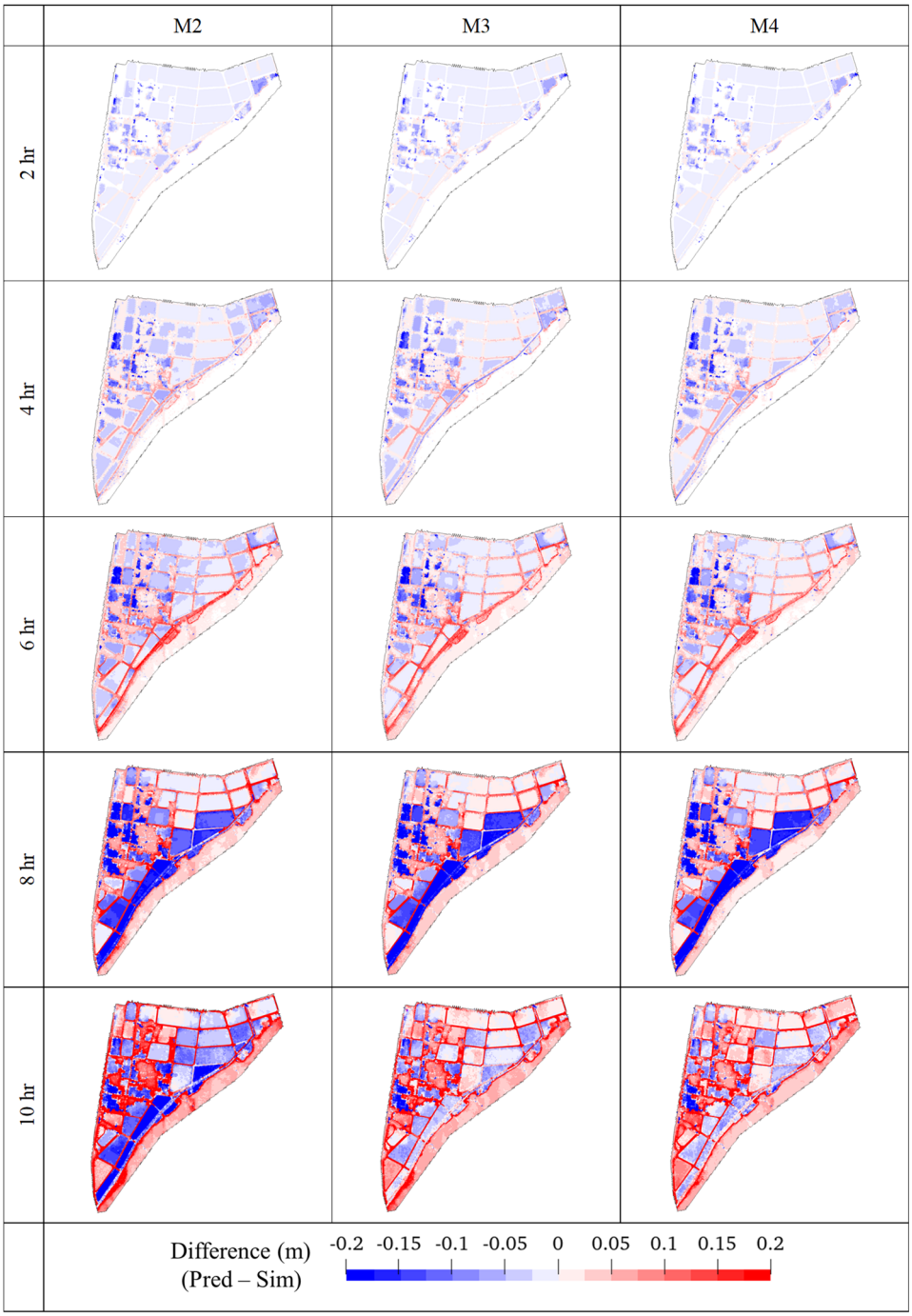


Fig. 4-20 Difference contour of prediction and simulated water depth comparison over time.

significant around the time when overflow occurs, especially around the paddies located close to rivers with underestimations of more than 0.1 m at 8 hr.

To observe the minor differences between the prediction of three models, the difference contour compared to simulation is illustrated in **Fig. 4-20** with a range of -0.2 m to 0.2 m. During the first two iterations, all three models resulted in similar underestimation in paddies and overestimation over certain areas of roads. As errors accumulate over iterations, the difference becomes much notable at 6 h that the roads between paddies at the downstream are overestimated by approximately 0.10 m to 0.15 m. Comparing M2 with the other models, the overall accuracy is slightly lower as indicated by the darker colors in the difference contour maps and the difference is especially notable at 10 hr where the water depth in paddies along rivers are better predicted by the latter two models. By further including features indicating land use, the water depth over mountain areas aligns closer to simulated results for M4, but the general trend for falsely overestimating inundation areas is yet not fully resolved. The performance of all three RF models is most compromised in the prediction at 8 hr where majority of paddies located along rivers resulted in an underestimation by more than 0.2 m.

The quantitative evaluation in terms of RMSE of overall and with respect to each land use is shown in **Fig. 4-21**. As observed in the contour map, the overall RMSE gradually increases with time. The performance difference between models is not significant up until 6 hr with a significant rise at 8 hr. Comparing different land uses, it is concluded that this error is mostly

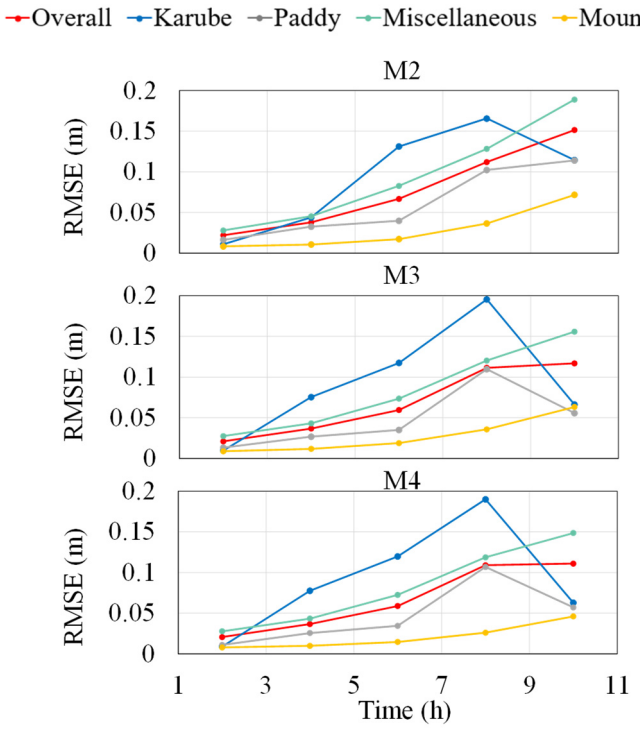


Fig. 4-21 RMSE comparison with respect of land use of iterated prediction.

contributed by paddy and Karube river, both of which having a steeper slope. The RMSE stables at 10 hr for M3 and M4 with a decrease in paddy while the that for M2 still increases. The RMSE for miscellaneous and mountain also approaches 0.18 m and 0.07 m for M2 while the other two models suppressed the values to approximately 0.15 m and 0.05 m respectively. And both the quantitative evaluations and contour maps suggest that including features on river boundary conditions better reflects the control volume of the system, such that the predicted water depth aligns better with simulation. Comparing the RMSE of miscellaneous and mountain for M3 and M4, it is implied that land use factors slightly improve model performance. On the other hand, the RMSE for the Karube river almost resulted in the highest value over time for all three models. This observation is mostly due to the imbalanced number of meshes for each land use. Since the Karube river only covers approximately 3.2% of meshes over the entire study area, a slight difference in the predicted value generates significant error when averaged. Another reason is that RF models make independent individual predictions without the actual logistics as those inherited in HD model. Shown in **Fig. 4-22** is the longitudinal water level extracted approximately every 50 m with the distance from downstream of the river basin of simulation and predicted results. The water level at 2 hr has yet to be stable even for simulation due to the closed downstream boundary, incoming flux from upstream residential areas and local drainage from paddies. The simulated water level stables starting 4 hr, yet the predicted results for all three models still vary significantly over the longitudinal direction without obvious trend. It is inferred that this difference induced by the nature of RF models that predictions are independent without investigating the actual hydraulic logistics.

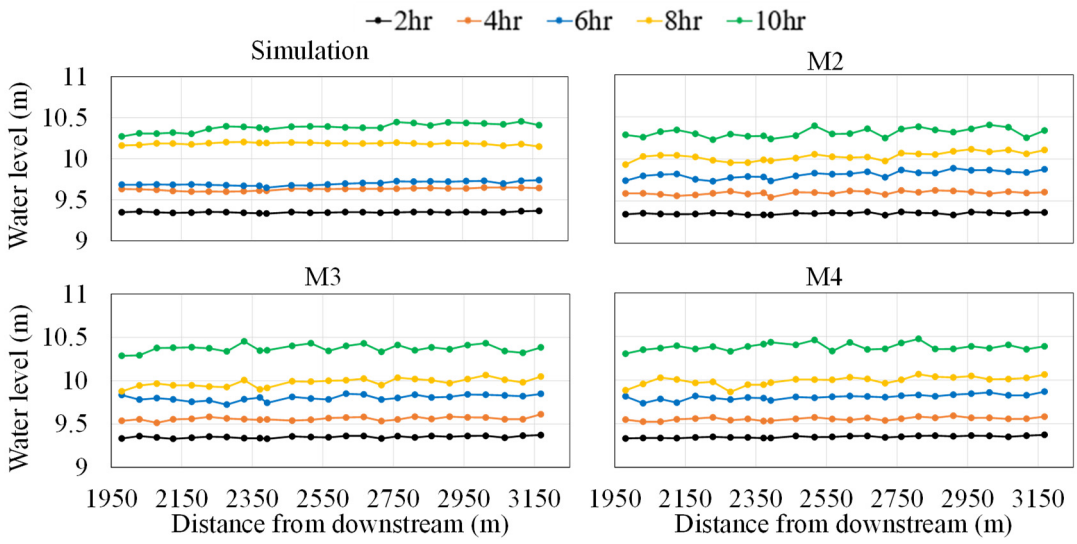


Fig. 4-22 Comparison of longitudinal water level in river.

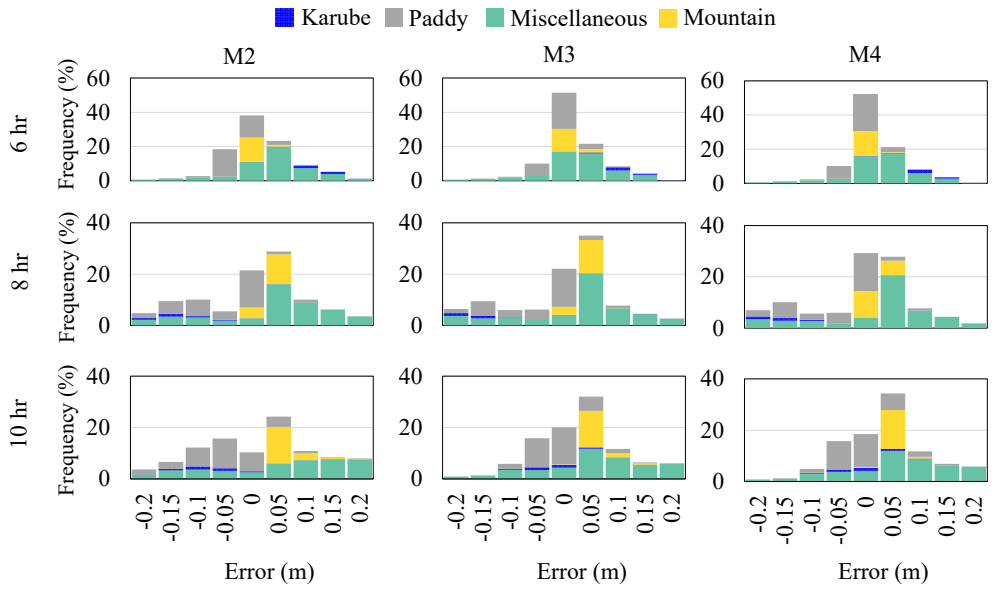


Fig. 4-23 Error histogram with respect to land use.

The error histograms with respect to land use of the three models are summarized in **Fig. 4-23** with the percentage against the total amount of mesh within computational domain on the vertical axis. In general, M3 and M4 outperform M2 as indicated by the larger percentage of meshes concentrated within errors of 0.05 m. The range of error gradually increases with iteration with around 10% meshes having errors of 0.2 m up until 10 hr, while more meshes have higher accuracy in M3 and M4. On the other hand, the prediction of water depth over miscellaneous land uses results in a tendency to overestimation, which led to the blurred land use delineation in the water depth contour map. Underestimations are especially notable in paddy areas, and the errors are spatially distributed differently. **Fig. 4-24** depicts a contour map of difference for the prediction of M4 at 6 hr with a smaller legend range. It is observed that the difference of predicted water depth within each paddy located at the upstream region is mostly distributed spatially evenly, with slightly smaller errors in paddies at locations close to the Karube river. Yet an inner gradient of error is noted in the three long and narrowly shaped paddies along the downstream, which suggests the actual change of water level within these paddies is not properly reflected in this RF model. It can be inferred that compared to the upstream, the hydraulic characteristics at the downstream are highly affected by other factors that the current RF model failed to incorporate, for example, pump drainage and discharge from paddies. Similar to the significant longitudinal variance observed in the predicted water level in rivers, this error is implied to be induced by the inherited nature of assuming independence of data without consideration of physical logistics. The significant increase in the range of error between 6 hr and 8 hr, while partially contributed by error accumulation due

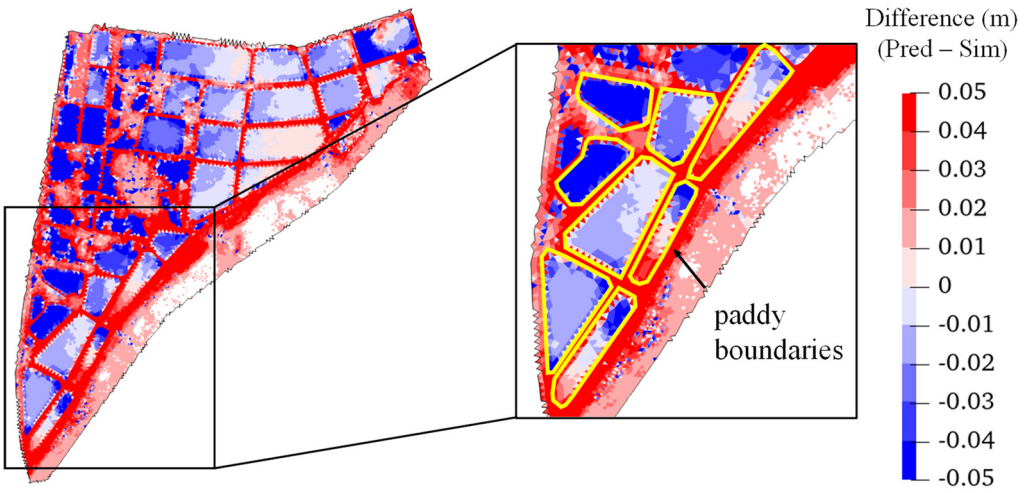


Fig. 4-24 Detailed difference contour of M4 at 6 hr.

to iterations, is mostly induced by the variation in hydraulic characteristics before and after overflow. As previously summarized in **Fig. 4-17**, accuracy decreases significantly from 6 hr to 8 hr even provided with accurate prior state information. Shown in **Fig. 4-25** and **Fig. 4-26** are the corresponding water depth and difference contour at 6 hr and 8 hr respectively carried out without iteration. Eliminating errors induced by iteration, the delineation between land uses in the predicted results aligns better with simulations at 6 hr for all three RF models. The misclassified inundation areas over mountains are also resolved comparing to the results of iteration prediction approach. Yet certain overestimations are still observed at areas located along rivers. When it comes to the prediction at 8 hr, accuracy is significantly compromised in terms of large areas of paddies being underestimated by more than 0.2 m. Observed from simulated results, the characteristic in the rise of water depths varies significantly before and after overflow even under constant rainfall since the contributing factor shifts from precipitation to the combined effect of flow exchange. Although including features indicating river boundary conditions in M3 and M4 slightly improves prediction accuracy, such significant underestimation still suggests a compromised model performance. To obtain better predictions, the prediction interval should be shortened to 10 min, for example, such that the hydraulic characteristics do not change significantly between intervals and prior state information could correctly reflect current property. A performance assessment was then evaluated by reconstructing the RF model with 10-min interval for prior state using the same features of M4 (hereafter denoted by M4_10min). In order to eliminate the error accumulation induced by different number of iterations, the result comparison was only carried out without long-term iteration (i.e. predicted with the simulated prior state information at each prediction). Illustrates in **Fig. 4-27** is the temporal RMSE evaluation against the same 25 mm/hr constant

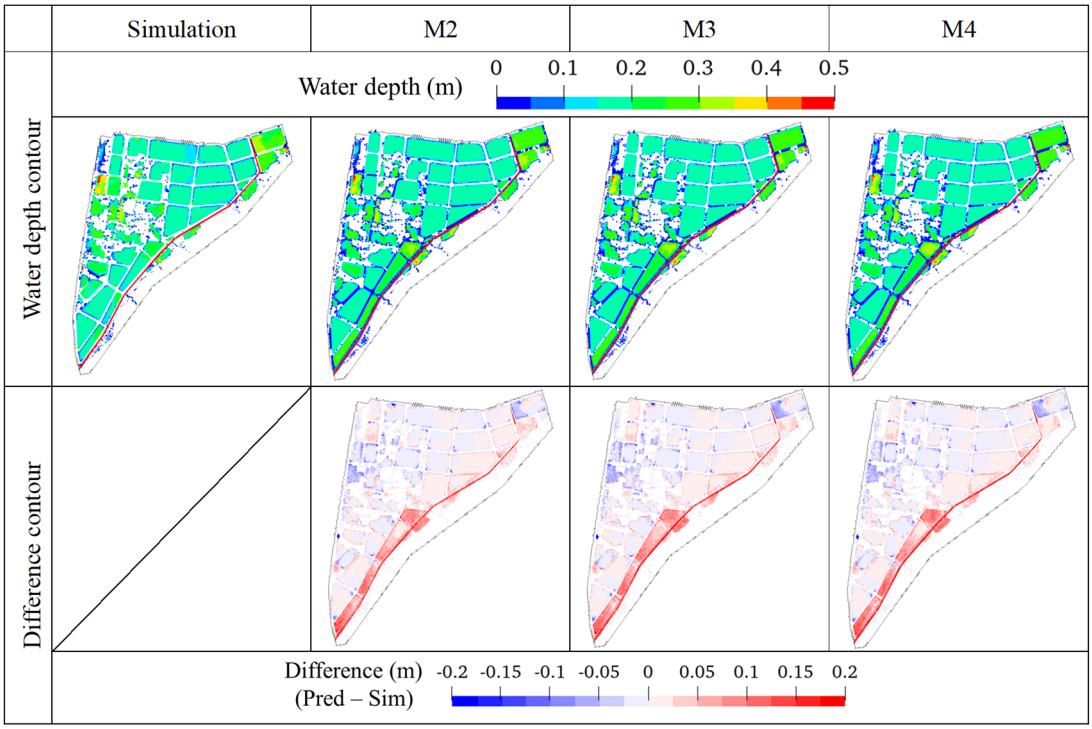


Fig. 4-25 Water depth and difference contour comparison for 6 hr without iteration.

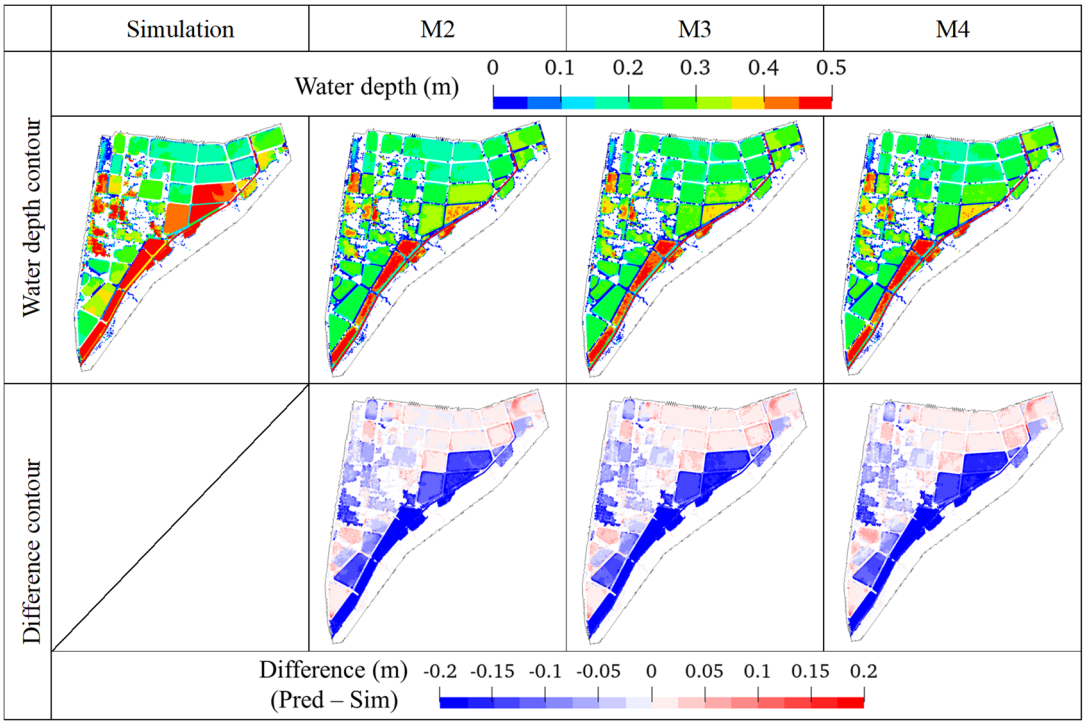


Fig. 4-26 Water depth and difference contour comparison for 8 hr without iteration.

rainfall. The prediction results indicated high performance with RMSE less than 0.05 m. A slight significant incremental increase is observed around 7 hr, which corresponds to the time when overflow from the Karube river occurs. The water depth and difference contour maps for the prediction at 6 hr and 8 hr are depicted in **Fig. 4-28**. Compared with the results predicted by models trained with 2-hr interval (**Fig. 4-25** and **Fig. 4-26**), the performance of M4_10min is significantly improved to the extent that the majority of error is distributed within -0.05 m to 0.05 m. This observation concludes that increasing temporal resolution allows the RF model to better capture the short-time dynamics and rapid changes of water depths more accurately.

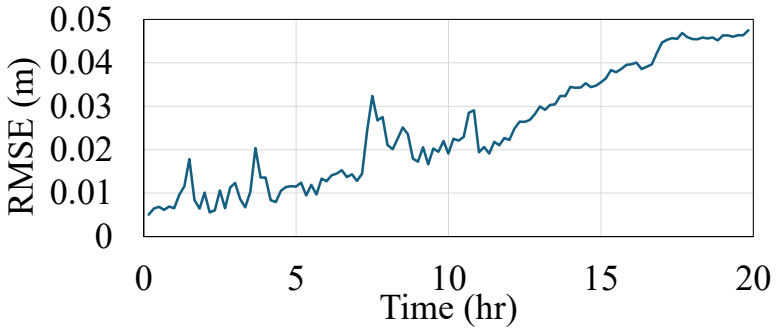


Fig. 4-27 RMSE evaluation for 25 mm/hr rainfall predicted by M4_10min without iteration.

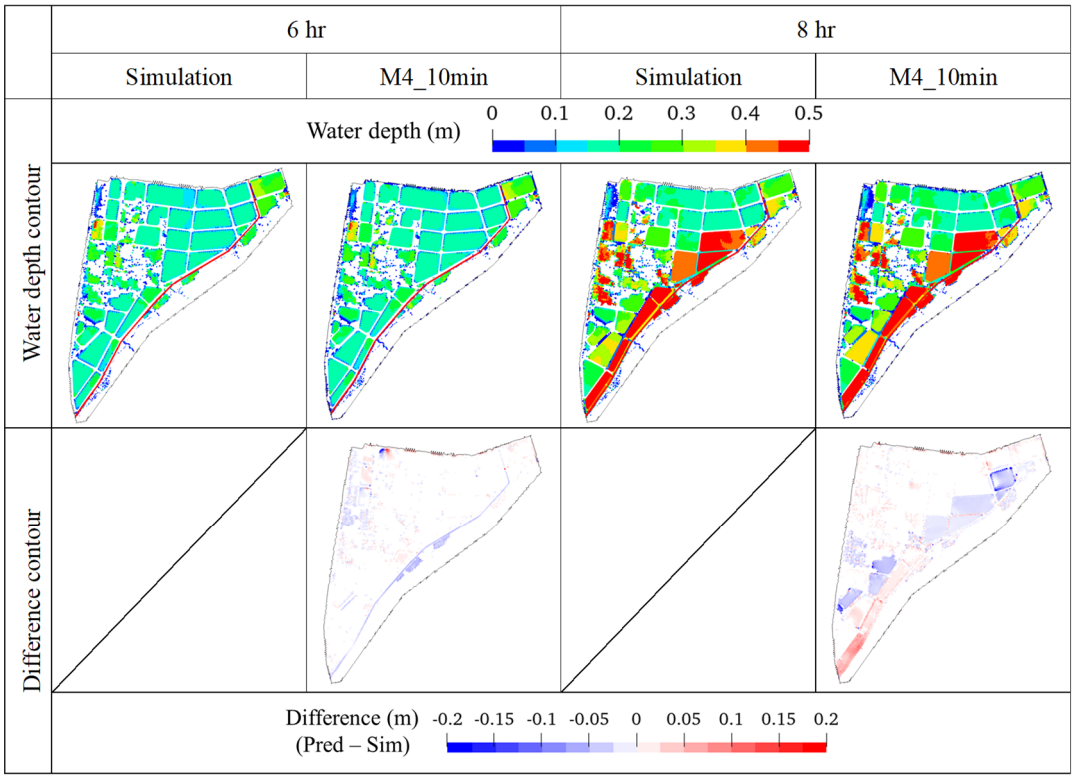


Fig. 4-28 Water depth and difference contour predicted with M4_10min without iteration.

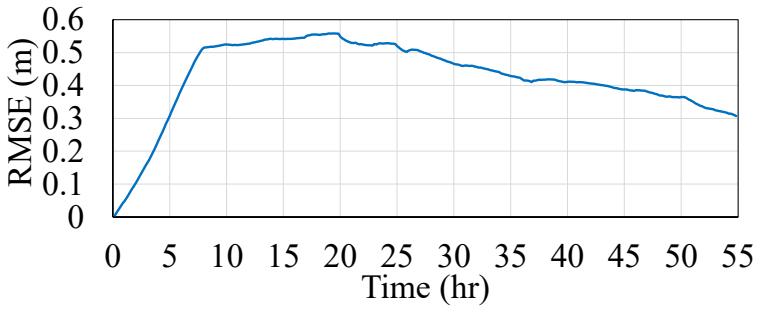


Fig. 4-29 Temporal variation of RMSE for 2021 rainfall iteratively predicted by M4_10min.

The current RF model, constructed with training data under constant rainfall conditions, on the other hand, struggles to generalize for the prediction of real rainfall events. **Fig. 4-29** shows the temporal RMSE for the iterative prediction of 2021 rainfall event using M4_10min. A significant linear increase in RMSE is observed with values greater than 0.5 m during the first 7 hr of rainfall events. Since it is observed from simulation results that the majority of area is only inundated by depths less than 0.1 m during this period (previously illustrated in **Fig. 4-11**), the significant RMSE values suggest that the current model fails to capture the dynamic hydrological behaviors of real rainfall events. With only constant rainfall conditions as the training data, the nonlinear response of flood propagation under rainfalls with temporal varying intensities is not properly reflected by the current RF model. As a result, evaluations regarding the hypothetical rainfall patterns in the training data are required to advance the model's generalization capabilities for practical rainfall patterns. The results also reveal that the iterative prediction process is sensitive to initial error. The finer iteration interval, while contributing to a higher accuracy in reproducing the hydrodynamic properties, also leads to significant error accumulation for long-term predictions. Therefore, further investigations into features selection are also essential to enable the current approach for flood predictions of long-term real rainfall events.

4.3.5 Computational cost evaluation

A comparison of computational cost under constant rainfall condition was carried out for efficiency evaluation. Summarized in **Table 4-4** are the hardware information and computational cost for the flood analysis of HD model and predictions for RF model. As stated in previous literature, the computational cost of the comprehensive flood analysis HD model had been significantly reduced with the assistance of parallel processing technique of GPGPU to be dozens of times faster than carried out with CPU alone[13]. Yet the computational time for a 20-hr event using HD model with the computational domain depicted in **Fig. 4-6(a)** was 48 min. In contrast, the training phase of RF model targeting the downstream area of interest (**Fig.**

Table 4-4 Comparison between hardware and computational cost.

Model	Hardware information	Computational time
HD model	GPU: NVIDIA Geforce RTX3090	48 min
RF model	CPU: 13th Gen Intel Core i7-13700, RAM 64 GB	Training: 73 min Prediction: 4.9 s/iteration

4-12(c)) required approximately 73 min with a total of 100 hr of training data and only took about 5 s to make one prediction using CPU. While the iteration prediction process requires reconstructing predicted outputs into the input for the next prediction between each prediction, the RF model still significantly reduces computational cost such that the prediction of a 20-hr event could be performed within minutes. Yet based on the discussion in the previous subsection, using 2.0-hr interval as the iteration prediction interval is not sufficient enough to capture the complicated variation in hydraulic characteristics and the accuracy for models with such interval is compromised in comparison to models constructed with 10-min intervals. Therefore detailed efficiency evaluation should be further carried out after improving the general performance of the RF model when targeting real rainfall events.

4.4 Chapter summary

This chapter carried out a baseline analysis to investigate the feasibility of applying the augmentation approach of HD model and RF model to make flood predictions over a practical terrain setting the Karube river basin under constant rainfall conditions. Summarized below are the main findings of this section.

1. The Karube river basin area is prone to flood, especially at paddies located downstream and along the Karube river. Under constant rainfall of 25 mm/hr, paddies gradually reach full capacity as precipitation and runoff from mountain accumulates. Together with overflow from the river, flood propagates with water depths greater than 0.5 m after 10 hr.
2. The feature importance of the configurated RF models resulted in similar trend as observed from the numerical experiments from the previous chapter and is consistent with the logistics inherited in HD models. As indications on control volume of the system besides precipitation, information of inflow and outflow at river boundaries is considered important in the configuration of RF model. The corresponding factor for rivers outweighs the other land uses due to the complexity of flux exchange at not only boundaries of computational domain but also within adjacent meshes. The prior state of previous water depths of adjacent meshes resulted in the highest feature importance, while the importance for the elevation of neighboring meshes was rather negligible.

3. Using an iteration prediction interval of 2.0-hr, the accuracy of prediction decreases significantly around the time when overflow occurs, even provided with accurate prior information. Since the characteristics in water rise changes from accumulation due to precipitation to the comprehensive flux exchange between meshes with large volume incoming and outcoming from computational boundaries, prior information before overflow occurs could not correctly capture the characteristics afterwards.
4. Including features reflecting inflow and outflow information of rivers and indications of land use improves model accuracy using 2.0-hr iteration interval. The RF models resulted in a tendency to falsely overestimate areas that are not inundated, and the error of overestimation accumulates over iterations. On the other hand, the prediction of water depths of paddies are underestimated. The water level in rivers also varied significantly longitudinally, which again suggested the inherited nature of RF models in terms of making independent predictions based on data-driven statistics instead of the governing equations in HD models.
5. Although the training phase of RF model was slightly time consuming due to large number of training data, the actual prediction process significantly reduced the computational cost to an average of 5 s per predictions carried out under CPU. The augmentation approach of RF model configuration and iteration prediction process indicated the potential for real-time applications, provided that the iteration interval and features to be further investigated to improve prediction performance.

Based on the observations obtained from this chapter, the following summarizes the limitations of the current model and necessary procedures to be carried out in the following chapters in order for the augmentation approach and iterated prediction to be applied under practical situations where rainfall conditions and flood characteristics are more complicated.

1. The rainfall intensities used in the flood simulation were too extreme and lacked diversity. While peak intensity of 70 mm/hr could be observed under rare situations, an intensity of such lasting over hours is rather unrealistic. The current RF model also fails to reproduce the results under practical rainfall conditions due to the limited variety of training data and further investigation should be carried out to increase the diversity of hypothetical rainfall patterns used to construct training data.
2. While iteration interval of 2.0-hr reduces computational cost, this interval fails to properly capture intense changes in hydraulic characteristics, especially before and after overflow occurs. Shortened iteration intervals allow for better reflection of immediate dynamic changes and increase local performance, while also inducing the potential for error accumulation over long-term predictions. Therefore, further discussions regarding features selection are also necessary to enable the stability and reliability of the model for practical

applications. Based on the observations from this chapter, the following aspects should be investigated specifically. The elevation of adjacent meshes resulted in almost negligible feature importance for all the three RF models. Since the idea was to use both the elevation and water depth to represent the water surface such that the RF model follows the logistics of HD model, investigations should be carried out replacing the features as the water level of neighbor meshes. The prediction process of M3 and M4 included the inflow and outflow at river boundaries with the assumption that such information is available. However, discharges are rather difficult to be observed compared to water levels. Therefore, the possibility of replacing discharge to water level information is to be further studied considering practical applications of the method. Finally the current approach of binary indication of land use resulted in a misclassification of false positive and blurred boundaries between different land uses. Since the iterated prediction process is subject to error accumulation and sensitive initial error, other methods to represent land use should be investigated to improve the performance while maintaining the interpretability of RF models.

References

1) Homepage of Soja City: *History of disasters in Soja City*. https://www.city.soja.okayama.jp/data/open/cnt/3/12320/1/P06.pdf?20250904181510 (Accessed September 28, 2025).
2) Homepage of Okayama Prefecture: *Flooding records maps of Karube river basin*. https://www.pref.okayama.jp/uploaded/life/689339_6119017_misc.pdf (Accessed July 1, 2025).
3) Ministry of Land, Infrastructure, Transport and Tourism: *Water Information System*. http://www1.river.go.jp/ (Accessed November 11, 2024).
4) Homepage of Okayama Prefecture: *Flood records of the Karube River Basin*. https://www.pref.okayama.jp/uploaded/life/689339_6119017_misc.pdf (Accessed July 1, 2025).
5) Crisis Management Department of Soja City: *Memorial Journal of Heavy Rain Disaster Response of July 2018.* https://www.city.soja.okayama.jp/s/kikikanri/kurashi/bousai/h30gouusaigai_kiroku/h30gouusaigai_kiroku_top.html (Accessed May 22, 2025)
6) Data provided by Soja City: *Flood extent during heavy rainfall of July 2020*.
7) Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., Allaire, M., & Matthew, R. A. (2018). What Is Nuisance Flooding? Defining and Monitoring an Emerging Challenge. *Water Resources Research*, 54(7), 4218–4227. https://doi.org/10.1029/2018WR022828
8) Ministry of Agriculture, Forestry and Fisheries Rural Development Bureau Development Department (2022). *Handbook for Rice Paddy Dams*.
9) Geospatial Information Authority of Japan: Basic map information download service.
10) Data provided by Soja City: *Information of the longitudinal riverbed of the Karube river*.
11) Akoh, R., Takuno, T., T., Matusi, D., & Maeno, S. (2024). Proposal of Efficient Analysis Method for Paddy Field Dam for River Basin Flood Control and Field Application. *Japanese Journal of JSCE*, 80(16), 23-16073.
12) Committee of Hydraulic Engineering (2018). *Formula of Hydraulics*, JSCE.
13) Zeng, X., Akoh, R., Ishikawa, T., & Nakamura, T. (2013). On the Efficiency of Gpgpu Acceleration of Tsunami Simulations Using AN Unstructured Triangular Mesh System. *Journal of Japan Society of Civil Engineers*, Ser. B1 (Hydraulic Engineering), 69(4), I_619-I_624.

CHAPTER 5 INVESTIGATION ON THE EFFECT OF HYPOTHETICAL RAINFALL IN TRAINING DATA AND FEATURE REFINEMENT

5.1 Overview

The previous two chapters investigated the applicability of the augmentation approach of HD model and RF model for efficient spatiotemporal flood predictions under constant rainfall conditions. Configurated with flood simulation results from HD model as the training data, the RF model incorporated features of static properties of river basin and prior state water depth of adjacent meshes, as the spatial factor. An iteration prediction method where previous predicted outcomes are reconstructed into the input dataset for succeeding predictions was proposed aiming for detailed temporal water depths predictions. Further including features indicating river boundary conditions and land use, the RF model was able to capture the general trend of flood propagation and iteratively predict water depth with relatively high performance up until the time when overflow occurs as well as maintaining certain interpretability of features. On the other hand, the results also indicated certain limitations regarding the reality of hypothetical rainfall condition, iterative prediction interval and feature selection in terms of data availability and reducing computational cost.

This chapter further carried out the following investigations to assess the potential of using high-accuracy flood analysis results from HD model under hypothetical design rainfall conditions to train RF model such that flood predictions under actual rainfall events could be obtained spatially and temporarily. First, the rainfall characteristics of the Karube river basin were summarized and hypothetical rainfall patterns in terms of magnitude, intensity and distribution of peak were designed accordingly. The hypothetical rainfalls were then applied spatially evenly over the Karube river basin and inundation analysis with HD model was performed. The simulated results for the area of interest were picked up and summarized as individual datasets. Regarding the configuration of RF models, different selections of features and cross validation methods were first investigated using one pattern of hypothetical design rainfall to optimal prediction performance while maintaining interpretability. The combination of features and validation methods that resulted in highest performance was then adopted to

further investigate the influence of training data. To be specific, RF models were trained with various subsets of simulated database and tested against untrained patterns and intensities of design rainfall to investigate the differences in the generalization capability of models.

5.2 Rainfall characteristics of the Karube river basin and design of hypothetical rainfall

Three important factors to characterize rainfall patterns include the magnitude, spatiotemporal distribution and peak intensity of rainfall. Located at the downstream of the Takahashi river, the designated design rainfall magnitude in terms of river management and determination of inundation prone area takes the magnitude of the Takahashi river system, which includes the planned rainfall of L1 scale and maximum foreseeable rainfall of L2 scale, according to the revised Flood Control Act published in 2015[1)]. The cumulative rainfall magnitude of L1 with return period of 100 yr and L2 scale, reported by Okayama prefecture as of August 30, 2024, is 233 mm and 674 mm over a 48-hour period, respectively. Another two-day cumulative rainfall information with return period of approximately 150 yr is 248 mm as provided by Soja city[2)] and it can be noted that the rainfall event in 2018 is an equivalent of this scale. The distribution of the rainfall intensity significantly affects when the peak of flood occurs within a river basin. The temporal distribution of rainfall in Japan is typically characterized into the following categories depending on the relative arrival time of peak intensity during a precipitation event: front-loaded, centralized, back-loaded and flat-typed patterns. For large river basins, representatives of spatial distribution also include middle-upstream, middle-stream and downstream patterns regarding the location where precipitation concentrates[3)]. The peak intensity for rainfall with different return periods is obtained through the intensity-duration-frequency (IDF) equation, also known as the Kimijima equation, as the following.

$$r = \frac{a}{T^n + b} \tag{5.1}$$

where r (mm/hr) is rainfall intensity, T (hr) is rainfall duration, and a, b, n are the empirical coefficients obtained from observatories. Two observatories located close to the Karube river basin are the Takahashi and Yakage observatories and summarized in **Table 5-1** and **Table 5-2** are the peak rainfall intensities of rainfalls and 5-hr cumulative rainfall with different returen periods calculated based on the coefficients listed in the Okayama Prefecture Forest Land Development Permit Application Guide (issued April 2023)[4)], respectively. A centralized rainfall hyetograph could then be generated through the conventional method that consists of

Table 5-1 Peak intensity (mm/hr) of design rainfall with different return periods.

Observatory	Return period (yr)				
	2	5	10	20	30
Takahashi	27.59943	36.99863	43.30174	49.30193	52.80018
Yakage	24.23018	32.53724	38.04445	43.25085	46.3526

Observatory	Return period (yr)				
	50	70	100	150	200
Takahashi	57.10222	59.90251	62.90456	66.30412	68.70245
Yakage	50.05979	52.54320	55.16470	58.16827	60.26884

Table 5-2 5-hr cumulative rainfall (mm) for design rainfall with different return periods.

Observatory	Return period (yr)				
	2	5	10	20	30
Takahashi	85.85076	111.3466	128.2396	144.4805	153.8663
Yakage	77.93045	101.5071	117.077	131.9985	140.7175

Observatory	Return period (yr)				
	50	70	100	150	200
Takahashi	165.4946	173.072	181.1951	190.3729	196.8968
Yakage	151.3659	158.4048	165.8888	174.3807	180.3549

the following procedures: (a) selection of imperical coefficients for the IDF equation according to location and corresponding flood design scale of the river, (b) determination of flood arrival time at locations where discharge for flood design scales were calculated based on land use conditions, (c) calculation of rainfall magnitude for each rainfall duration taking the flood arrival time as the unit time, and (d) generation of hyetograph by assigning large magnitudes at the center of rainfall duration and the remaining magnitudes sequentially distributed before and after the peak[5)].

With details summarized previously in Chapter 4, the Karube river basin suffered a few major floods historically. Comparing the designiated peak intensities to the scales of observable historical rainfall records at the Soja City, it is concluded that while the peak intensities of historical precipitation events only takes the scale of approximately 2-yr return period, the long-lasting duration leads to significant cumulative rainfall with scales close to 150-yr scales. One major advantage of the augmentation approach of HD model and RF model is the flexibility in the generation of training data. Instead of depending purely on limited and low-quality historical observations, the configuration of RF models in this research used

accurate hypothetical flood simulations with high resolution obtained from HD models as the training data. Considering the complexity of real rainfall events, the design of hypothetical rainfall patterns took into account the magnitudes of design rainfalls as characteristics above with simplified temporal distributions using the following equation.

$$P(t) = R_{peak} \cdot \exp\{-a(t - b \times R_{time})^2\} \tag{5.1}$$

where $P(t)$ (mm/hr) is the rainfall intensity at time t in hr, R_{peak} (mm/hr) is peak rainfall intensity, R_{time} (hr) is total calculation time, and a, b are coefficients determining the shape and peak arrival time of rainfall respectively. This study investigated 2 types of shapes, normal distribution (hereafter denoted by ND) and narrow normal distribution (hereafter denoted by NND), and 3 types of peak time, early, center and late (denoted by E, C, L, respectively), with the corresponding coefficients summarized in **Table 5-3**. The total calculation time was first set to be 5 hr and illustrates in **Fig. 5-1** is a normalized representation of the hypothetical rainfall patterns. The selection for peak intensity referred the magnitudes of design rainfalls using the IDF equation with different return periods, which are 20 mm/hr, 30 mm/hr, 40 mm/hr, 50 mm/hr, 60 mm/hr, 70 mm/hr. With a range of cumulative rainfall from 35.3 mm to 173.7 mm, the hypothetical rainfall patterns cover approximately the same scale of the design rainfall intensities and cumulative rainfalls at the Karube river, as well as the different categories of

Table 5-3 Notations of hypothetical rainfall and corresponding coefficient settings.

Hypothetical rainfall	Coefficients	
	a (shape)	b (peak time)
ND_C	0.5	0.5
ND_E		0.4
ND_L		0.6
NND_C	1.0	0.5
NND_E		0.4
NND_L		0.6

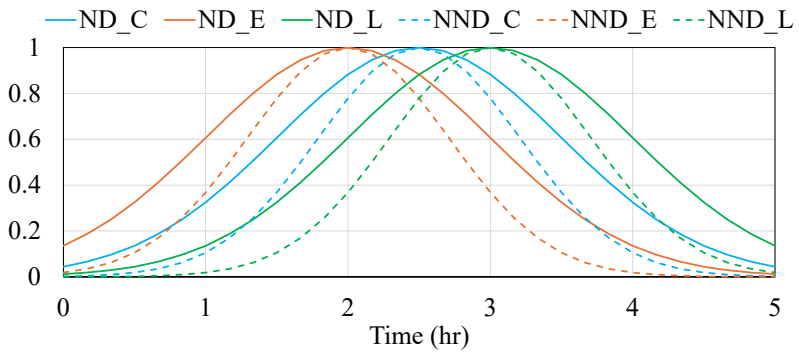


Fig. 5-1 Schematic of the temporal distribution of normalized 5-hr hypothetical design rainfalls.

peak arrival time and temporal distribution. The influence of spatial distribution of precipitation in the case of the Karube river basin was inferred to be rather insignificant considering the area of the region. Therefore, rainfall intensities were applied spatially evenly over the river basin for all caluclations. With a 30-min shift in peak time, the cumulative rainfall regarding rainfall patterns with different peak time is similar to each other. But the inundation characteristics observed at the downstream target area, such as when overflow from the Karube river occurs and how long inundation lasts, varies significantly due to flux interaction at river boundaries. Other calculation conditions of HD models such as boundary conditions of rivers, land use parameter setting, and initial water depth follows the same situations as described in the previous chapter. Regarding the interval for iterated prediction, findings were that shortened intervals are necessary when significant temporal variation in the characteristics of water depth are observed. Therefore the interval for prediction was refined to be the same interval for the output of HD models, which is 10 min. The data processing for results pick up regarding the same area of interest of the RF model from the previous chapter and summarization of training data were also carried out with 10-min intervals.

5.3 Refinement of RF model configuration

5.3.1 Investigation on feature selection

As discussed in the previous chapter, further investigation was necessary regarding features on elevation of adjacent meshes, representation of land use and the availability boundary conditions of rivers. Therefore, this section first configurated multiple RF models with different features to validate the augmentation approach under temporally varying rainfall conditions while optimizing prediction performance. Model configuration and evaluation were carried out with the same training data and tested against the same rainfall event to avoid biases induced by the difference in dataset. To be specific, the training data consisted of the 6 subsets of simulation results using normal-distributed rainfall pattern with centralized peak time (hereafter denoted by ND_C) with the above-mentioned intensities. Each of the rainfall patterns included 30 steps of simulation results, each containing details of the 18,687 meshes. The performance of RF models was tested against rainfall events of the same ND_C pattern but with untrained intensities of 45 mm/hr (ND_CR45).

The common features adopted from previous discussions are coordinates (x, y), elevation (z), rainfall at each step ($rain_t$). Instead of cumulative rainfall, this section selected the rainfall at previous step ($rain_{t-1}$). Indications here are that lagged rainfalls capture immediate influence of the input volume, and cumulative rainfall, as a representation of overall state of the system,

Table 5-4 Additional features used for model configuration.

Models	Features	Training data
M1	$h_{t-1,\,adj}$, z_{adj}, land use (binary)	ND_CR20, ND_CR30, ND_CR40, ND_CR50, ND_CR60, ND_CR70
M2	$H_{t-1,\,adj}$, land use (binary)	
M3	$H_{t-1,\,adj}$, $Fr_{t-1,\,adj}$, Fr_{adj}	
M4	$H_{t-1,\,adj}$, land use (binary), $Fr_{t-1,\,adj}$, Fr_{adj}	
M5	$H_{t-1,\,adj}$, $Fr_{t-1,\,adj}$, Fr_{adj}, H_{in}, H_{out}	

is already represented by the features indicating prior states of adjacent cells. With the same logistics, the same prior state at each mesh (h_{t-1}) was added as a common feature for all model configurations. As summarized in **Table 5-4**, the main discussion on feature refinements involves mainly on the selection of how to represent prior state information of adjacent meshes, land use and boundary information of rivers, and the selection of adjacent meshes are the same as stated in previous chapters. M1 adopted the same representation as discussed in the previous chapter where prior state water depth of adjacent meshes ($h_{t-1,\,adj}$) and elevations (z_{adj}) were separated as different features and indictors for land uses were one hot encoding binary features. M2 combines prior water depth and elevation into the water level of adjacent meshes ($H_{t-1,\,adj}$). Although the two representations stand for the same hydraulic property in physical logistics, differences in accuracy might be induced depending on how static and dynamic features are interpreted in RF algorithms. Investigation regarding M3 was carried out with the same considerations where representation of land use was revised into runoff coefficient. As detailed in Chapter 2, the runoff coefficient in HD model settings is switching variable dominated by saturated rainfall of each land use and cumulative rainfall. While the coefficient at each mesh shifts only once from unsaturated to saturated value, the time exceeding the threshold varies depending on land use and the characteristics of water rise differ before and after this shift. An indication of this temporal variation is considered important, and M3 incorporated the feature of runoff coefficients of adjacent cells both at previous step ($Fr_{t-1,\,adj}$) and current step (Fr_{adj}). M4 then further investigated whether including both static and dynamic representations of land use result in redundancy or improvement in accuracy. Finally, M5 built on the configuration of M3 with additional features on boundary information of rivers. The discharge of rivers is difficult to measure compared with water level in practical situations. Yet such information is considered an important contributing factor that determines the control volume within the system and model accuracy. As a result, M5 investigated the feasibility of refining the features of river boundary conditions into the water level at upstream and downstream of river (H_{in}, H_{out}). Assumptions were made that the temporal varying river boundary water levels are observable, and the same boundary water level values were assigned to all meshes in the configuration of RF models. Hyperparameter settings were assigned to be

Table 5-5 Details of configurated RF models.

	M1	M2	M3	M4	M5
Number of features	33	21	32	35	34
Training time (s)	945	1381	1260	1052	1303
Prediction time (s)	12	14	15	22	20
Model size (GB)	1.26	1.51	1.44	1.48	1.48
RMSE (m) of validation set	0.008	0.007	0.007	0.007	0.006

the same values (*n_estimator* = 100, *max_depth* = 20, *max_features* = 'sqrt', splitting = *train_test_split*) for all RF models to maintain consistency between models other than features. Unless specified otherwise, all predictions were performed using iterated prediction process with 10-min interval.

Table 5-5 lists the details of the configurated RF models in terms of total number of features, training time, prediction time of validation set, model size and RMSE evaluated against validation set. Even using the same training dataset and hyperparameter, the time consumption in training varies from 15 min to approximately 23 min depending on the number of features.

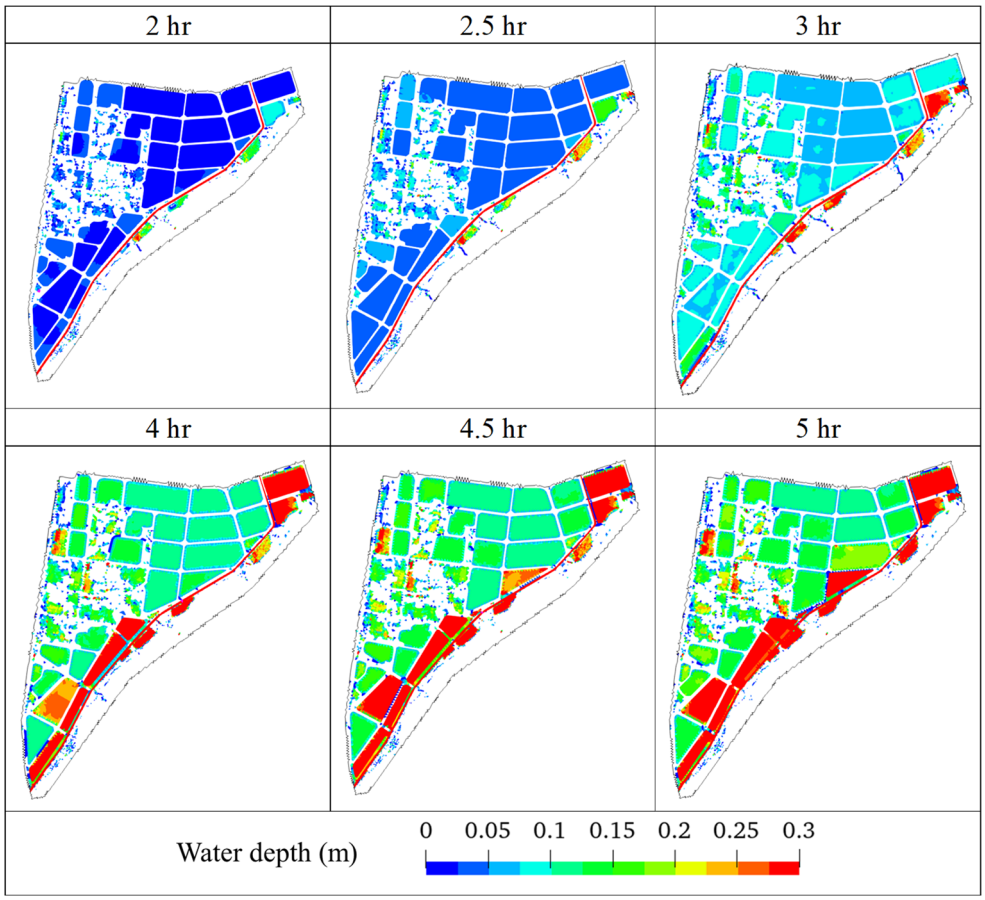


Fig. 5-2 Simulated water depth contour under rainfall ND_CR45.

But once being trained, the prediction time performed against the same subsets of validation dataset is similar for all 5 models. The size of configurated models also varies slightly, and fewer features do not necessarily result in small model size: for example M2 results in the largest model while having least number of features. Explanations are that although the maximum possible depth of trees is defined, decision trees may not reach this threshold. With more features to select, each split of trees shows more randomness such that shallow and effective trees are constructed. As for the evaluation against validation dataset, all models result in higher accuracy with RMSE within 0.01 m. Therefore, the models were further tested over rainfall pattern ND_CR45 for the evaluation of prediction capability over untrained events.

First, depicted in **Fig. 5-2** are the simulated water depth contour maps of the area of interest. The water depth in paddy gradually accumulates and the ones located at the upper right corner of computational domain reaching full capacity at 3 hr. Since this area is located at the foot of the Karube mountain, additional surface inflow from mountain areas naturally drains into these paddies, leading to a higher water depth compared with other areas. Slight overflow from the Karube river over roads at the downstream most region is also observed at 3 hr. Because of this overflow and decreasing of storage capacity over time, flood gradually propagates from river to distant areas and eventually results in pooling condition with a water depth of approximately 0.3 m at the downstream. As for areas slightly distant from the Karube region, the water depths at the final step in paddies, induced mainly due to direct precipitation, ends up to be around 0.15 m and the majority of road areas is not inundated.

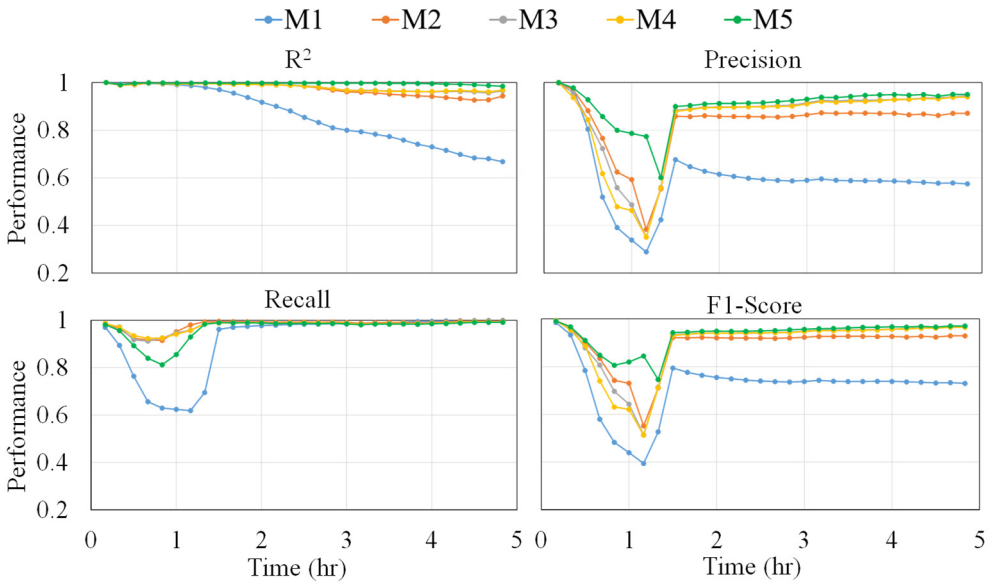


Fig. 5-3 Performance comparison of RF models tested against rainfall ND_CR45.

The performance metrics of RF models predicted against ND_CR45 rainfall event are shown in **Fig. 5-3**. Similar to observations from previous chapter, the R^2 values decrease as time progresses and the performance for M1 is rather compromised by the end of calculation period while all other models still having values greater than 0.9. Evaluated with a threshold of 0.01 m for the classification of inundation areas, all models indicated a slightly low performance during the first 1.5 hr. The low precision and recall values suggest that capability for identifying inundation or not is slightly compromised during the first few steps by falsely alarming non-inundation areas as well as missing inundated areas. Yet from the simulation results in **Fig. 5-2**, it can be observed that even at 2 hr the majority of the area is only inundated by a shallow depth of 0.02 m to 0.03 m. That is, the low performance scores and misclassification when actual water depth is only under several centimeters are more likely to be induced from the arbitrary choice of threshold and numerical errors. Other evidences for this inference will be carried out with discussions regarding detailed quantitative evaluation of inundation area and volumes and visualization of contour maps. Nevertheless, the precision score for M1 after 1.5 hr only reaches 0.6 while that of all other models bounce back to approximately 0.9, suggesting that the features of how the prior state information is represented introduced a significant difference. The RMSE evaluation, illustrated in **Fig. 5-4**, also indicates a significant linear increase of error in M1 when other models suppressed the overall error within 0.1 m at the end of prediction.

A comparison of predicted results using different prediction approaches was carried out to evaluate whether this increase of error is induced by the nature of the model or error accumulation over time. Depicted in **Fig. 5-5** is the comparison in terms of inundation area and inundation volume, where areas having water depths greater than 0.01 m are classified as inundated. Here 'Iterated' denotes the iterated prediction method proposed in this research, and 'StepPred' represents the process where correct prior state information, as simulated from HD models, is used as the input for each prediction such that error accumulation between steps

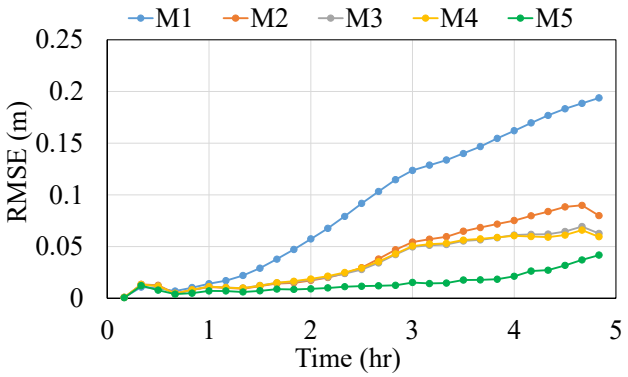


Fig. 5-4 RMSE comparison of RF models tested against rainfall ND_CR45.

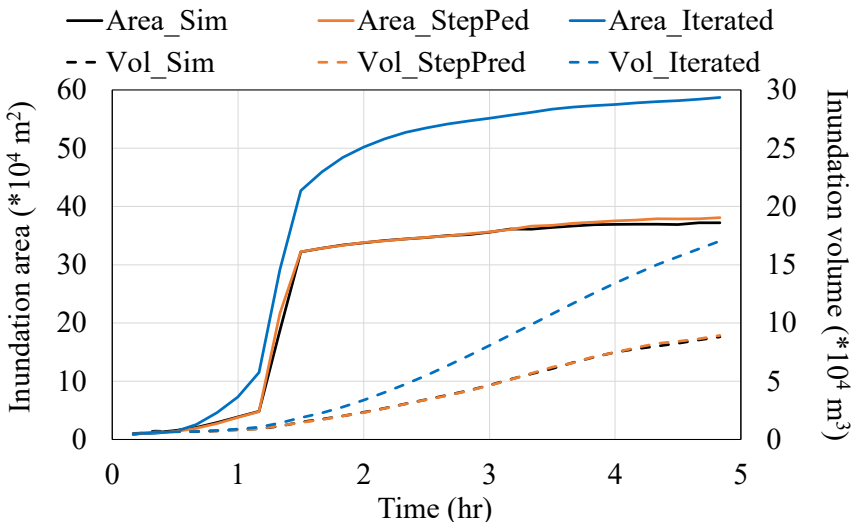


Fig. 5-5 Comparison of simulated and predicted inundation area and volume using different prediction approach of M1.

is eliminated. Consistent with the low precision score, the iterated process tends to falsely overestimate inundation areas during the first 1.5 hr. The predicted inundation volume, on the other hand, does not diverge significantly compared to simulated results since the inundation depths are rather shallow at this period. This again provides evidence that the misclassification is inferred to be induced by numerical error and could be considered rather insignificant. The error in inundation area and volume using iterated prediction approach becomes notable after 1.5 hr. As observed from simulated contour maps, the flood propagation, even after overflow from the Karube river, concentrates mostly along road areas close to the river which corresponds to the gradually flattened increase in simulated inundation area. Yet the inundation area using iterated prediction still increases with a relatively steep gradient, as well as the inundation volume, and the predicted area and volume is 1.5 and 1.9 times the simulated results at the end of 5 hr respectively. In comparison, the results of StepPred coincide highly with simulated results over the entire periods. That is, while the accuracy of M1 is high when provided with correct input step by step, the iteration process introduces significant error accumulation over time in terms of both area and volume.

Despite error accumulation over time, the main reason leading to the low accuracy of M1 is inferred to be the selection of prior state of features. A comparison of water depth contour maps at 2 hr and 3 hr of M1 and M2, both using iterated prediction approach, is illustrated in **Fig. 5-6** to observe the spatial differences. It can be clearly observed that the overestimation of inundation area distributes mostly over areas with miscellaneous land uses to the extent that the differentiation for paddy boundaries for M1. The inundation depths along rivers are also significantly overestimated. The overestimation accumulates over time where misclassification in mountain areas is also observed at 3 hr and the error in depth prediction

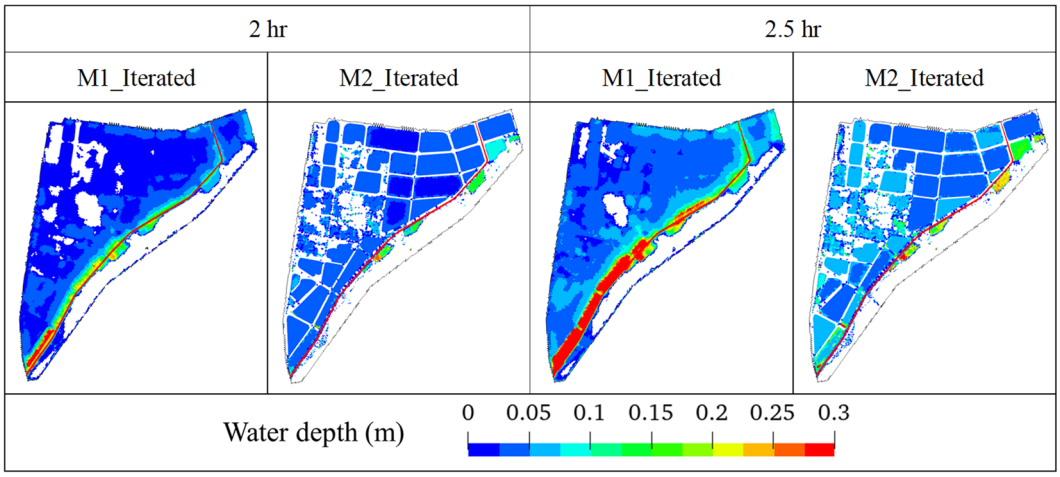


Fig. 5-6 Comparison of depth contour of M1 using different prediction approach.

increases as well. In contrast, M2 provides better identification of inundation areas, even though the prediction of water depth is slightly off compared to simulated results. In other words, although the features of M1 and M2 contain the same physical information, they are not interpreted equivalently in RF models. The feature importance of the elevation and prior water depth of neighbor meshes is depicted in **Fig. 5-7**. The importance of elevation is almost negligible compared to prior water depth since there is no temporal variation in elevation. As RF models select only one feature at each split of trees, the most effective factor in predicting current inundation depth is therefore captured through prior depth information in M1. Deep and complicated hierarchical splits are required in order for the RF model to correctly capture the relationship between the two features, yet the current configuration fails to build such correlations. By contrast, the representation in terms of prior water level of M2 directly represents the physical property, which allows the RF model to capture the influence of features more easily. Because water surface elevation is the physical driver of actual flood propagation, M2 results in better prediction results compared to M1.

The representation of land use was then investigated through M2, M3 and M4. Based on the

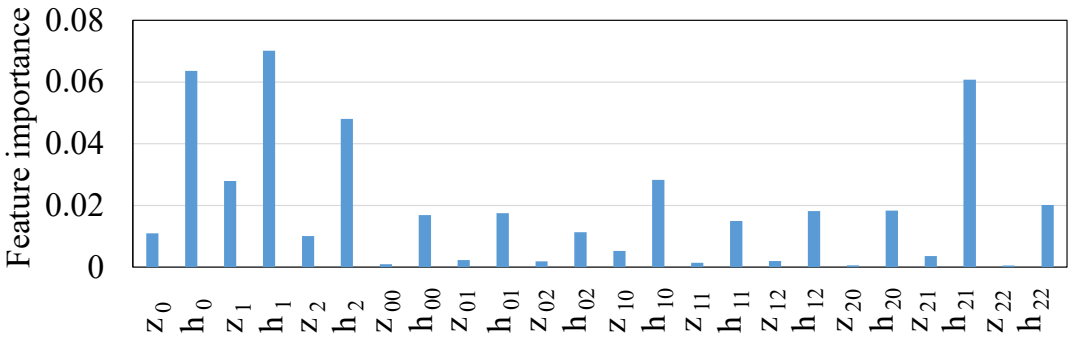


Fig. 5-7 Feature importance of prior depth and elevation of adjacent meshes in M1.

performance metrics and RMSE evaluations, M3 and M4 resulted in similar predictions, both slightly overperforming M2 (**Fig. 5-3** and **Fig. 5-4**). The low performances of precision, recall and F1-score for all three models during the first 1.5 hr are, again, inferred to be contributed from numerical error and selection of thresholds. In terms of RMSE, the difference between models is quite negligible up until 3 hr and the accuracy diverges afterwards. The predicted water depth contour maps after 3 hr are depicted in **Fig. 5-8**, and the difference maps of prediction over simulations are illustrated in **Fig. 5-9** for easier comparison. The inundation area at 3 hr is generally captured correctly by all three models while the depths, especially

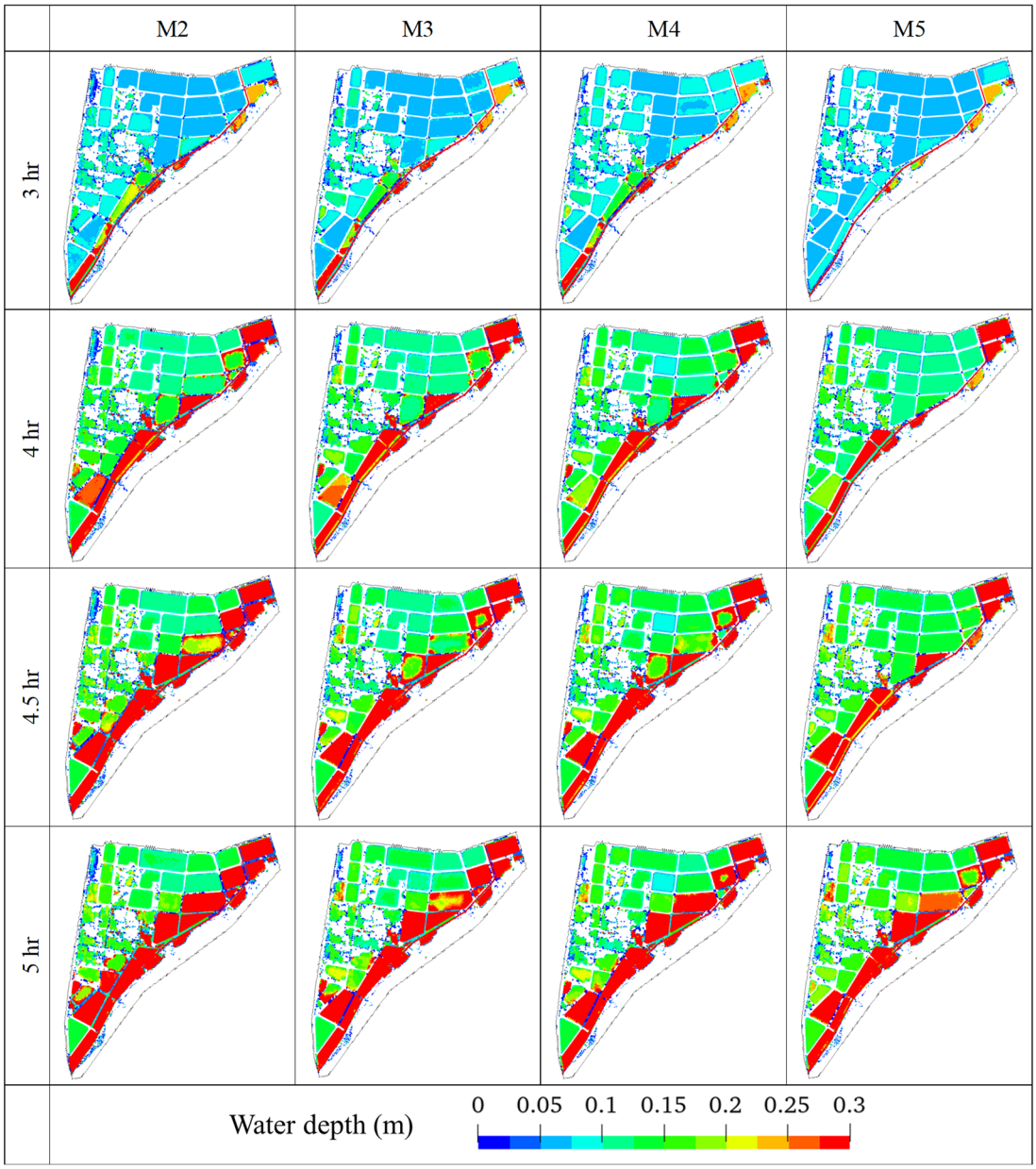


Fig. 5-8 Comparison of predicted water depth contour maps.

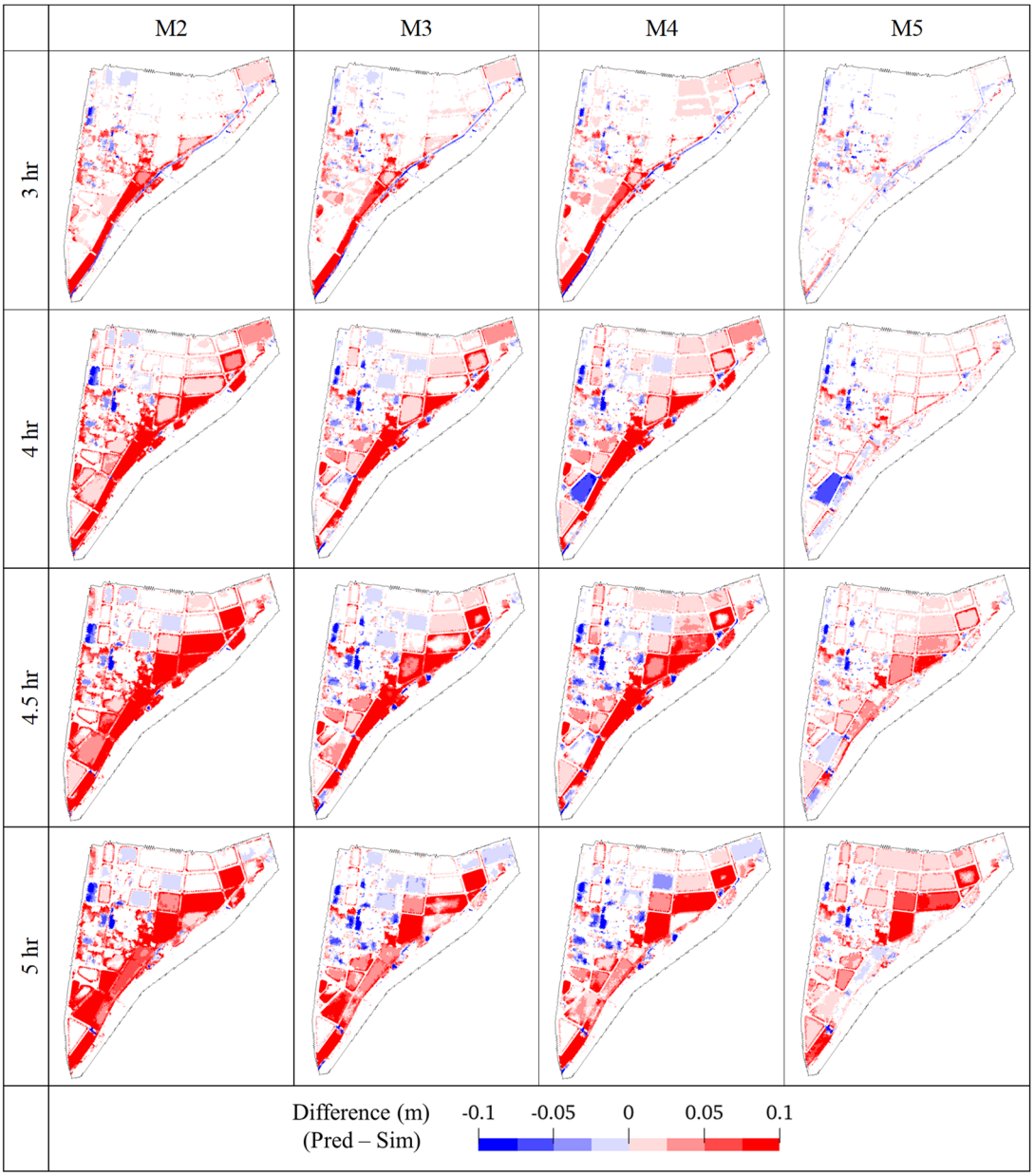


Fig. 5-9 Comparison of difference contour maps.

locations at the downstream, are significantly overestimated. This overestimation accumulates as time progresses due to iteration induced error accumulation, and differentiation of miscellaneous land use at the downstream, having a simulated depth less than 0.3 m, is no longer observed in predictions due to this overestimation. From the difference contour maps, it can be concluded that M3 and M4 slightly outperform M2 as indicated by the lighter color. The feature importance of land use features is summarized in **Fig. 5-10** to compare the influence of land use indicators. The overall importance value of Fr, both prior and current, outweighs the binary indicator of land use. A qualitative evaluation of RMSE in terms of each land use is shown in **Fig. 5-11**. By converting binary indications to temporal varying features,

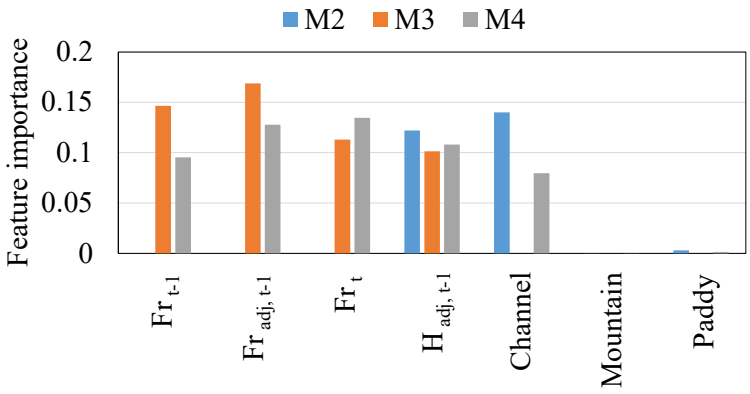


Fig. 5-10 Feature importance of land use features.

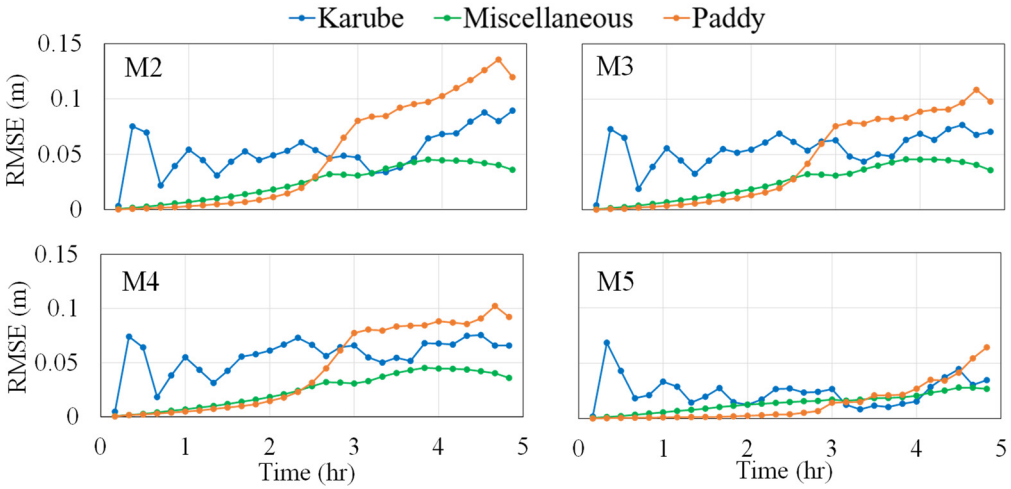


Fig. 5-11 Temporal variation of RMSE with respect to land use.

the RMSE in paddy and miscellaneous decreases especially after 2.5 hr, which is consistent with the one-time shifting logistics of how this coefficient is incorporated in HD models. On the other hand, the difference between M3 and M4 is rather insignificant both in terms of spatial distribution and RMSE. Even though the importance of channel in M4 results in a value of 0.07, which could be considered influential, the prediction performance in the Karube river is not significantly improved compared to M3. Therefore, including both feature indicators of land use is not necessary, and the representation of Fr is adopted for further discussions.

From the water depth contour and difference maps, it is concluded that the water depth predictions are significantly overestimated for M4, and a significant increase in the RMSE of paddy is observed at approximately 2.5 hr when overflow from the Karube region occurs. Also proved in the investigation from the previous chapter, river boundary conditions are important dominating factors for the control volume of the computational domain besides precipitation, and including features reflecting such information improves the predictability of RF models. Previous discussions incorporated the discharge at upstream/downstream of river, which is

now refined to be the water level in M5, considering the availability of observable data. The predicted contour maps, difference maps and RMSE with respect to land use are depicted in **Fig. 5-8**, **Fig. 5-9**, and **Fig. 5-11**, respectively. The predicted inundation area and detailed water depth coincides better with simulated results compared to the other models up until 4.5 hr. The error in the majority of paddies at the final prediction is retained within 0.05 m, with a few paddies along rivers indicating a slightly higher error of 0.1 m. The RMSE over road areas is under 0.03 m during the entire prediction period, and the significant increase of error in paddies is also resolved, again proving that features reflecting overall control volume is necessary in improving prediction accuracy. As for the water depth in rivers, the temporal variation of error is rather stable except for a slightly high initial error. The reasons are first due to the small percentage of meshes of rivers, approximately 3% out of 18,687 meshes, and also the cold start from initial state of normal water level. While the initial state of water level, obtained through runup calculations, follows the natural gradient of riverbed and a greater longitudinal gradient is observed during the early stage of simulation. As illustrated in **Fig. 5-12**, the water level at the downstream rises first in comparison to the upstream due to the wall condition at the downstream sluiceway and propagates backwards. The gradient gradually diminishes over time and the longitudinal water level approximately flattens and increases evenly afterwards starting 2 hr. Therefore, the prediction during the first hour involves longitudinal gradient change and starts from a significant change, both of which leading to a rather notable error even with features representing the boundary conditions.

Finally, a quantitative summarization of inundation area and inundation volume of the models was carried out and illustrated in **Fig. 5-13**. All of the models suggest a trend of overestimation, especially during the first 1.5 hr, yet the error is mostly inferred to be induced by choice of threshold and numerical error since the predictions of corresponding inundation volume are consistent with simulations. Divergence of inundation volume between models is

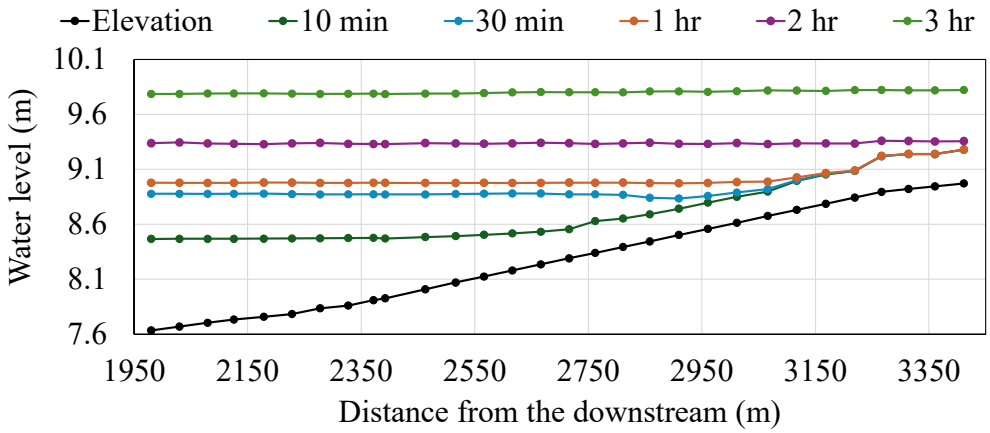


Fig. 5-12 Temporal variation of simulated longitudinal water level of the Karube river.

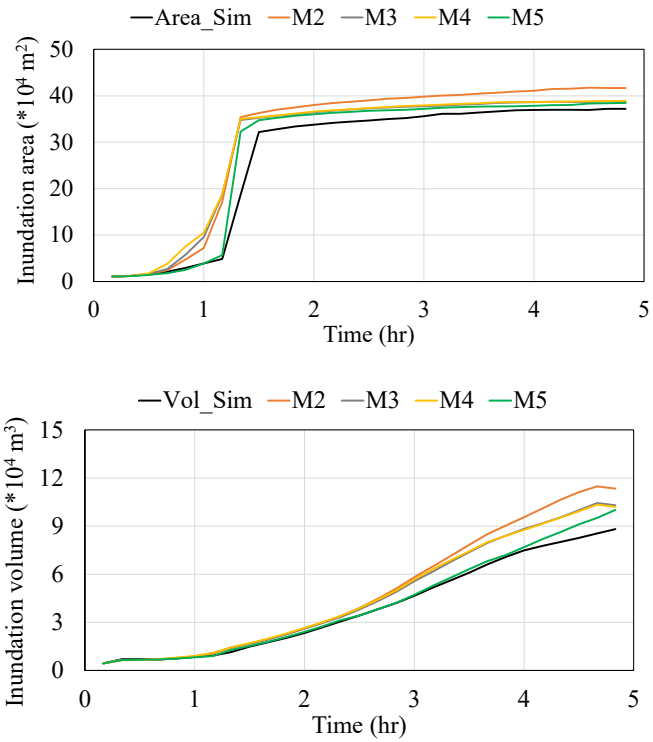


Fig. 5-13 Comparison of simulated and iteratively predicted inundation area and volume.

observed after 2.5 hr. Comparing different indication of land use, the usage of temporal varying features slightly improves model accuracy, and the most effective factor is the inclusion of water level at river boundaries, which contributes to high coincidence to simulation until 4 hr. Although the last hour of prediction of M5 still suggests a trend of overestimation, this level of error is deducted to be induced from both the error accumulation through iteration, and a limited number of training data. Since RF models configurated in this section only used 6 sets of training data, all having the same rainfall distribution, there still exists the possibility of overfitting. Therefore further discussions are performed in the next few section investigating different combinations of training data.

Overall, the following observations are concluded from this section in terms of feature selection. First, direct representation of prior water level information improves prediction accuracy compared to features of elevation and prior depth, even if the physical meaning of the two representations is the same. Features with temporal variations better reflect land uses, especially by reducing RMSE in paddy and miscellaneous after overflow occurs in comparison with models using binary indicators. As an indication for the overall control volume within the system, water levels at river boundaries further improve model accuracy. Even involving iteration process induced error accumulation, the best model is able to provide higher performance predictions with an overall RMSE under 0.05 m.

5.3.2 Discussion on cross-validation method

In the configuration of RF models, how and what portion of the dataset is partitioned into training data and validation data could influence the performance of prediction. So far the model configuration discussed in this study utilized the common splitting method of *train_test_split* function of the sklearn library, which by default assumes independent and identically distributed data. A short investigation was carried out in Chapter 3 regarding the ratio of data split and conclusions were drawn that impact is rather insignificant when the size of training data is large enough. Discussions on cross validation methods, on the other hand, were not performed. Especially considering that rainfall conditions are temporally varying variables, the corresponding variation of water depth is expected to inherit correlations with respect to time, and the assumption for data independence might be compromised. Therefore, investigations were carried out to evaluate how cross validation methods affect model performance.

The basic dataset for model configuration adopted the same amount of data from the previous subsection (rainfall pattern of ND_C with 6 peak intensities) and the features for M5,

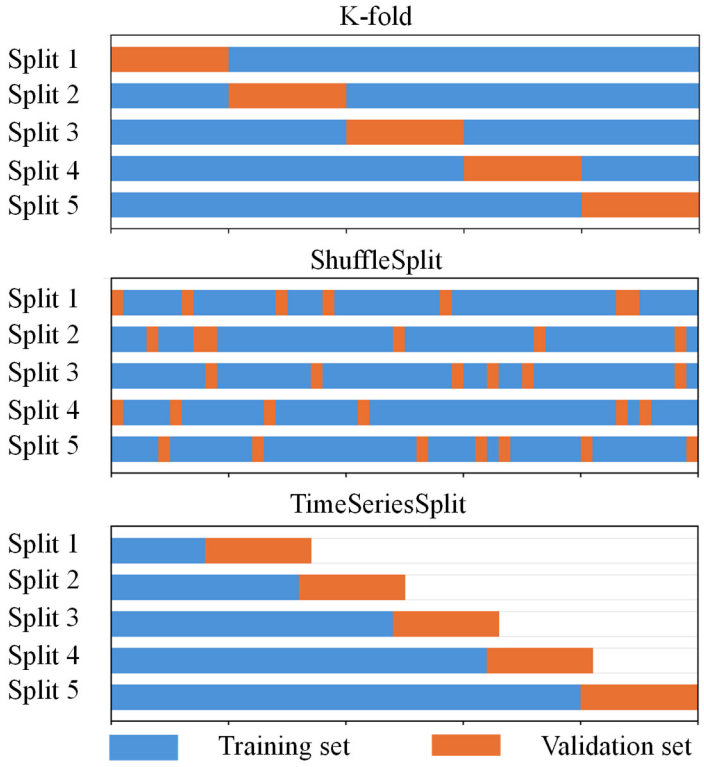


Fig. 5-14 Schematic of cross validation methods.

having the best performance of the investigations in the previous section, were selected. The cross-validation methods investigated here include *train_test_split*, *K-fold*, *ShffleSplit* and *TimeSeriesSplit*[9], and the number of splits were selected to be 5. **Fig. 5-14** illustrates the schematic of each cross-validation method. Since the size of training and validation set is different for each method, direct evaluation of the prediction performance against the validation set may introduce bias. Instead, the evaluation was carried out by testing the models over rainfall events of ND_CR45, which was the event used in the previous subsection. All other conditions regarding hyperparameter setting and iteration predictions process were set to be the exact same condition as the previous subsection. Shown in **Fig. 5-15** are the performances in R^2 and F1-score using different cross-validation methods. It can be observed that both metrics for the 4 methods result in high consistency, with slight diverges during the

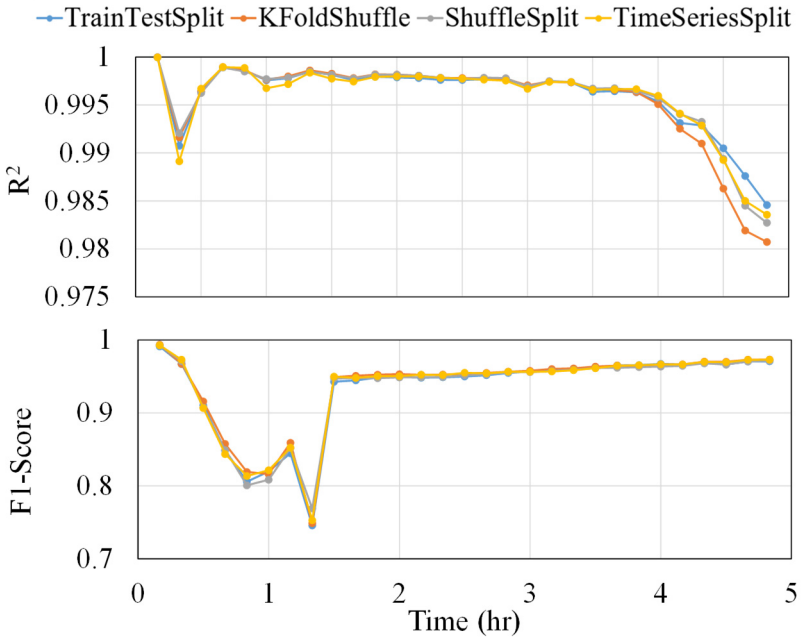


Fig. 5-15 Performance comparison tested against ND_CR45.

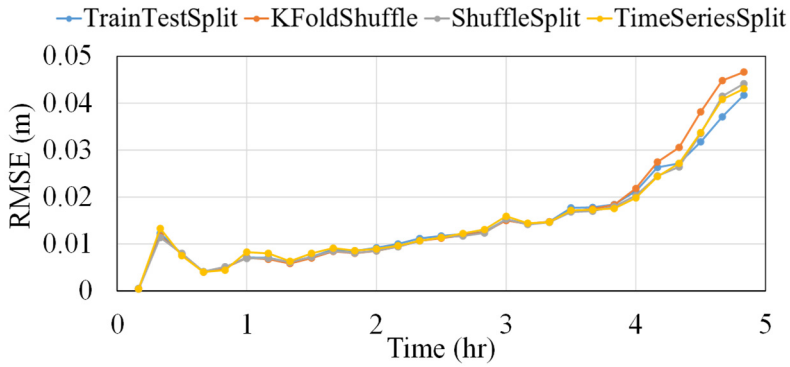


Fig. 5-16 RMSE comparison tested against ND_CR45.

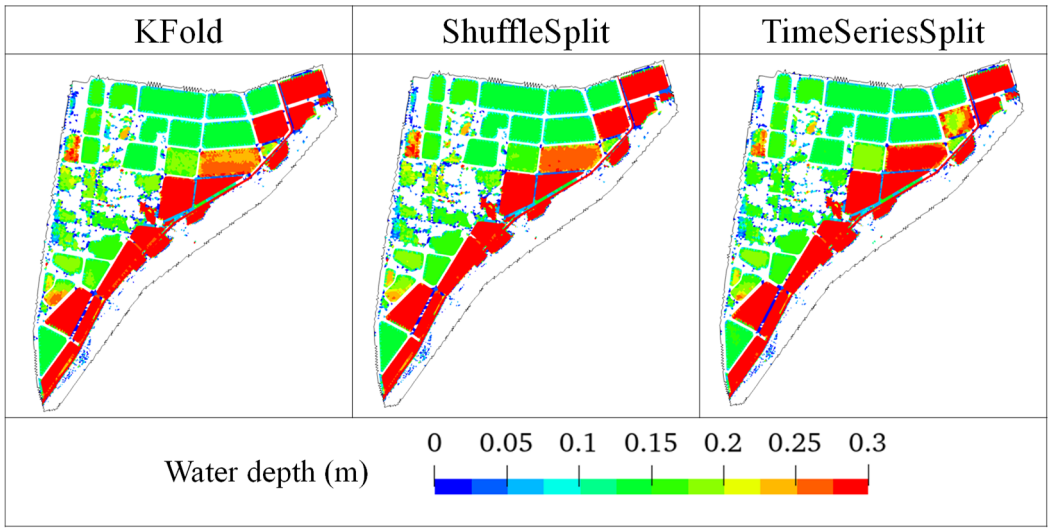


Fig. 5-17 Comparison of water depth contour at 5 hr using different cross validation method.

last hour of prediction in terms of R^2. The temporal variation of RMSE also indicates similar coincidence of results and the *train_test_split* method has slightly better performance than the other three methods (**Fig. 5-16**). The inundation depth contour maps at 5 hr using different cross-validation methods is depicted in **Fig. 5-17**. Despite small variations in a few paddies, the prediction of both inundation area and depth also coincides highly with the results of the original *train_test_split* methods (**Fig. 5-8**), and the performance of *TimeSeriesSplit* method is not significantly improved. Indication for this observation is that the influence of timeseries correlation is already reflected through the selection of features. The prior state features, as being carefully investigated with both numerical experiments and applications under actual terrains, preserves the lagged information necessary to capture the temporal dependence within the samples at each time step. That is, the original sequentially dependent datasets are approximately transferred to be independent ones with the use of prior information as features. The influence of correlation across timeseries becomes less significant, resulting in similar result against test events. Therefore, the following discussions still take advantage of the assumption of independency in samples and use the original *train_test_split* method. On the other hand, the influence of temporal correlation might become significant when testing over rainfall events with significantly different characteristics, different shapes and distribution of peak for example. Therefore investigations are necessary to evaluate the feasibility of the current configuration against other patterns of rainfall.

5.4 Selection of rainfall patterns in training data

Similar to other data-driven approaches, the performance of RF models is highly influenced by the quality, range and diversity of conditions represented in the training data. This section further examined how the selection of training data affects the performance of RF models when tested against other rainfall patterns. To be specific, different combinations of rainfall in terms of shapes and peak distribution of peak intensity were selected for the configuration of models and the predictive performance was tested against multiple hypothetical rainfall events. The peak intensity in training data was the same 6 patterns ranging from 20 mm/hr to 70 mm/hr. As summarized in **Table 5-6**, a total of 9 models were configurated to evaluate how the inclusion or exclusion of particular rainfall patterns affects the capability to represent diverse conditions and predicts under scenarios with similar or different characteristics from the training data. Based on the conclusions from previous investigations, features and cross-validation methods for model configuration adopted the conditions of M5 in section 5.3.1 and *train_test_split* method. In other words, ND1 was the exact same model as M5 and was used as a baseline for model evaluation. Since the total time for the calculation of each rainfall pattern was 5 hr (a total of 30 time steps), the E and L patterns of peak time, as previously defined in **Equation 5.1** and **Table 5-4**, correspond to a peak time at 2 hr and 3 hr respectively.

5.4.1 Impact of peak time of rainfall

Investigations were first conducted by performing predictions against rainfall events of the same shape. Models ND1 through ND4 were tested against rainfall ND_CR45 to evaluate whether increasing the number of training data to include rainfalls of different characteristics

Table 5-6 Combination of different selections of training data for model configuration.

Model name	Rainfall patterns in training data		Total time steps in training data	Model size (GB)
	Shape	Peak time		
ND1	ND	C	180	1.59
ND2		C + E	360	2.31
ND3		C + L	360	2.19
ND4		C + E + L	540	2.69
NND1	NND	C	180	1.46
NND2		C + E	360	2.12
NND3		C + L	360	1.99
NND4		C + E + L	540	2.47
CMB4	ND + NND	C + E + L	1080	3.45

of peak time affects the prediction accuracy compared to the original model.

Illustrated in **Fig. 5-18** are the water depth contour and difference maps at the final step of prediction. The corresponding R^2 and F1-score for the baseline model ND1 was 0.985 and 0.971 respectively. Comparing to ND1, the performance metrics were increased for all three models, yet the spatial distribution of prediction varies between models. For example, while having the lowest overall performance of the three models, ND3 resulted in similar high accuracy in predicting the water depth of the paddy at the upper right corner and long narrow shaped paddy as ND1. In comparison, ND2 and ND4 indicated underestimations in these paddies. On the other hand, significant overestimations were also observed at other locations for ND3, suggesting that the models have higher variance in prediction error. When quantitively evaluated through RMSE, it can be observed that the accuracy of models was highly consistent except for the last hour especially in terms of paddy and the Karube river

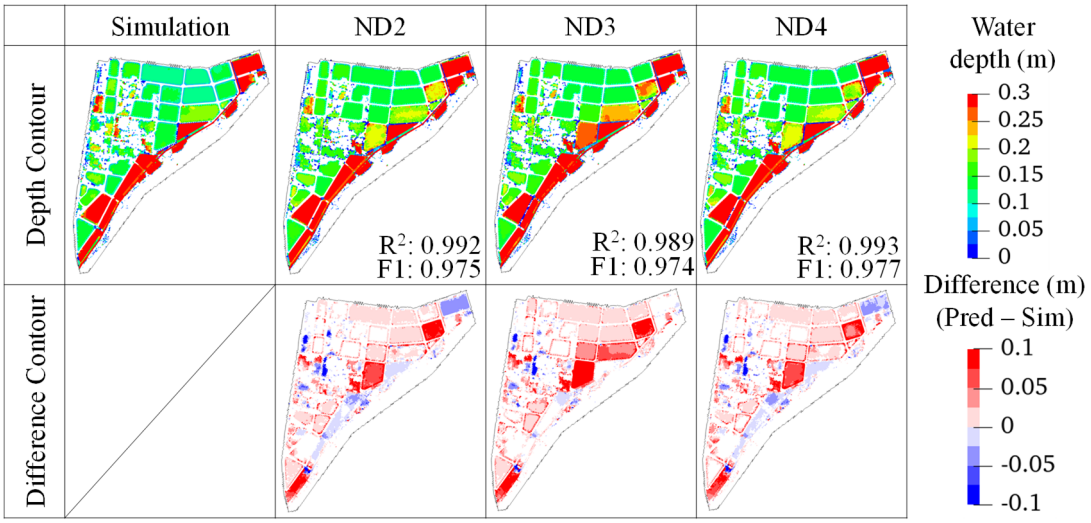


Fig. 5-18 Water depth and difference contour maps for ND_CR45 predictions at 5 hr.

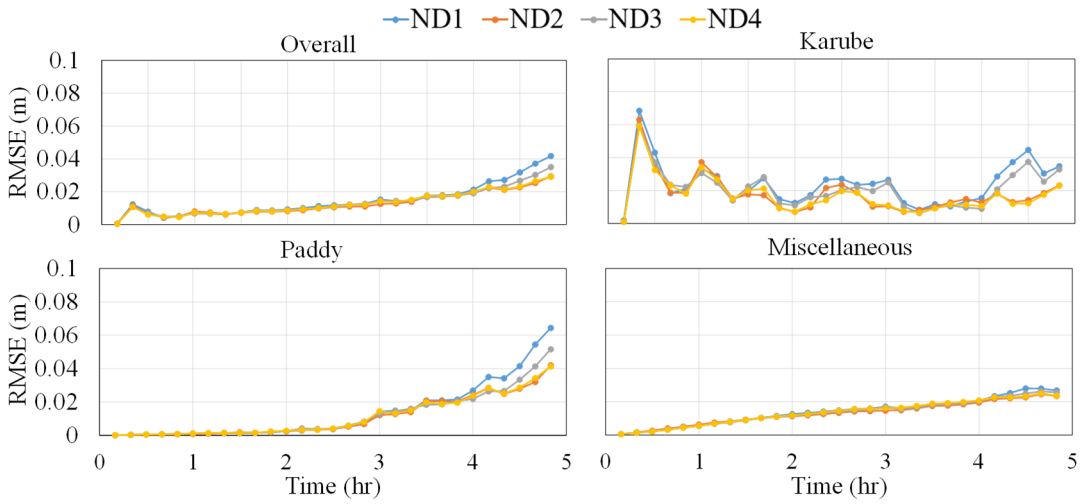


Fig. 5-19 Comparison of RMSE of different ND models predicting on ND_CR45.

(**Fig. 5-19**). The improvement of prediction accuracy from overall RMSE of 0.04 m to 0.03 m in ND4, however, is not significant considering that the number of training data is tripled.

This finding is further supported by testing the models against rainfall patterns with other peak times. For example, the models were further testing against ND_ER45 and ND_LR45, with overall RMSE shown in **Fig. 5-20**. Although it is not reasonable to compare the values directly since inundation depths are different depending on rainfall, the general trend of temporal variation is observed that the models without training data of similar characteristics resulted in compromised accuracy, especially during the first few hours. And the high error is mostly induced from the prediction of depth in rivers. Shown in **Fig. 5-21** are the RMSE of the Karube river for the two rainfall patterns respectively, and the error during the first few predictions is extremely significant. The performance of models also varies between rainfall patterns and even the best model ND4 resulted in significantly high RMSE greater than 0.15 m in prediction of water level in rivers for rainfall pattern ND_ER45. As discussed previously, the longitudinal water level rises first at the downstream then propagates upstream and the time reaching flatten water surface gradient varies between rainfall patterns. For rainfalls with the same shape and peak intensity, front-loaded rainfalls lead to a significant temporal variation in water level. **Fig. 5-22** illustrates the temporal variation of simulated longitudinal water level of ND_ER45 and ND_LR45 respectively. Compared with that of ND_CR45 depicted previously in **Fig. 5-13**, the ND_ER45 pattern already reaches leveled water surface

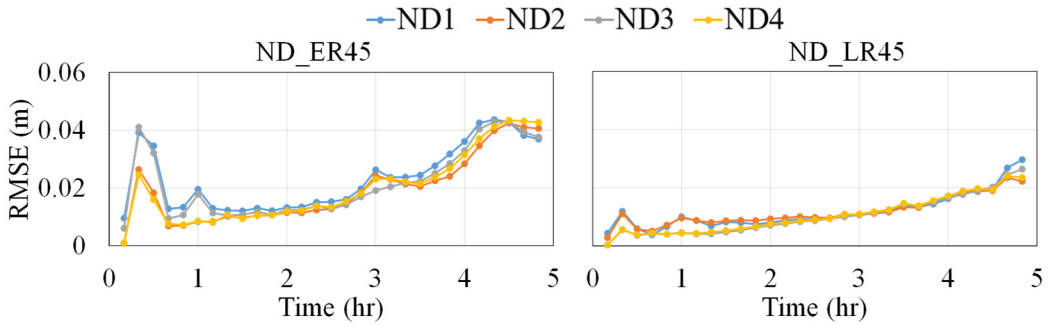


Fig. 5-20 Comparison of overall RMSE of ND models predicting for ND_ER45 and ND_LR45.

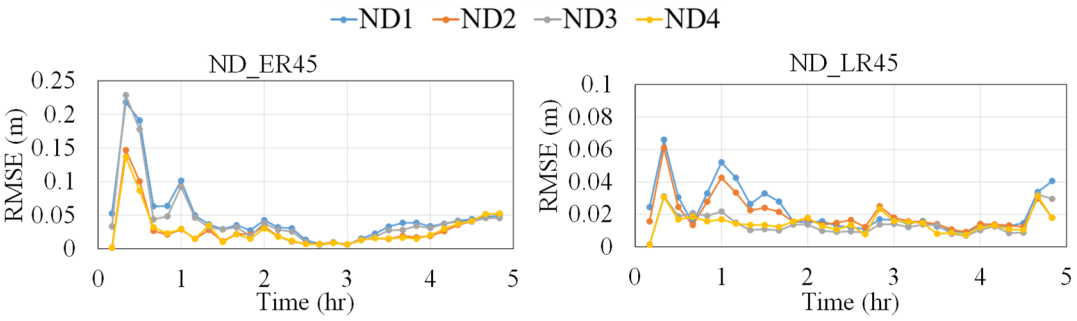


Fig. 5-21 Comparison of RMSE over the Karube river predicting for ND_ER45 and ND_LR45.

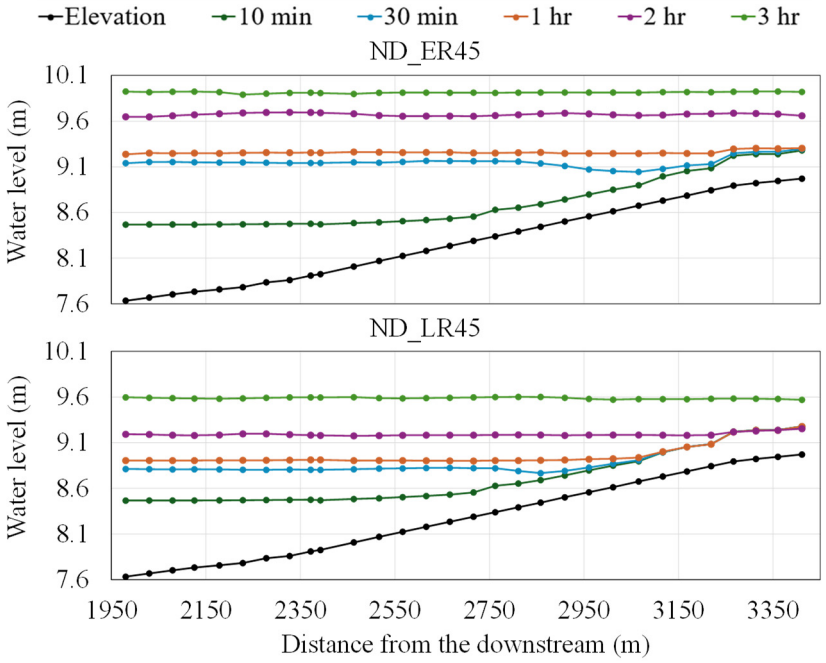


Fig. 5-22 Temporal variation of simulated longitudinal water level of ND_ER45 and ND_LR45.

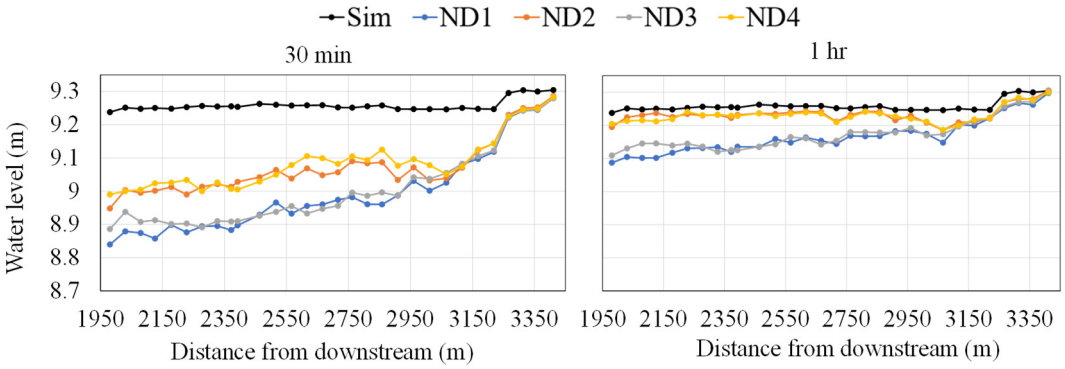


Fig. 5-23 Temporal variation of predicted longitudinal water level for ND_ER45.

approximately 30 min after rainfall starts while the other two rainfall patterns require somewhere between 1 hr and 2 hr before the gradient diminishes. **Fig. 5-23**, **Fig. 5-24** and **Fig. 5-25** are the temporal change of longitudinal water level, illustrating both simulated and predicted results, of three rainfall patterns respectively. As shown in the figures, all of the RF models tend to predict for sloped water level during early stage of a rainfall event regardless of the number of training data. Yet due to the extremely shortened period before leveling, the predictions of ND_ER45 are significantly underestimated towards the downstream, especially for ND1 and ND3 models that did not incorporate similar front-loaded patterns of training data. The predictions then reach flattened longitudinal gradient and the difference between prediction and simulation gradually diminishes, leading to the significant RMSE during first few steps. In contrast, the variation in the longitudinal water level of ND_CR45 and ND_LR45

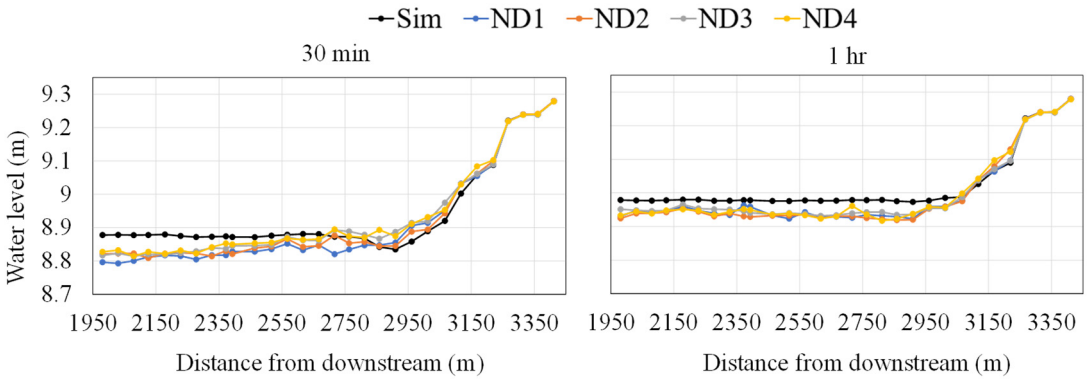


Fig. 5-24 Temporal variation of predicted longitudinal water level for ND_CR45.

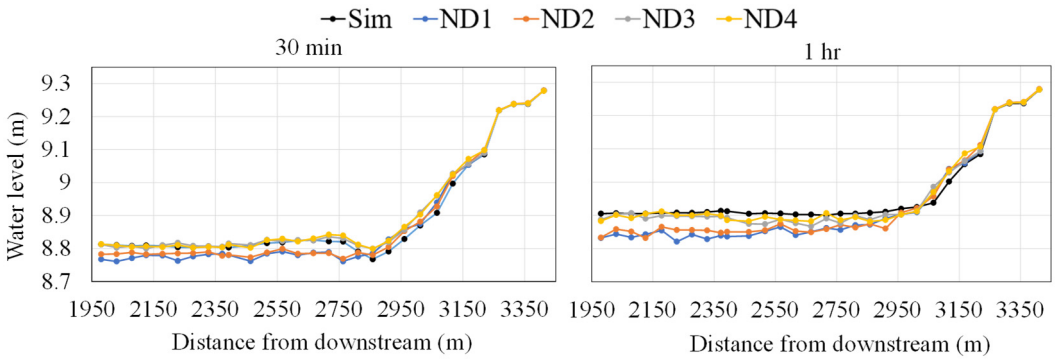


Fig. 5-25 Temporal variation of predicted longitudinal water level for ND_LR45.

patterns are rather smoothened over time, and the predictions of RF models coincide better with simulations.

Both having an initial water depth of 0 m and gradually accumulation of inundation depth, miscellaneous and paddies resulted in a different trend of temporal error variation. First depicted in **Fig. 5-26** are the temporal variation of RMSE over miscellaneous for ND_ER45 and ND_LR45. Together with the results for ND_CR45 summarized in **Fig. 5-19**, it can be concluded that prediction performance between models coincide highly regardless of prediction event. The incremental error accumulation over time is also rather constant without

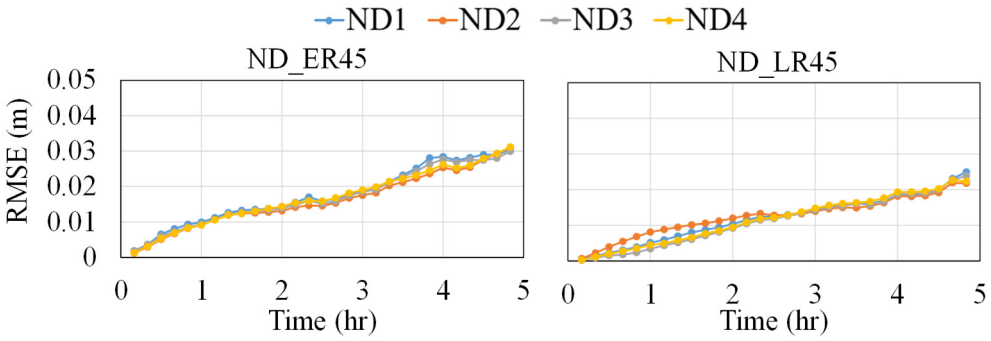


Fig. 5-26 Comparison of RMSE of miscellaneous for ND_ER45 and ND_LR45 rainfalls.

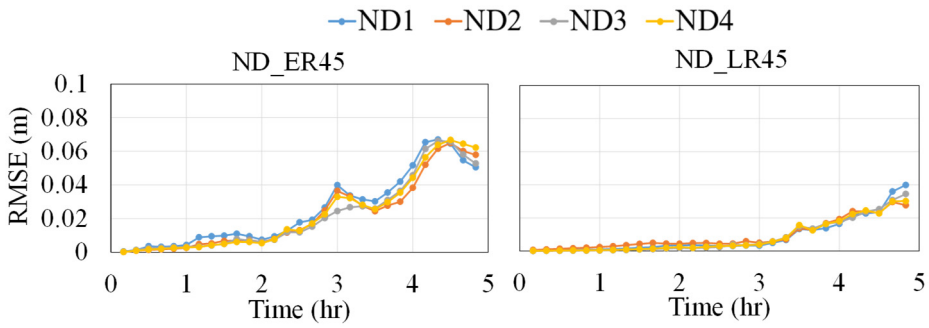


Fig. 5-27 Comparison of RMSE over paddy for ND_ER45 and ND_LR45 rainfalls.

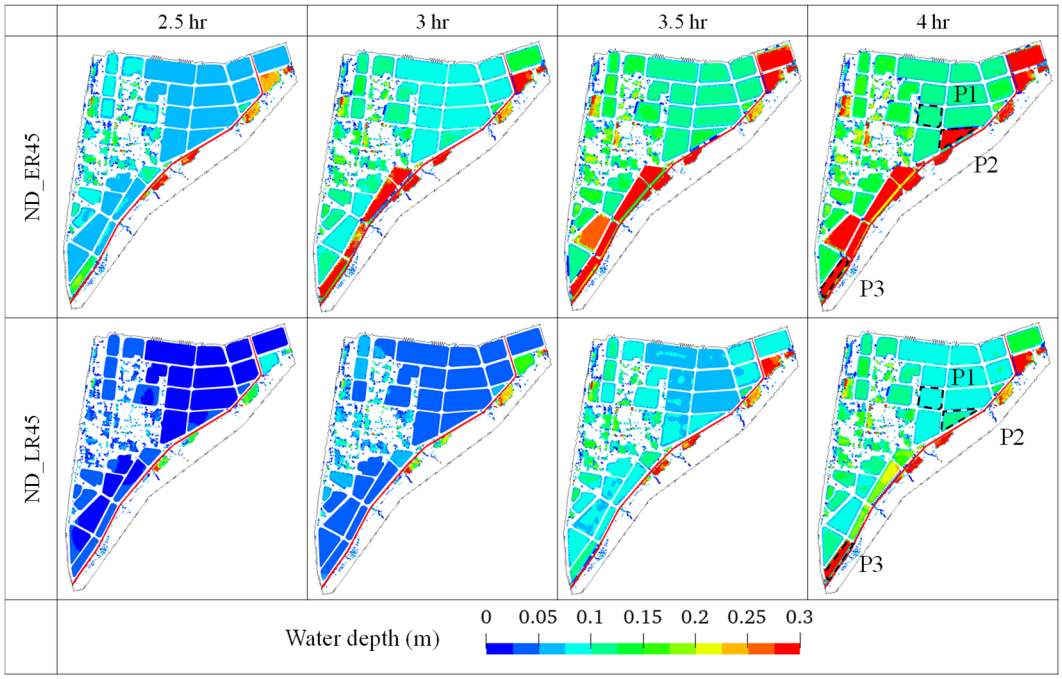


Fig. 5-28 Simulated water depth contour under rainfall ND_ER45 and ND_LR45.

sudden variations, and all models indicate prediction error starting the first prediction. On the other hand, the temporal variation of RMSE over paddy, as shown in **Fig. 5-27**, starts with a low value of less than 0.01 m and maintains the high performance for a certain period of time until a significant increase in incremental error is observed. The time for the sudden change in incremental error varies between target rainfall event. The peak times for front-loaded and back-loaded rainfall patterns were shifted 30 min with respect to centered rainfall to be 2 hr and 3 hr respectively. As illustrated as the contour maps for simulated water depths of ND_ER45 and ND_LR45 in **Fig. 5-28**, it can be observed the overflow at downstream occurs approximately 30 min after the peak time of rainfall, which corresponds to the time when significant increase RMSE of paddy starts to be notable. The corresponding water depths at the final step also varies significantly depending on peak time as shown in **Fig. 5-29** and **Fig. 5-30** respectively. It is observed that the prediction performance decreases as the water depth

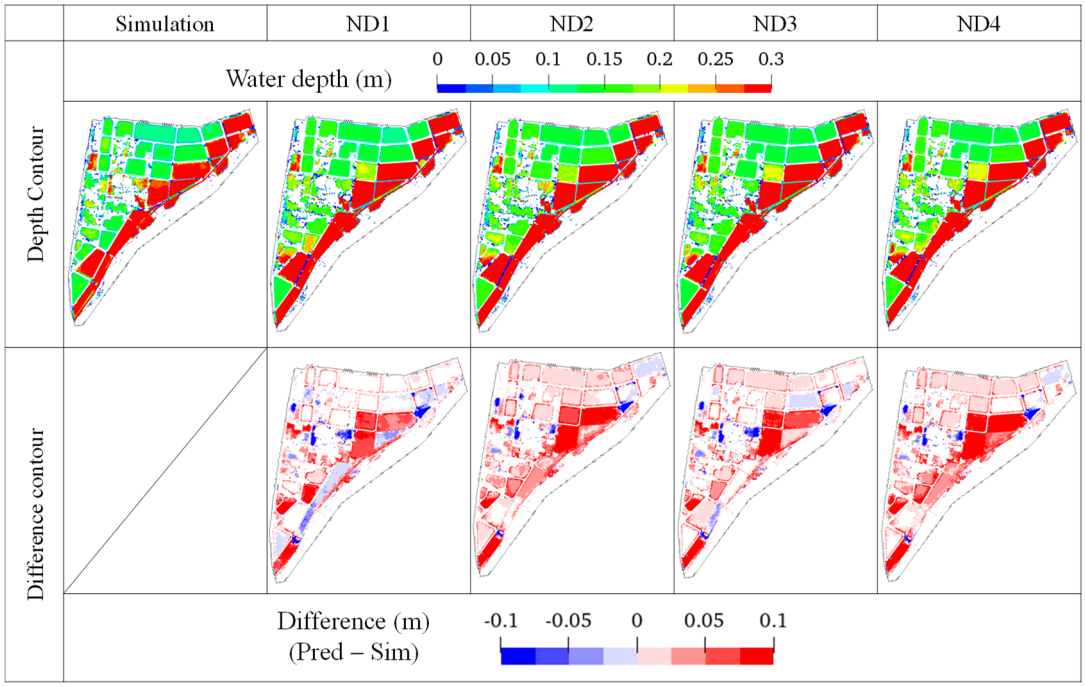


Fig. 5-29 Water depth and difference contour maps for ND_ER45 predictions at final step.

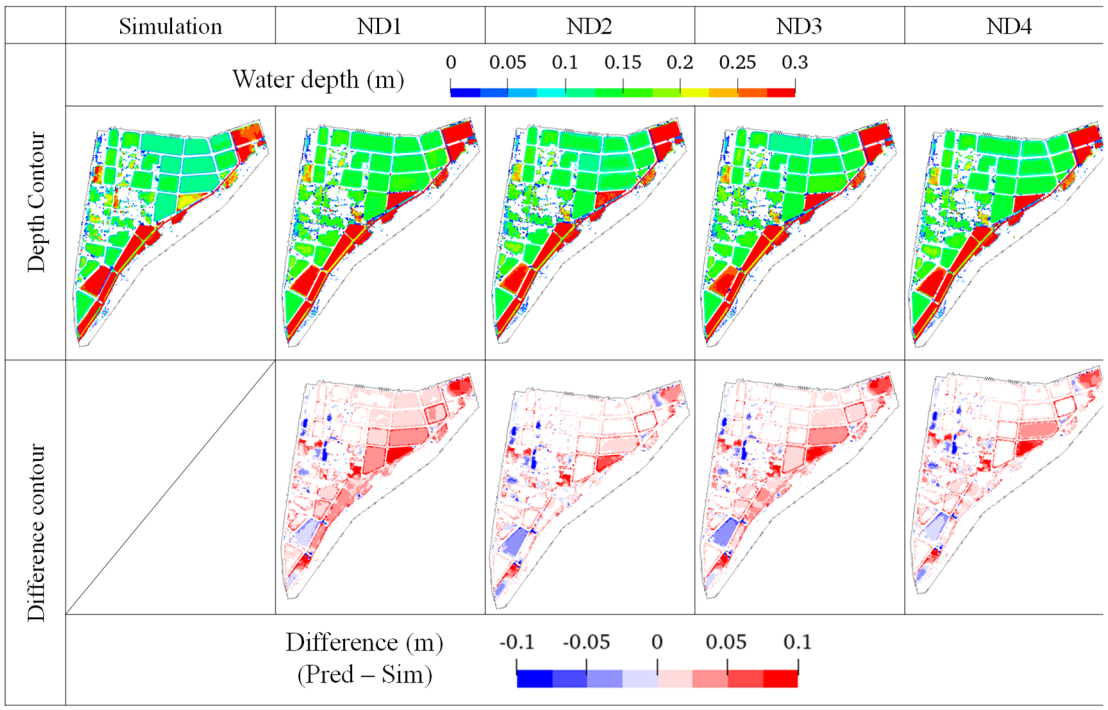


Fig. 5-30 Water depth and difference contour maps for ND_LR45 predictions at final step.

increases and variance of error between paddies is also more notable for rainfall events with greater inundation even using ND4 that had the greatest number of training data covering all types of hypothetical rainfall patterns. Indications are that overflow from the Karube region

introduces another dominating factor for the rise of water depth other than precipitation for the paddies located close to river. Another reason is that RF model performs better regarding interpolations, and the performance of models tends to decrease as the target event approaches the upper limit of the training range. To compare in detail how the performance differs between rainfall patterns and between paddies, three of the paddies (enclosed with dashed black lines in **Fig. 5-28**) were selected with the following consideration: P1 locates distant from the Karube river and was not influenced by overflow regardless of rainfall pattern, P2 suffered certain overflow during the last few steps, and P3 received most significant effect of overflow for all rainfall patterns. **Fig. 5-31** plots a comparison of average water depths of each paddy as obtained from simulation and prediction using ND4 for different target rainfall events. With the height of paddy ridge set to be 0.3 m, water depths greater than this threshold indicate overflow from paddy occurs. For all three rainfall patterns, the prediction is highly consistent with simulation results when the water depths are shallow when the dominating factor is direct rainfall. Divergence of results for P2 and P3 starts to become notable when the water depth increases significantly, and it can be observed that the time for this increases corresponds to the time when the paddy receives influence from overflow. On the other hand, the performance

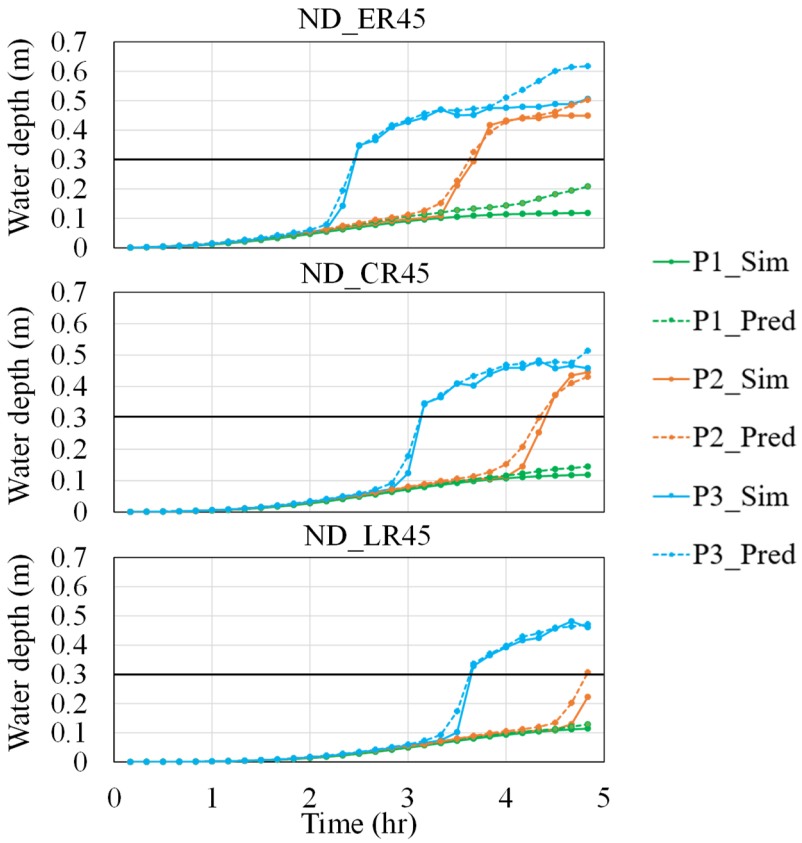


Fig. 5-31 Comparison of average simulated and predicted water depth of selected paddies under different rainfall conditions.

regarding the last hour for ND_CR45 and ND_LR45 are rather high while that for ND_ER45 is significantly compromised even for P1 that is not affected by overflow. The reason contributing to the difference between rainfall patterns is inferred to be the incapability of RF model to make extrapolations. As also concluded by previous literatures with applications in multiple disciplines, the performance of RF models degrades when making predictions near or beyond the range of training data due to its non-parametric ensembled nature[10), 11), 12)]. As a result, the long-lasting overflow from the Karube river of ND_ER45 induced larger inundation volume compared to other rainfall patterns, and the number of training data exceeding this scale is limited, both leading to a degraded performance in the prediction of the last hour.

Similar trends of performance variance between models and between prediction events are observed when using NND1 through NND4 to test rainfalls of NND_CR45, NND_ER45, and NND_LR45. **Fig. 3-32** through **Fig. 3-34** plots the overall and land-use-wise RMSE of the

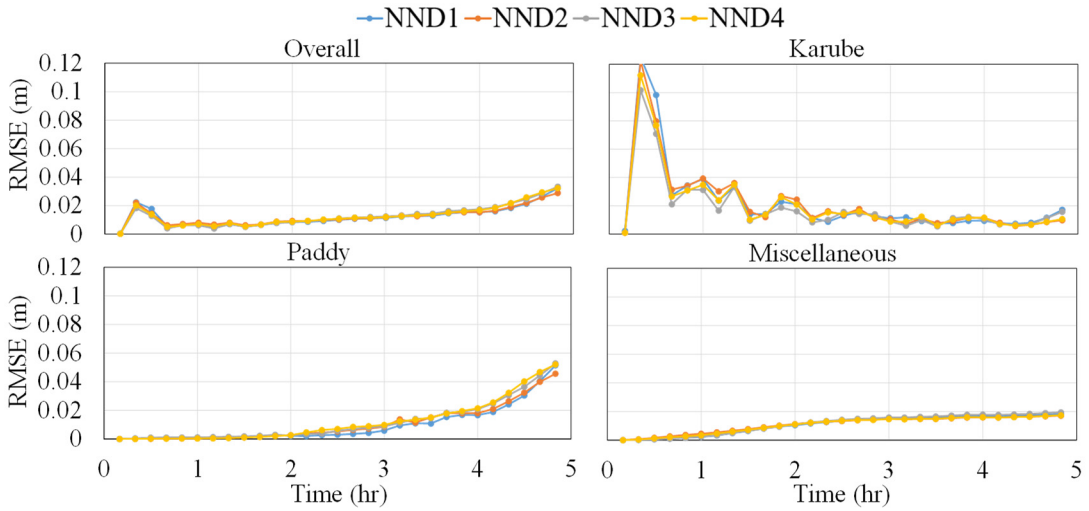


Fig. 5-32 Comparison of RMSE of different NND models predicting on NND_CR45.

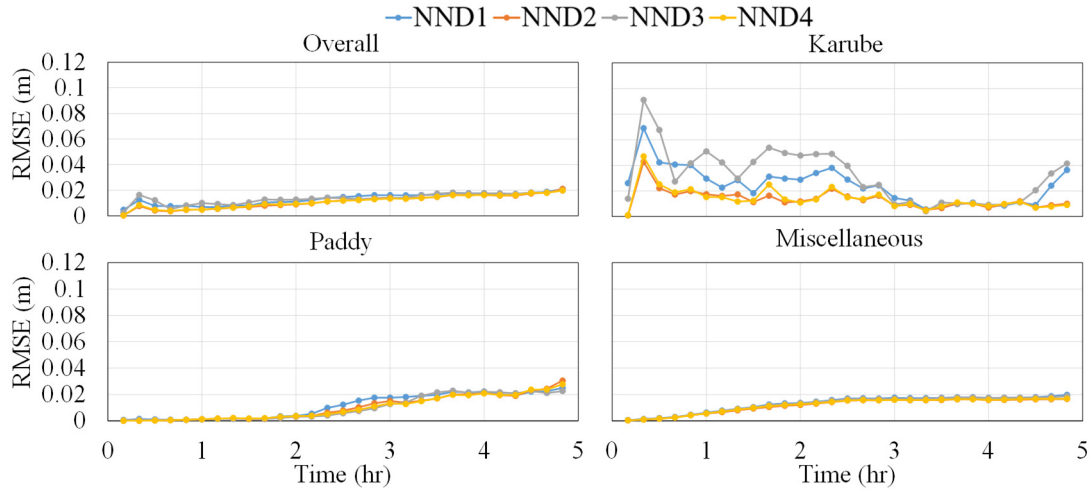


Fig. 5-33 Comparison of RMSE of different NND models predicting on NND_ER45.

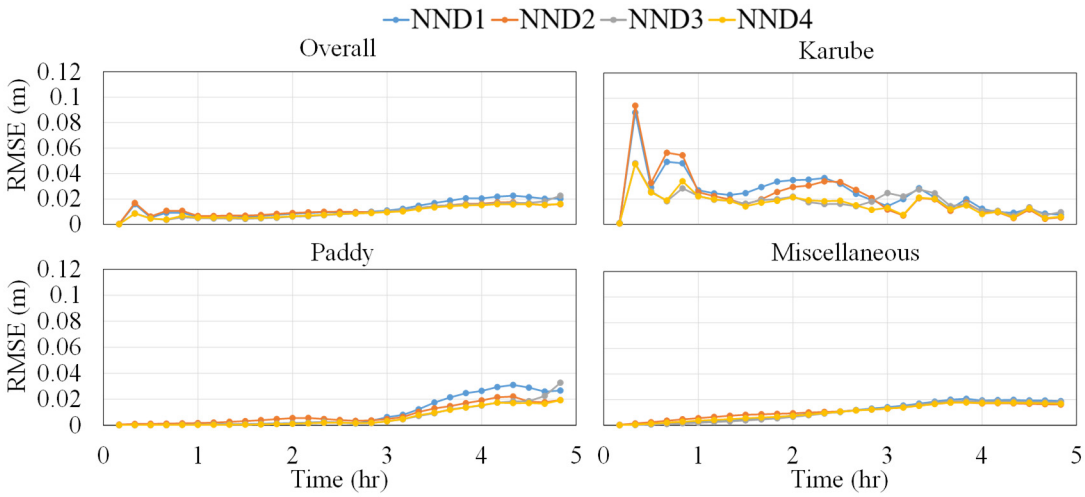


Fig. 5-34 Comparison of RMSE of different NND models predicting on NND_LR45.

predictions for the respective rainfall events. When predicting for NND_CR45, the results evaluated as RMSE highly coincide with each other. But the prediction over the Karube river for rainfalls with shifted peak time showed degraded performance of models configurated without similar rainfall patterns in the training data. Compared with the results predicting for ND rainfalls, the differences between models are even less notable since the overall inundation is shallower due to small amount of accumulative rainfall. Yet both sets of comparisons suggest that models ND4 and NND4, configurated with the largest number of training data that covers different rainfall patterns of peak time, resulted in the highest performance compared with other models, suggesting the necessity of including rainfalls with multiple peak times.

5.4.2 Impact of rainfall shapes and peak intensity

As stated in the previous section, one common limitation of RF models is the degraded generalization capability when predicting for events approaching or exceeding the range of training data. Therefore this section carried out further evaluation using the configurated models to predict for rainfalls with different shapes, that is using ND models to predict NND-shaped rainfall and vice versa, and then rainfalls with peak intensities closer to the upper limit of the training data, 55 mm/hr and 65 mm/hr specifically. Based on previous performance evaluation, this results comparison was conducted using the high performance models ND4 and NND4 for simplicity.

The predicted results using ND4 to predict for untrained rainfall shapes NND_CR45 in terms of RMSE are illustrated in **Fig. 5-35**. Compared with the performance evaluations predicted with NND4 as depicted in **Fig. 5-32**, ND4 resulted in significantly high RMSE

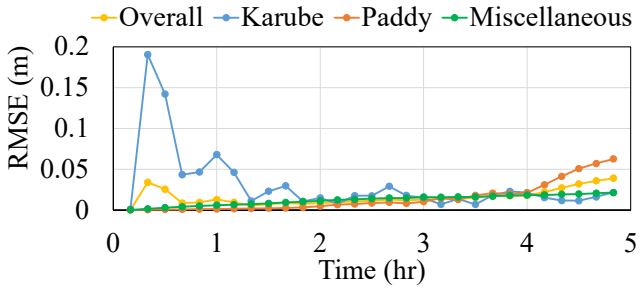


Fig. 5-35 RMSE of ND4 testing over rainfall NND_CR45.

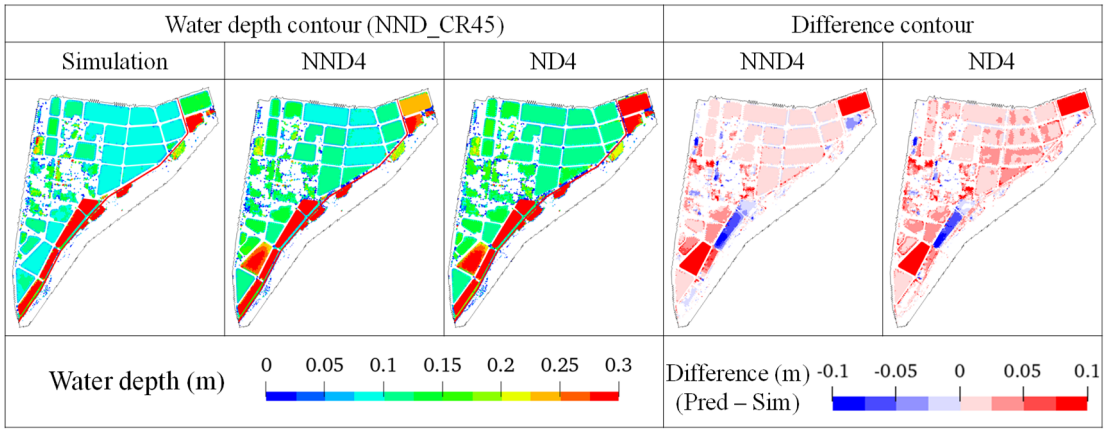


Fig. 5-36 Comparison of water depth and difference contour maps for final step of NND_CR45.

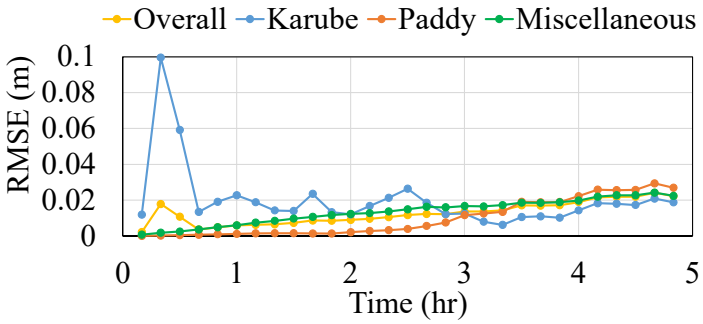


Fig. 5-37 RMSE of NND4 testing over rainfall ND_CR45.

approaching 0.2 m for the Karube river during the first few predictions. The RMSE for paddy for final steps using ND4 is slightly higher than that predicted using NND4, yet the overall performance is rather similar. **Fig. 5-36** shows a comparison of the water depth and difference contour maps at the final step, and it can be observed that the spatial distribution of error between the two models is also highly consistent. Similar trend is also observed using NND4 to predict for ND_CR45 that the prediction performance coincides with predictions using models trained with the same rainfall shapes. The RMSE comparison of **Fig. 5-37** and previous results shown in **Fig. 5-19** suggests that the differences between model predictions are not significant except for the high initial error over the Karube region. The contour maps at the

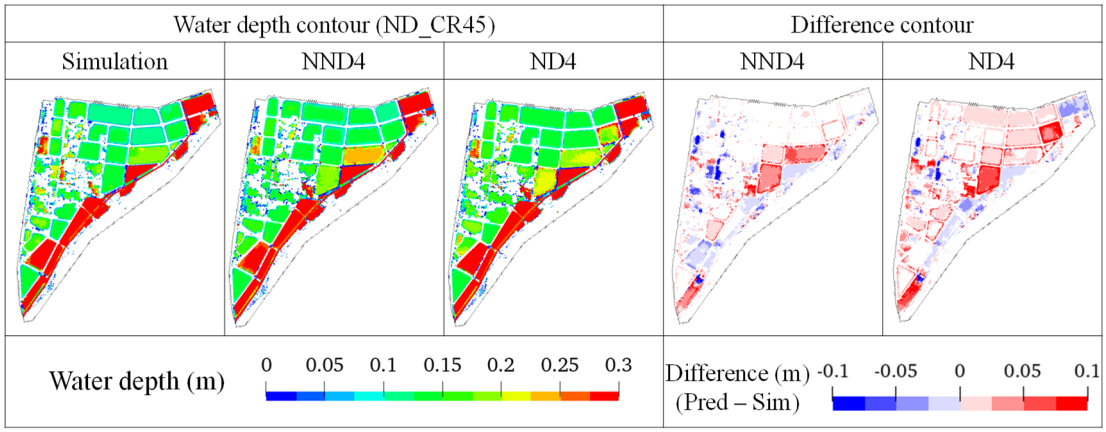


Fig. 5-38 Comparison of water depth and difference contour maps for final step of ND_CR45**.**

final step also provide evidence for this observation (**Fig. 5-38**), and the NND4 model, trained with concentrated rainfall shapes, even resulted in slightly better predictions in certain paddies. This observations, while slightly against the common findings of RF models, could be explained by the overlapped range of training data and selection of target rainfall event. The cumulative rainfall covered in training data for ND4 ranges from 49 mm to 173 mm, and that for NND4 ranges from 35 mm to 124 mm. And the cumulative rainfalls for the prediction events of ND_CR45 and NND_CR45 are 111 mm and 79 mm, respectively. In other words, the cumulative rainfalls of the target events fall within the range of training data of both models, which resulted in similar results at the final stage of prediction, while the limited capability to be applied for untrained patterns is reflected by the degraded performance of rivers during the first few predictions.

This limitation is more notable as the prediction events approach or exceed the range of training data. The ND4 and NND4 models were then tested against rainfall event of ND_CR65. With a peak intensity of 65 mm/hr, the cumulative rainfall for the target event reaches 160 mm, which is close to the upper range of training data of model ND4 and already exceeds that of

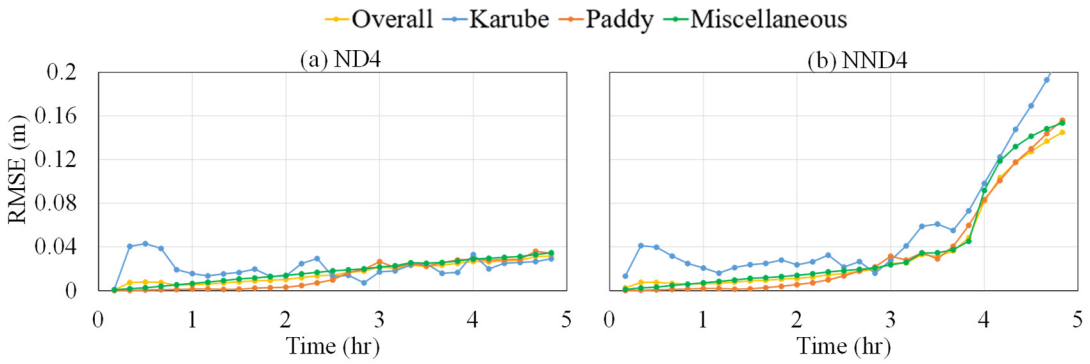


Fig. 5-39 Model performance in terms of RMSE tested over rainfall ND_CR65 predicted with (a) model ND4 and (b) model NND4.

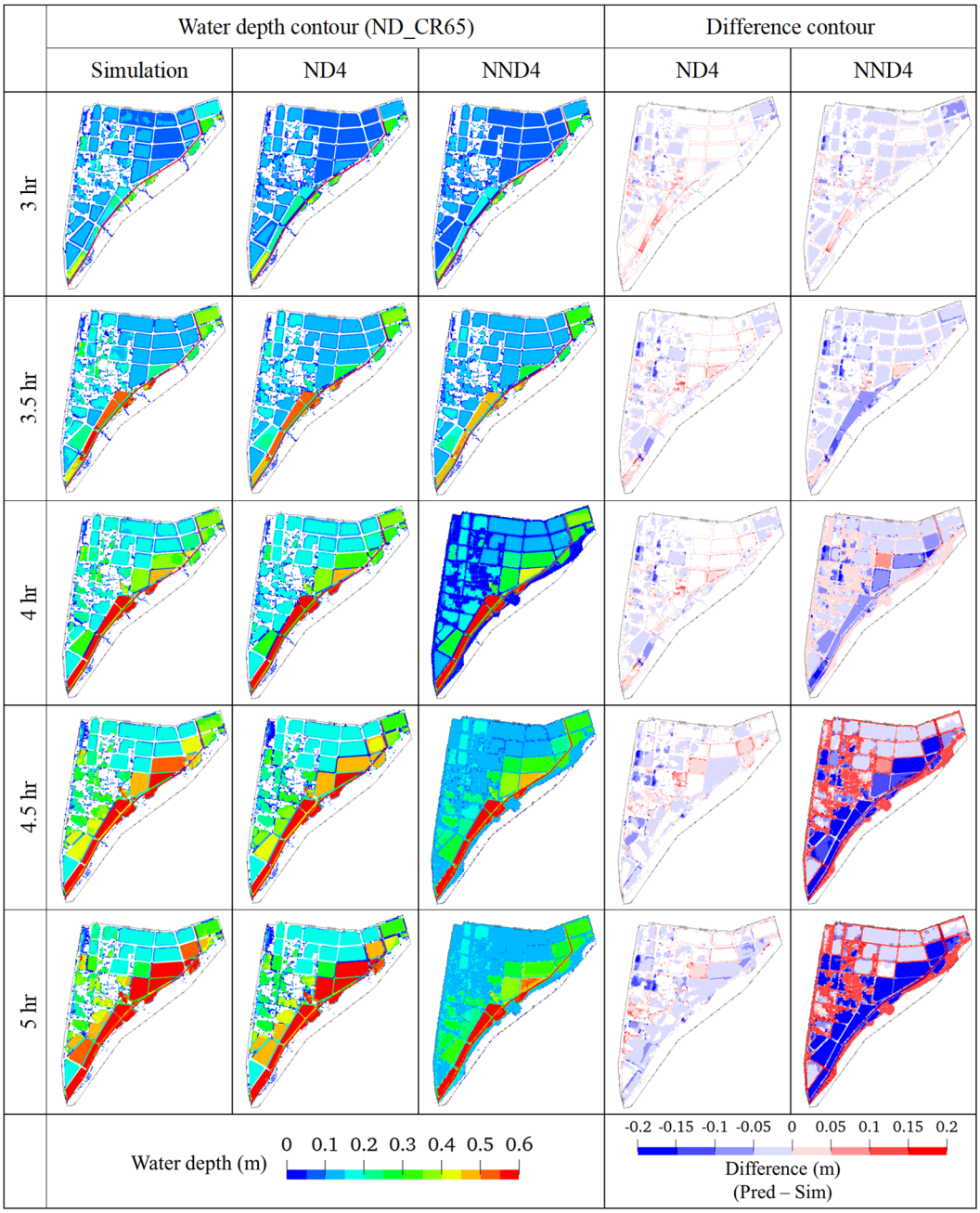


Fig. 5-40 Temporal variation of water depth and difference contour maps for rainfall pattern of ND_CR65 predicted by different models.

model NND4. The RMSE performance using the two models is illustrated in **Fig. 5-39**. The performance of the two models during the first 3 hr is rather similar to each other, with a slightly higher RMSE in the Karube river for NND4. On the other hand, the performance significantly diverges afterwards. The temporal error accumulation of ND4 is still gradual without any abrupt change while NND4 results in a notably steep gradient in the increase of RMSE for all land uses and leads to a final overall RMSE of 0.14 m. The temporal variation

of the simulated and predicted water depth contour and the difference contour maps are illustrated in **Fig. 5-40**. Based on simulation results of HD models, overflow from the Karube river occurs at approximately 3 hr and the paddies located along the downstream of river reaches full capacity due to this overflow. Due to the large peak intensity, large volume of water inflows from river boundaries and flood propagates fast with time such that large areas of miscellaneous land use also suffered inundation at 5 hr. The depth contour of ND4 overall captures how the flood propagates and error for water depth prediction is mostly suppressed within 0.05 m, with slight underestimation over paddies during the last hour. In comparison, NND4 indicated a tendency of underestimation for almost all paddy areas and the extent of underestimation intensifies as time progresses. Significant compromised predictions over miscellaneous land use are also observed, both in terms of classifying inundated areas and predicting for water depths. The majority of roads between paddies is misclassified to be inundated starting 4 hr, and the corresponding error for the prediction of water depth accumulates with maximum absolute difference exceeding 0.2 m for the final prediction. And the significant divergence of predicted values to simulation results observed in NND4 is induced mostly by the limited capability for extrapolation rather than error accumulation of the iterated prediction process. Shown in **Fig. 5-41** is the quantitative comparison of simulated and predicted inundation area and volume. Notations of ND4_Ite, NND4_Ite and NND4_StepPred denote results of iterated prediction using ND4, iterated prediction using NND4 and one-step prediction of NND4, that is assuming the correct values of prior state information are provided at each step. With slight overestimation in inundation area, ND4 overall generates highly consistent predictions with simulated results even with the

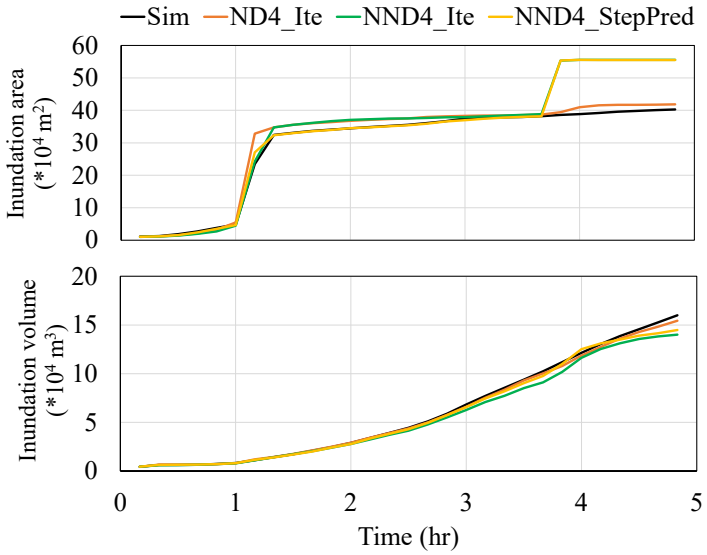


Fig. 5-41 Comparison of simulated and predicted inundation area and volume with different models and prediction approaches.

accumulation of error over time. NND4 also resulted in a similar high performance during the first 3 hr. Comparing the results with and without iteration, it can also be concluded that the overestimation in predicting inundation area is due to error accumulation. Induced mainly through direct precipitation, the inundation depth during this period is still shallow and the corresponding predicted inundation volume do not diverges significantly compared to simulation results. A significant overestimation of more than 150,000 m^2 in inundation area is observed after 3.5 hr, and this error is not introduced from iteration since NND4_StepPred resulted in the same trend of prediction divergence. On the other hand, the predicted inundation volume resulted in a trend for underestimation even provided with correct input for prior state information, and the underestimation is distributed mostly over paddies as illustrated in the previous contour maps. Based on the settings for hypothetically designed rainfall patterns, 3 hr is approximately the time when the cumulative rainfall for ND_CR65 exceeds the maximum range covered in the training dataset for model NND4. In other words, the capability of NND4 for predicting inundation area and water depth is severely degraded when the conditions of target event falls outside the characteristics of training data, further proving the limitations of RF models in making extrapolations.

5.4.3 Incorporation of all rainfall patterns

The previous two subsections investigated the performance of RF models configurated with unified shapes of training data. It was observed that the variety and representativeness of training data are essential factors determining the model's reliability due to the compromised capabilities for extrapolation. Therefore, it is reasonable to include large and diverse training data, for example configurating RF models using all hypothetical rainfall patterns investigated so far, to ensure the model's generalization capability while maintaining robustness. On the other hand, the inherent ensembled learning mechanism of RF regression also induces the potential to converge toward a mean behavior observed in training data. Since each individual decision tree produces a discrete uncorrelated estimate and the final prediction output is the arithmetic average across all trees, the averaging mechanism tends to smooth out localized extremes. In other words, incorporating all rainfall patterns in the training data is expected to improve prediction performance against rainfall events of intermediate scales, yet trade off might be introduced when targeting extreme events due to underrepresentation of such patterns in the training data.

Therefore, this section further configurated an integrated RF model incorporating all patterns of hypothetical rainfall patterns and evaluates its performance relative to previous models training with single-sourced rainfall patterns. Configuration of the integrated model

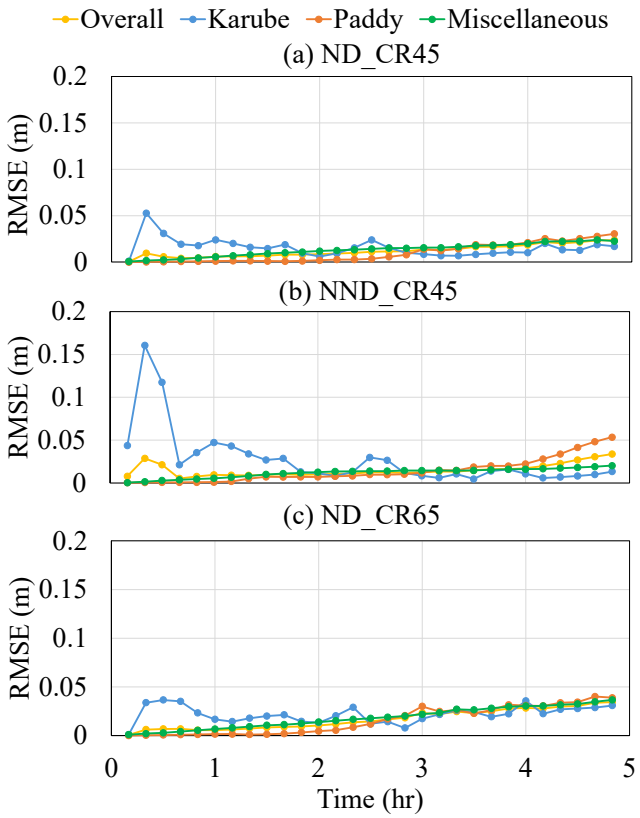


Fig. 5-42 RMSE of CMB4 evaluated against rainfall patterns of (a) ND_CR45, (b) NND_CR45, and (c) ND_CR65.

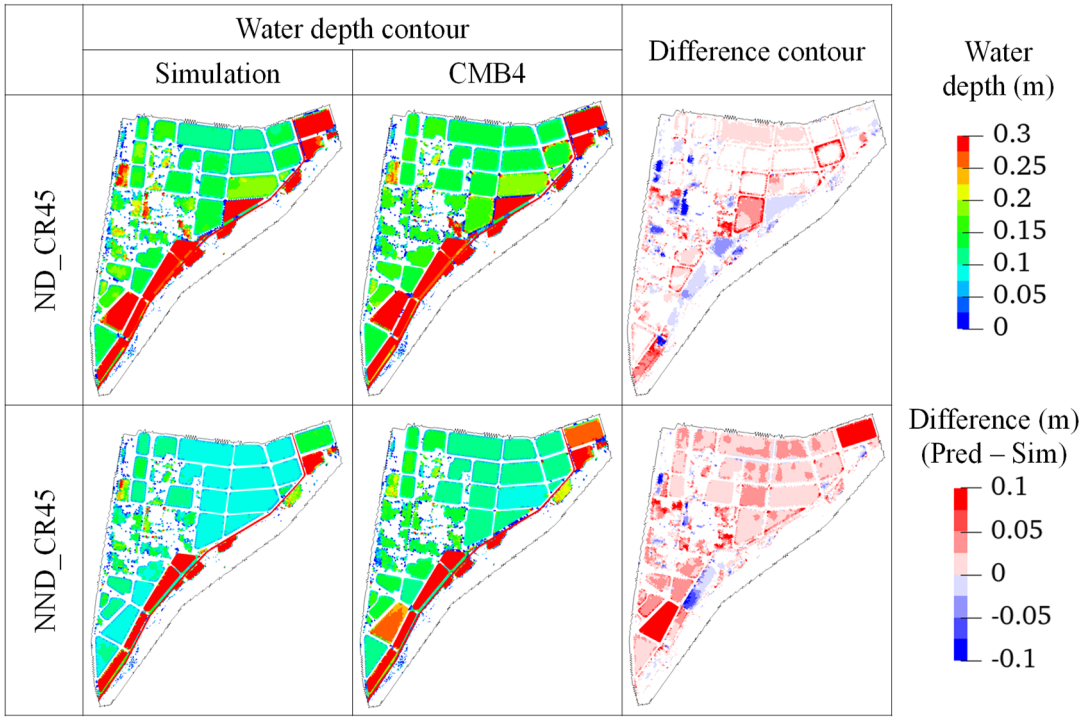


Fig. 5-43 Water depth and difference contour maps for final steps predicted with CMB4.

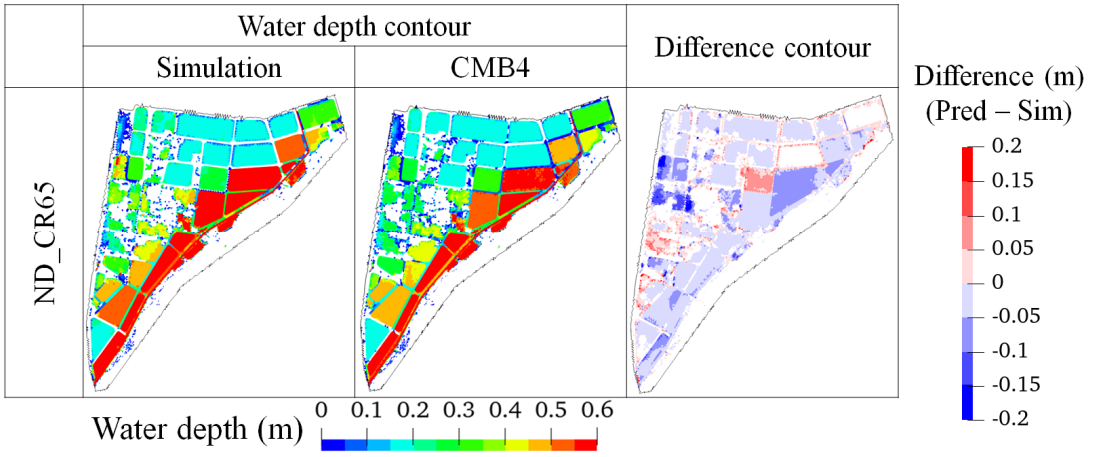


Fig. 5-44 Water depth and difference contour maps for final steps of ND_CR65 predicted with integrated model CMB4.

(denoted by CMB4) combined all sets of training data in terms of shape and peak time as summarized in **Table 5-6**, which resulted in 1080 total times steps of training data. The target events for predictive performance were the same hypothetical rainfall events of ND_CR45, NND_CR45, and ND_CR65 investigated previously and ND4 model was used as the baseline models for performance comparison. The first two events were selected to evaluate the generalization capability of the integrated model, and the last rainfall event was used to examine the potential trade off when the extra training data does not contribute to the extremes of target predictions. The RMSE of with respect to each land use is illustrated in **Fig. 5-42**. Comparing with the RMSE evaluations of the baseline model ND4 depicted in **Fig. 5-18**, **Fig. 5-35** and **Fig. 5-39**, it can be observed that the performance of the two models is highly consistent for the first two target rainfall events. While the initial error over the Karube river is not fully resolved, the integrated model CMB4 slightly reduces the RMSE by approximately 0.01 m for the prediction of ND_CR45 and NND_CR45. The spatial distribution of prediction in terms of contour maps at the final step of prediction (**Fig. 5-43**) also suggests that CMB4 better coincides with simulation results compared to the results from model ND4 (**Fig. 5-18** and **Fig. 5-38**), which is consistent with the expectation that increasing diverse training data improves the overall generalization capability of model. The evaluation results for the prediction against ND_CR65, on the other hand, indicates a slightly degraded performance with a greater bias towards underestimation as shown in **Fig. 5-44**, compared with results of the baseline model ND4 (**Fig. 5-40**). This observation is consistent with expectation that averaging mechanism of RF models tends to converge prediction results. Since the target ND_CR65 falls on the extreme end of the range in the training data, the extra data of NND rainfall patterns with smaller magnitudes in CMB4 leads to an underrepresentation of extreme events and the corresponding prediction therefore biases towards underestimation.

In summary, incorporating training datasets with diverse characteristics enhances the model's overall performance against a wider range of scenarios in comparison to models trained with single-patterned rainfalls. Yet the intrinsic mechanisms also induce the potential for conservative estimations over events that fall near the range of training data (e.g. biases towards overestimation for rainfalls with shallow depths and underestimation for events with large scales).

5.5 Chapter summary

This chapter further investigated hypothetical rainfall patterns of temporal varying intensities and configuration of RF models with different combinations of simulation results as training data. The following conclusions are drawn from the evaluations by testing models against multiple hypothetical rainfall events.

1. Based on the characteristics of actual rainfall events at the study area, hypothetical rainfall patterns were designed considering magnitude, temporal distribution and peak intensity in order to construct the datasets need for the configuration of RF models.
2. Investigations regarding feature selection further proved that RF models interpret static and dynamic properties differently where the importance for properties with dynamic values and temporal changes are higher in comparison. Features that directly represent physical properties, for example prior state of water level and runoff conditions, improved the predictive performance of RF models. Including features reflecting inflow and outflow conditions at river boundaries also increased the consistency of RF model both in identifying inundation area and predicting for inundation volume. The influence of cross-validation methods on reflecting the temporal correlation of data in this research was rather insignificant since the selection of prior state as features to some extent reflects the sequential dependence of data.
3. The performance of RF models significantly depends on the selection of training data due to their limited capability of extrapolation. The performance of RF models was compromised if either the shape or temporal distribution of the target event falls outside the patterns in the training data. Including diverse patterns of training data enhanced the generalization capability of the model. On the other hand, such large database also led to the potential for conservative prediction due to the ensembled nature of RF regression algorithms such that predicted results tend to average towards mean values when targeting extreme or rara events.

References

1) Homepage of Okayama Prefecture: *Flood Inundation Maps*. https://www.pref.okayama.jp/page/548036.html (Accessed September 28, 2025).
2) Crisis Management Department of Soja City: *Memorial Journal of Heavy Rain Disaster Response of July 2018*. https://www.city.soja.okayama.jp/s/kikikanri/kurashi/bousai/h30gouusaigai_kiroku/h30gouusaigai_kiroku_top.html (Accessed May 22, 2025).
3) Ministry of Land, Infrastructure, Transport and Tourism: *Variations in precipitation patterns due to climate change*. https://www.mlit.go.jp/river/shinngikai_blog/chisui_kentoukai/dai04kai/02_4_kouupatternhenka.pdf (Accessed September 29, 2025).
4) Citizens' Affairs Department, Mountain and Regional Development Division of Okayama Prefecture (2020). *Okayama Prefecture Land Conservation Ordinance Guide*.
5) Ministry of Land, Infrastructure, Transport and Tourism, River Bureau, Flood Control Division (2005). *Guidelines to creating maps of areas expected to be inundated by small and medium-sized rivers*.
6) Homepage of Soja City: *History of disasters in Soja City*. https://www.city.soja.okayama.jp/data/open/cnt/3/12320/1/P06.pdf?20250904181510 (Accessed September 28, 2025).
7) Homepage of Okayama Prefecture: *Flooding records maps of Karube river basin*. https://www.pref.okayama.jp/uploaded/life/689339_6119017_misc.pdf (Accessed July 1, 2025).
8) Ministry of Land, Infrastructure, Transport and Tourism: *Water Information System*. http://www1.river.go.jp/ (Accessed November 11, 2024).
9) Scikit learn User Guide (n. d.). *Cross-validation: evaluating estimator performance*. https://scikit-learn.org/stable/model_selection.html (Retrieved October 11, 2025).
10) Booker, D. J., & Whitehead, A. L. (2018). Inside or outside: Quantifying extrapolation across river networks. *Water Resources Research, 54*(9), 6983-7003.
11) Takoutsing, B., & Heuvelink, G. B. (2022). Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. *Geoderma, 428*, 116192.
12) Hateffard, F., Steinbuch, L., & Heuvelink, G. B. (2024). Evaluating the extrapolation potential of random forest digital soil mapping. *Geoderma, 441*, 116740.

CHAPTER 6 PREDICTION FOR FLOODS INDUCED BY REAL RAINFALL EVENTS USING RF MODELS TRAINED WITH HYPOTHETICAL DESIGN RAINFALL

6.1 Overview

In chapter 5, the feasibility of iterated prediction targeting hypothetical rainfall events with temporal-varying intensities using the augmentation approach was investigated. Configured with simulation results from HD models, the RF model with features directly reflecting prior state of adjacent meshes and river boundary conditions resulted in high predictive performance when the target rainfall event holds similar characteristics, in terms of peak time and rainfall shape, as the training dataset. The evaluations against untrained hypothetical rainfall conditions provided insights on the importance of integrating diverse training datasets to maintain the generalization capability of RF models. On the other hand, the practical applicability of the simulation-informed RF model when subjected to observed real rainfall events remains unverified. Unlike idealized hypothetical rainfall patterns explored in the previous chapter, the temporal variation of real rainfall events is much more complicated, and uncertainties could be introduced in the iterative prediction process.

Therefore, this chapter further assessed the capability of the RF models, trained with simplified hypothetical rainfall patterns, in reproducing spatial and temporal propagation of inundation under actual rainfall events. Two rainfall events that led to inundation at the Karube region in recent years were selected as the target event for prediction and two RF models trained with different training data were used for prediction. Model evaluations using both performance metrics and quantitative metrics as spatial and temporal assessment of predictive accuracy were carried out and comparisons between models and rainfall events were performed. Through these analyses, this chapter aims to identify how the inclusion diverse training data covering extended periods of rainfall influences the robustness of models. The behavior of model performance between rainfall events at different scales intends to provide insights for the design of data-driven models using HD simulated datasets under practical conditions.

6.2 Details of target rainfall events and RF models

Two historical rainfall events, 2018 and 2021 rainfall, were selected as the target event for evaluation. These two precipitation events induced significant flood damages even with the facility of drainage pumps at the downstream and details of rainfall periods were summarized previously in Section 4.2.2. A comparison of hourly rainfall intensity and cumulative rainfall observed at the Soja City observatory is illustrated in **Fig. 6-1**[1)], with timestamps shifted at 0 hr for simplicity. The magnitude of 2021 rainfall event was relatively moderate while that of 2018 rainfall was considered extreme both in terms of peak intensity and cumulative rainfall. Another major difference in the inundation characteristics of the two events, as observed from the simulation results using HD model (depicted in **Fig. 4-10** and **Fig. 4-11** respectively), is whether the flood propagates upstream of the prefecture road in the target area for RF model only covers the most frequently inundated area, the assumption in the construction of training data is that the flow exchange at the outline of target area is negligible except for rivers. Previous discussions regarding feature selection suggested that including features that reflect river boundary conditions improves better performance and higher consistency of inundation volume. As a result, it is expected that the predictive performance would be degraded under extreme rainfall events once the areas along computational boundaries are inundated.

As concluded from previous chapter, incorporating training datasets that cover diverse inundation characteristics enhances the model's overall performance against a wider range of scenarios. It can be inferred that the prediction for real-world rainfall events further requires the robustness and generalization ability of RF model due to the complexity of temporal variation of rainfall events and corresponding spatiotemporal variability in flood propagation process. As a result, the RF model used in the investigation of this chapter adopted the *CMB4* model which covered all of the hypothetical rainfall events and resulted in the highest overall generalization capability in the discussion of the previous chapter. While incorporated with

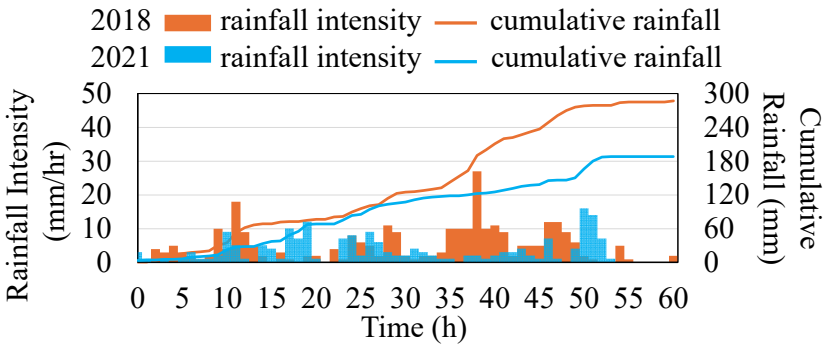


Fig. 6-1 Comparison of rainfall intensity and cumulative rainfall information for target events.

large variations of rainfall patterns in the training data, the database with 5-hr calculation period that only covered a maximum of cumulative rainfall of 173.8 mm. The magnitude of actual rainfall events, for example both of the target events, exceeds the range covered in training data due to long-lasting duration. Considering the limited capability of RF model to extrapolate, it is reasonable to include extra training data that covers larger range of cumulative rainfalls. Therefore, another model CMB4_EXT was configured with extra datasets simulated with 10-hr duration (R_{time}) using hypothetical rainfalls with empirical parameters a = 0.125 and b = 0.5 (Equation 5.1; **Table 6-1**). Both models take advantage of the same features as described in **Section 5.3.1** that resulted in the highest overall performance: coordinates (x, y), elevation (z), rainfall at each step ($rain_t$), rainfall at previous step ($rain_{t-1}$), prior water level of adjacent meshes ($H_{t-1,\ adj}$), runoff coefficients of adjacent cells both at previous step ($Fr_{t-1,\ adj}$) and current step (Fr_{adj}), and river boundary conditions as water level at upstream and downstream of river (H_{in}, H_{out}). Other hyperparameter settings were also set to be the same as *n_estimator* = 100, *max_depth* = 20, *max_features* = 'sqrt', splitting = *train_test_split* to maintain model consistency except for training data.

6.3 Results and discussions

6.3.1 Feature importance

First illustrated in **Fig. 6-2** is the feature importance comparison for the two RF models. The most important factors contributing to model accuracy are the prior water depth at each mesh, elevation and runoff coefficients for both models. Information of prior water level of adjacent meshes is also considered important with values close to 0.1, while that of rainfall resulted in the lowest feature importance of all. Both observations are consistent with the logistics of HD models. Compared to the effect of direct rainfall, the current water depth is primarily influenced by the prior state of the system, such as the preceding water depth and the flux exchange with adjacent cells. Boundary conditions of rivers, as of water level at inflow and downstream end, also dominate the control volume of water within the computational domain compared to direct precipitation, thus leading to the overall low feature importance of rainfall.

Table 6-1 Training data for model configuration.

Model name	Training data	Total time steps in training data	Maximum cumulative rainfall (mm)
CMB4	a = 0.5, 1.0; b = 0.4, 0.5, 0.6 (R_{time} = 5 hr)	1080	173
CMB4_EXT	a = 0.5, 1.0; b = 0.4, 0.5, 0.6 (R_{time} = 5 hr); a = 0.125; b = 0.5 (R_{time} = 10 hr)	1440	347

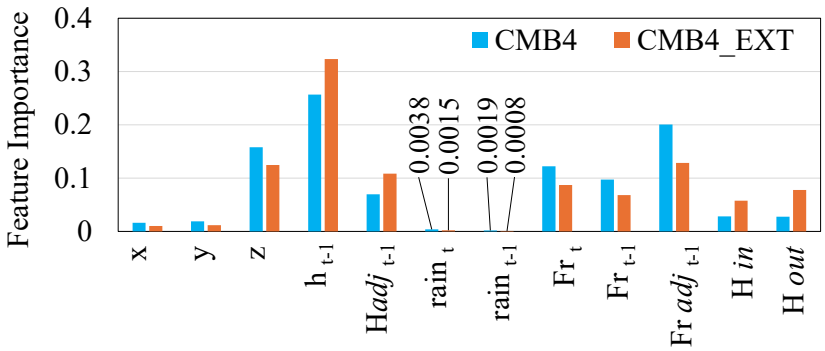


Fig. 6-2 Feature importance of configurated models.

In comparison, static features of coordinates resulted in low importance values since these features only serve as indication for location. Another observation is the slight variation of importance values between models due to the difference in training data. Even configured with the exact same features, CMB4_EXT indicated higher importance values for features of prior depth, prior water level of adjacent meshes and river boundary conditions, all of which represent prior state of hydrodynamic conditions and control volume of the system. This increase in feature importance arises since the additional training data in CMB4_EXT were generated with longer simulation durations where conditions with deeper inundation and flux interactions between neighboring meshes are dominates. In other words, extreme conditions with boundary-driven flux exchanges contribute a greater portion of the training data of CMB4_EXT, which lead to higher importance values.

6.3.2 Model performance against medium-scale rainfall

The temporal variation of simulated water depth contour under 2021 rainfall of medium scale is illustrated in **Fig. 6-3**. Overflow from the Karube river is observed starting 25 hr. Paddies located close to the river gradually reach full capacity of 0.3 m and flood propagates overtime with the majority of area inundated by more than 0.5 m at the end of precipitation.

As depicted in **Fig. 6-4**, the two models were first evaluated by the temporal variation of RMSE with respect to land use to observe in detail the different characteristics in inundation and between models. The predictive performance of the two model during the first 25 hr highly coincides with each other with overall RMSE suppressed at around 0.03 m. A slight increase in RMSE is observed starting 25 hr, approximately the time when overflow from the Karube river occurs, for both models and accuracy between models diverges. Both models illustrate relatively high accuracy up until 50 hr when a significant almost linear increase in RMSE up to 0.15 m is noted for CMB4. In comparison, CMB4_EXT still holds a low overall error of

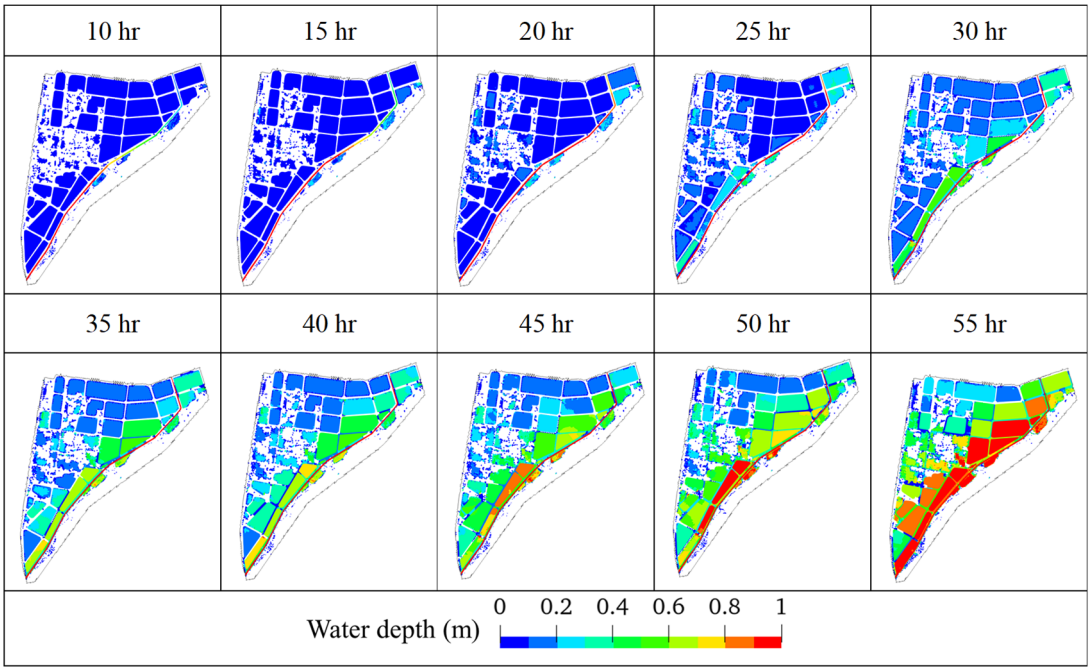


Fig. 6-3 Water depth contour from HD simulation (2021 rainfall).

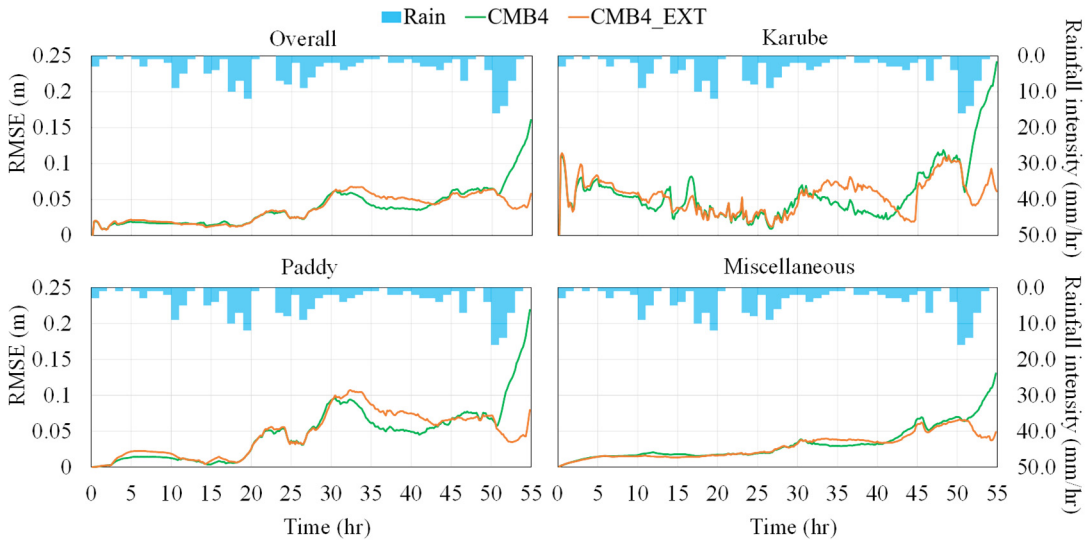


Fig. 6-4 Temporal variation of RMSE with respect to land use (2021 rainfall).

0.05 m at the end of prediction. When looking at the temporal variation with respect to each specific land use, the RMSE for river varies relatively significantly over time, especially with a slightly large error at the first few steps of prediction. Again, this initial high error is contributed from the small percentage of meshes as well as from the cold start from initial state of normal water level as specified also in previous sections. In comparison, the RMSE for paddy and miscellaneous land use starts with a low value and gradually accumulates overtime for both events. Overall, the RMSE comparison suggests that CMB4_EXT outperforms CMB4 except for a brief interval between 30 hr and 45 hr in which the predictive accuracy of the

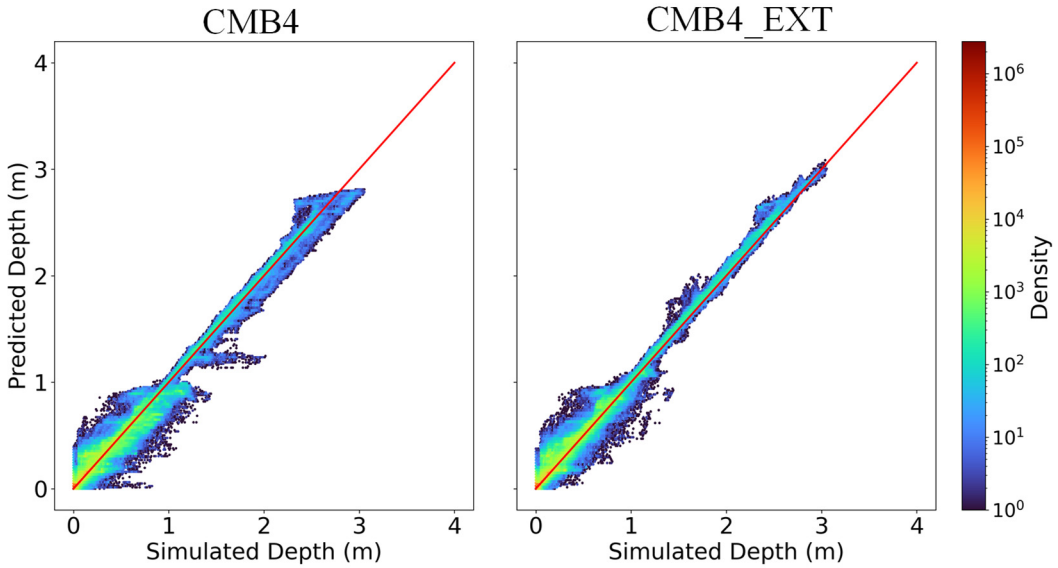


Fig. 6-5 Comparison of simulated water depth and predicted water depth with density.

former model is marginally compromised. To visualize the correlation of prediction and simulation results, a scatter plot with density of individual predictions is depicted in **Fig. 6-5**. The dense point cloud aligned closely to the perfect prediction line suggests overall high performance of both models. It is notable that the error diverges more significantly for predictions of shallow inundation depth, for example depths below 1.5 m. A trend for underestimation is also observed for CMB4 when water depth exceeds 1.5 m, suggesting a relatively low performance in comparison to CMB4_EXT.

To further investigate the detailed spatial and temporal distribution of error, the predicted water depth contour of CMB4 and CMB4_EXT and difference contour are depicted from **Fig. 6-6** through **Fig. 6-9**. From **Fig. 6-6** and **Fig. 6-7** it can be noted that CMB4 resulted in relatively high consistency with simulations during the first 25 hr where errors were suppressed within centimeters. As inundation depth increases and overflow occurs, the predictions reveal greater divergence spatial over land use of paddies and temporally between steps, which is expected from error accumulation over iterations. The most significant error for CMB4 is observed between 50 hr and 55 hr. The predicted contour maps suggest similar inundation depths while significant increase in water depth over paddy areas is observed from simulation, which leads to an underestimation of more than 0.2 m over the majority of areas at the final step of prediction. In comparison, the results predicted with CMB4_EXT illustrate a high performance during the first 50 hr (**Fig. 6-8** and **Fig. 6-9**). The final prediction at 55 hr, although still having a slight tendency towards underestimation, also indicates higher coincidence to simulation results compared to CMB4.

Another metric of critical success index (CSI) was performed evaluating how well the RF

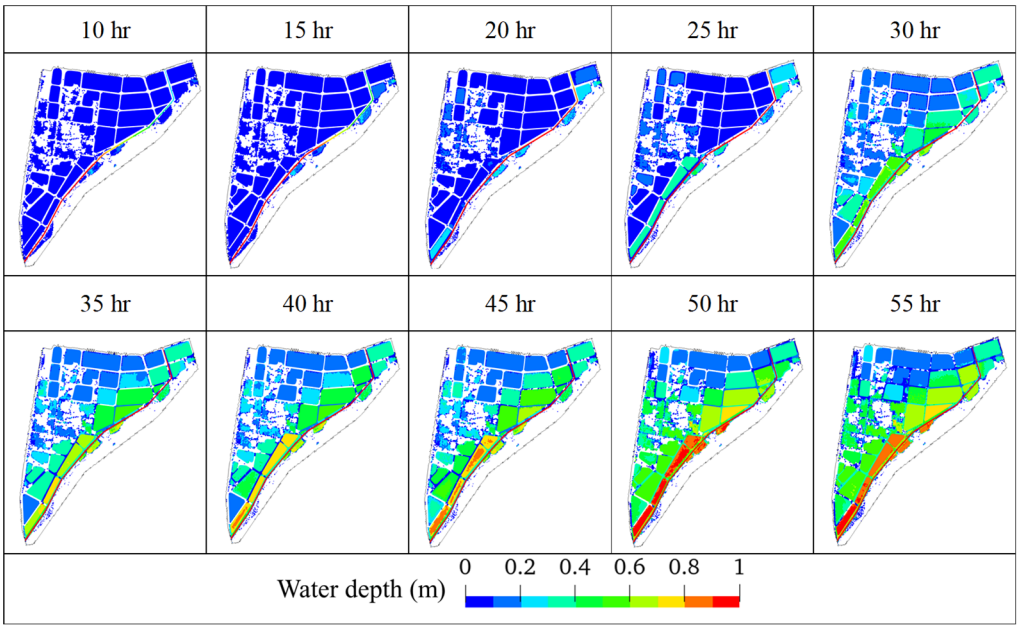


Fig. 6-6 Water depth contour predicted with iteration using CMB4 (2021 rainfall).

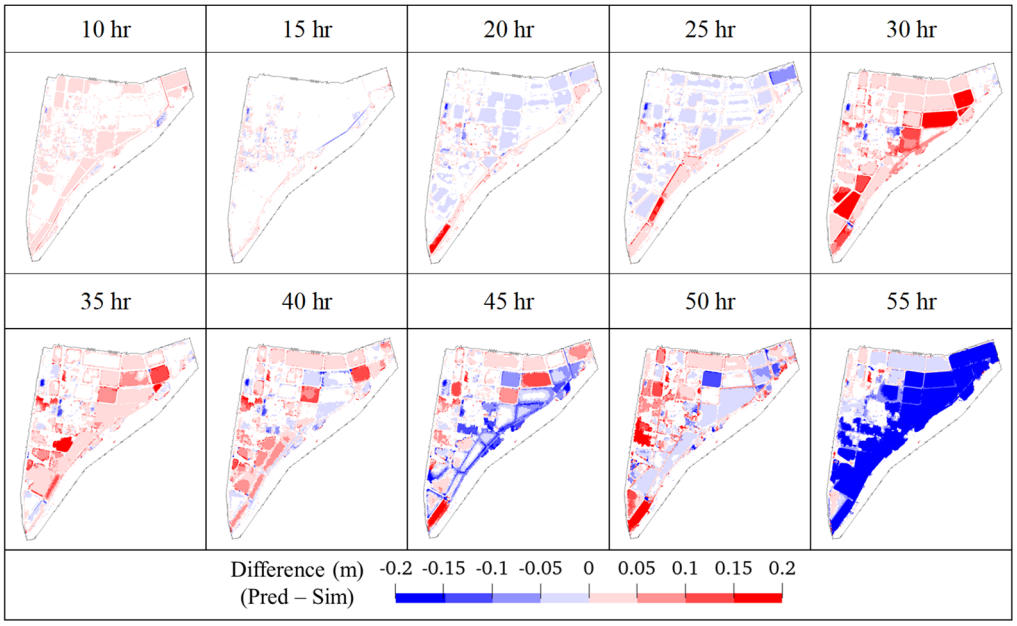


Fig. 6-7 Difference contour of CMB4 prediction compared to simulation (2021 rainfall).

model predicts for flood extent. CSI, also known as threat score, is a categorical agreement metric that measures the proportion of correctively predicted positives over the sum of total number of prediction and misses. The computation of CSI is performed by the following equation.

$$CSI = \frac{TP}{TP + FP + FN} \tag{6.1}$$

where TP, FP and FN represent true positive, false positive and false negative respectively,

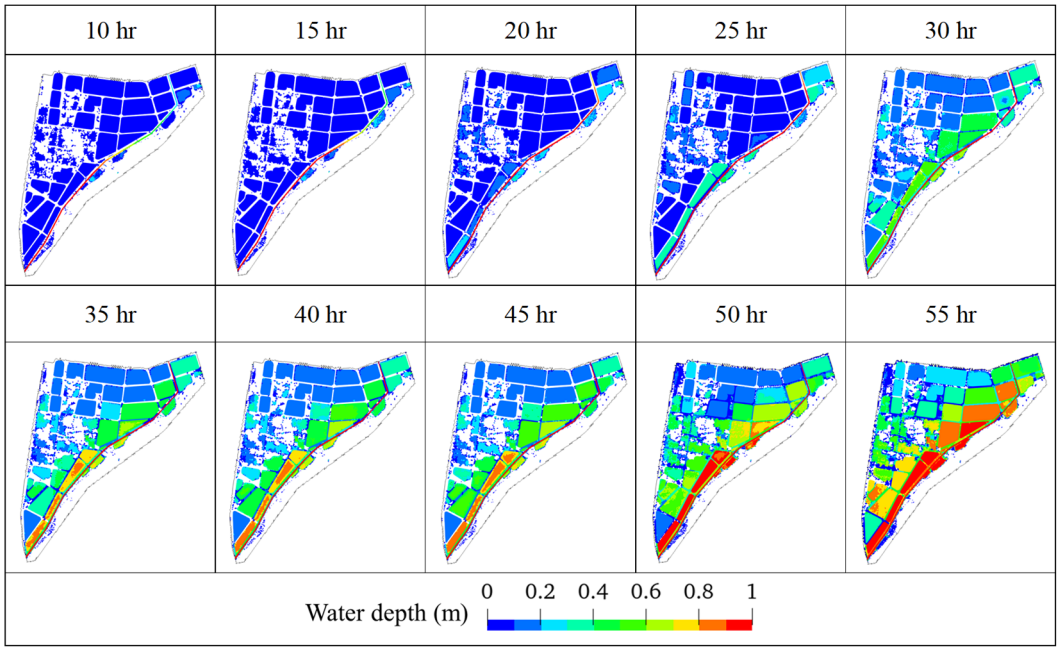


Fig. 6-8 Water depth contour predicted with iteration using CMB4_EXT (2021 rainfall).

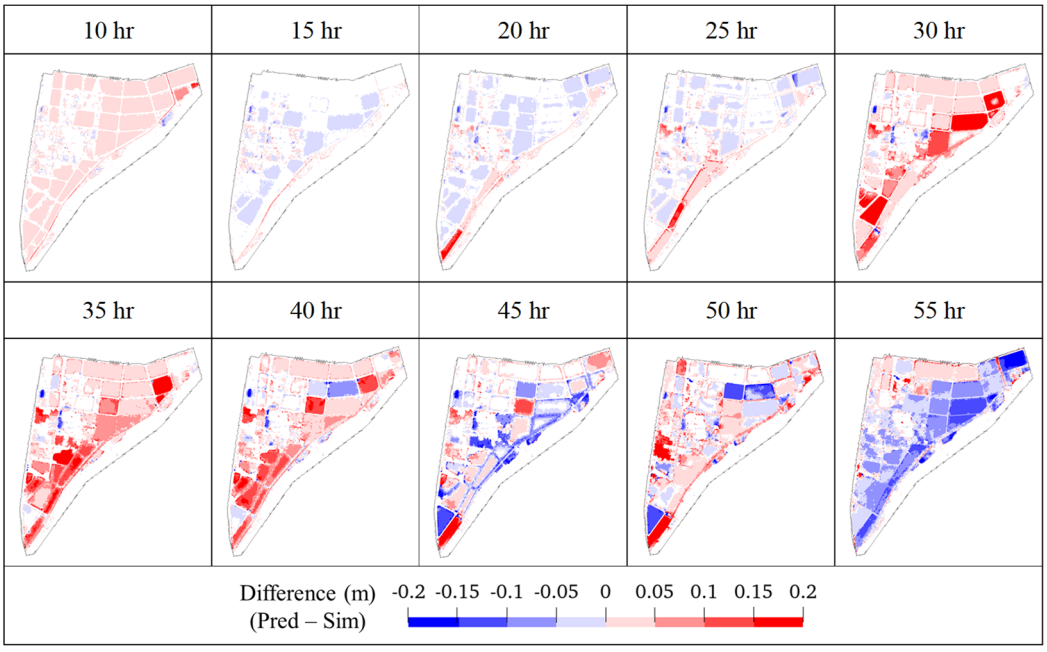


Fig. 6-9 Difference contour of CMB4_EXT prediction compared to simulation (2021 rainfall).

and the values are obtained the same way as confusion matrices described in Chapter 3. With values between 0 and 1, a higher CSI suggests better agreement of flood extent. Together with a contour map of flood arrival time, the spatial capability of RF model in assessing the occurrence of flood risks could be evaluated. Here multiple thresholds were selected for investigation: 0.01 m and 0.05 m, commonly used thresholds for the identification of wet and

dry processing, and 0.1 m and 0.3 m, where property damage may be induced and evacuation efficiency may be influenced. Summarized in **Table 6-2** and **Table 6-3** are the CSI metrics with respect to each land use for the two RF models respectively. Having values greater than 0.8 for all thresholds being evaluated, both models result in a relatively high overall performance and the difference between models is not significant. The evaluation with respect to land use, on the other hand, indicates a notable variation regarding different thresholds. For example, the CSI for miscellaneous land use does not change significantly with respect to the selection of threshold, while a degradation from 0.9 to less than 0.8 is observed for paddy when evaluated with a threshold of 0.3 m for both models. In comparison, the CSI for mountain starts with a relatively low value and ends with an almost perfect performance for threshold of 0.3m. This trend difference in land use is contributed from the various hydraulic flood characteristics. With the height of paddy ridge set to be 0.3 m, inundation depth of this thresholds indicates overflow from paddy occurs. That is, a significant change in the characteristics of water depth rise from depth accumulation to flood propagation occur at this threshold. In comparison, the flood characteristics of road areas do not inherit such temporal variation and the corresponding CSI metrics remain stable with respect to thresholds. The drastic change in mountain is induced by the overall shallow inundation, as observed from the simulated water depth contour (**Fig. 6-3**). The majority of mountain area is only shallowly inundated even at the end of precipitation and the number of meshes inundated with depths greater than 0.3 m is limited, which contributes to the extremely high CSI value. Comparing the two models, the general trend can be observed that the CSI scores of CMB4_EXT are relatively higher when evaluated with large thresholds regardless of land use. Since the extra training data cooperated in CMB4_EXT was simulated under more extreme precipitation

Table 6-2 CSI metrics for different land use of CMB4 (2021 rainfall).

Threshold	0.01 m	0.05 m	0.1 m	0.3 m
Overall	0.870	0.942	0.916	0.832
Miscellaneous	0.752	0.860	0.864	0.839
Mountain	0.576	0.513	0.416	0.997
Paddy	0.934	0.983	0.932	0.780

Table 6-3 CSI metrics for different land use of CMB4_EXT (2021 rainfall).

Threshold	0.01 m	0.05 m	0.1 m	0.3 m
Overall	0.859	0.935	0.916	0.846
Miscellaneous	0.758	0.852	0.867	0.842
Mountain	0.733	0.728	0.882	0.997
Paddy	0.908	0.976	0.929	0.804

conditions, this model performs better for inundation with greater depths while resulting in slightly compromised performance when inundation is shallow. This observation is also consistent with the previous illustrations of contour maps where CMB4_EXT resulted in higher differences between 10 hr and 25 hr (**Fig. 6-7** and **Fig. 6-9**).

The contour maps for time of inundation with thresholds 0.05 m, 0.1 m and 0.3 m are depicted in **Fig. 6-10** to evaluate whether the models successfully predict the temporal propagation of inundation, the arrival time of flood in other words. For the clarity of visualization, the figure only illustrates the results for areas with miscellaneous land use since such land use covering roads and residential areas is the most important land use relating to non-structural countermeasures under emergency situations such as evacuation route calculations. Under medium-scaled precipitation, the majority of road areas suffer inundation greater than 0.1 m and both models are able to capture the overall trend of flood propagation. Comparing the results of two models, CMB4 tends to miss certain areas of inundation over the upstream areas. The flood arrival time predicted by CMB4_EXT better coincides with

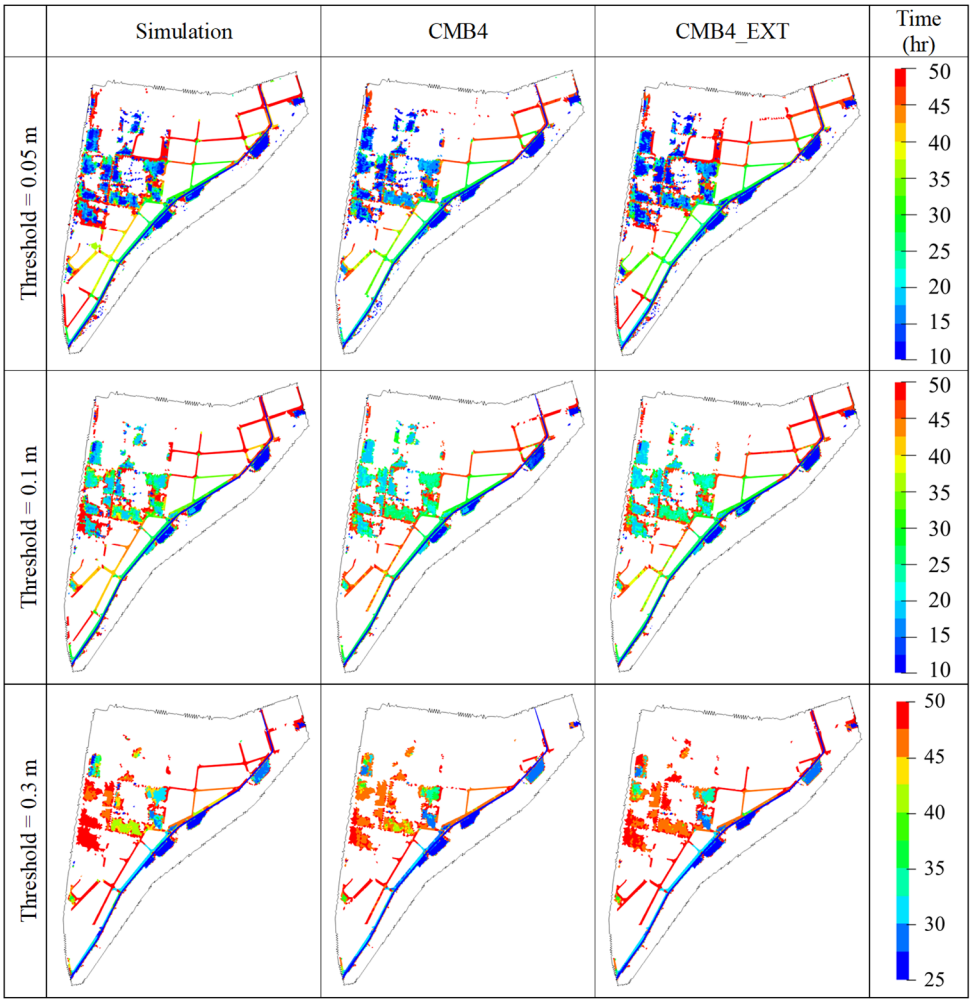


Fig. 6-10 Comparison of start time of inundation contours with different thresholds.

simulated results. But a slight delay is still observed in certain areas, which suggests the predictions falls on the danger end and might induce risk-underestimation in practical applications.

Finally, the performance was evaluated by metrics relating to the hydrologic validity and practical applicability of the models. Specifically, the inundation area and inundation volume were quantitively assessed, comparison of longitudinal water level of the Karube river at different times was carried out and the capability to predict for maximum inundation depth was investigated. Depicted in **Fig. 6-11** is the predicted inundation area, with 0.05 m threshold, and inundation volume using the two RF models compared with simulation results from HD model. During the early stage of rainfall events, the differences in predicted inundation area and volume between the two models are not significant, both having high alignments with simulated values. However, the values for predicted area and volume of CMB4 flatten when in reality flood propagates further with greater depth. Such divergence in prediction eventually leads to an underestimation in inundation volume by approximately 50,000 m^3. CMB4_EXT, in contrast, indicated overall high performance in predicting the temporal propagation of inundation area and volume, with some local errors for example during 30 hr to 45 hr. Indication for the different prediction behaviors is due to the selection of training data. As detailed previously in **Fig. 6-1**, the cumulative rainfall for the 2021 rainfall event approximately approaches 180 mm at 50 hr. The maximum cumulative rainfall used in the training data for CMB4 and CMB4_EXT is 173 mm and 347 mm respectively. The scale for 2021 rainfall exceeds the range covered by the training set of CMB4 after 50 hr, and the limited ability to extrapolate eventually leads to significantly compromised performance at the end of

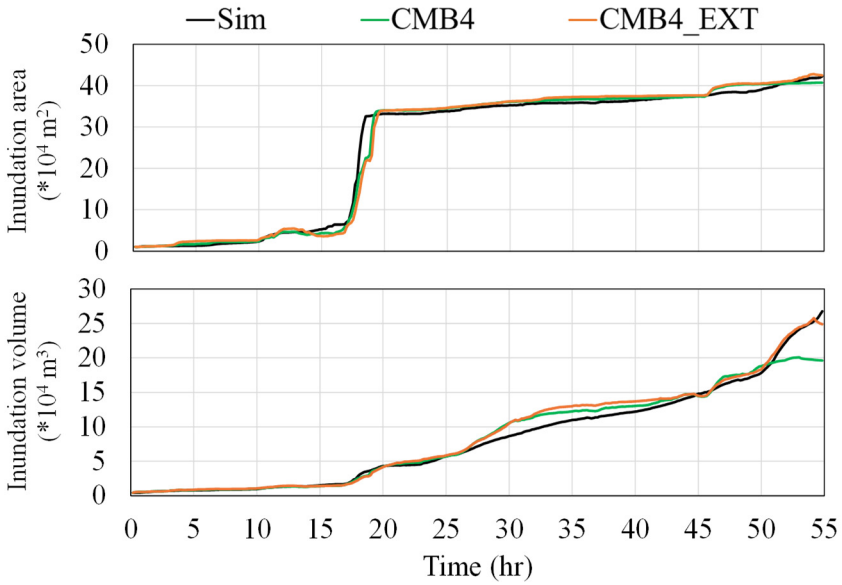


Fig. 6-11 Comparison of simulated and predicted inundation area and volume.

event. In contrast, the scale is still within the range of CMB4_EXT and significant incremental error variation is not observed.

The temporal variation of longitudinal water level of the Karube river was evaluated to hydraulic consistency of prediction to HD models. The evaluation focused on the early stage of precipitation event and the period within which where overflow from the river started to occur. Depicted from **Fig. 6-12** to **Fig. 6-14** are the longitudinal water level obtained from HD model simulation, prediction by CMB4 and prediction by CMB4_EXT respectively. Because of the closed sluice way condition at the downstream end, a longitudinal gradient is observed during the first few hours of precipitation and gradually levels starting approximately 5 hr for the 2021 rainfall event. The predicted outputs from both models, while overall capturing the temporal change of water level rise over time, resulted in a significant overestimation during

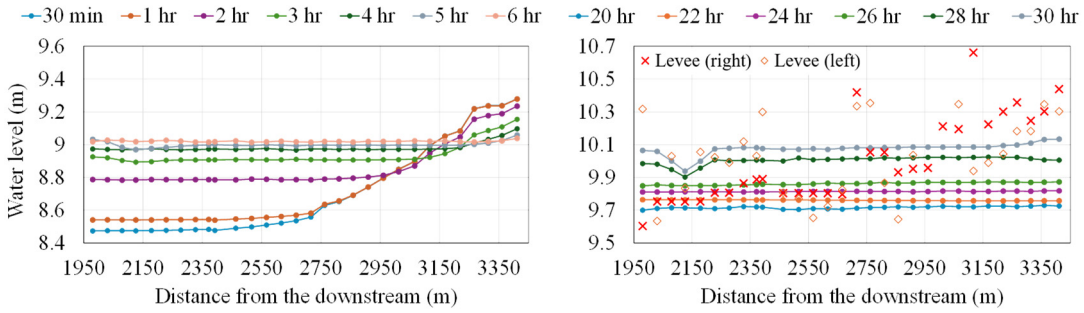


Fig. 6-12 Longitudinal water level of the Kaube river simulated from HD model (2021 rainfall).

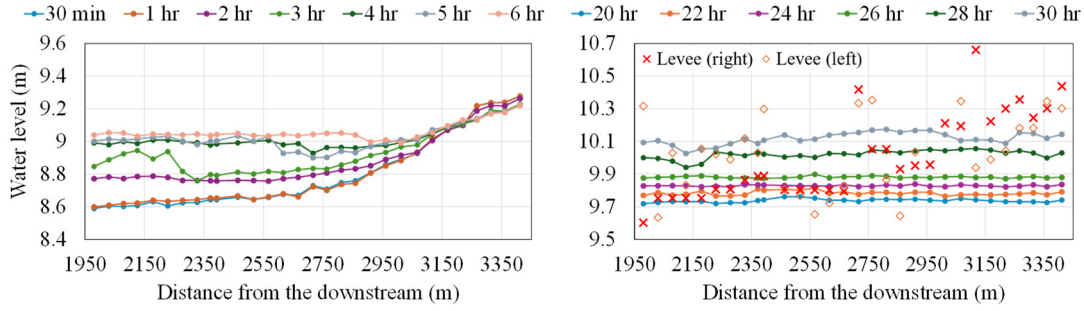


Fig. 6-13 Longitudinal water level of the Kaube river predicted by CMB4 (2021 rainfall).

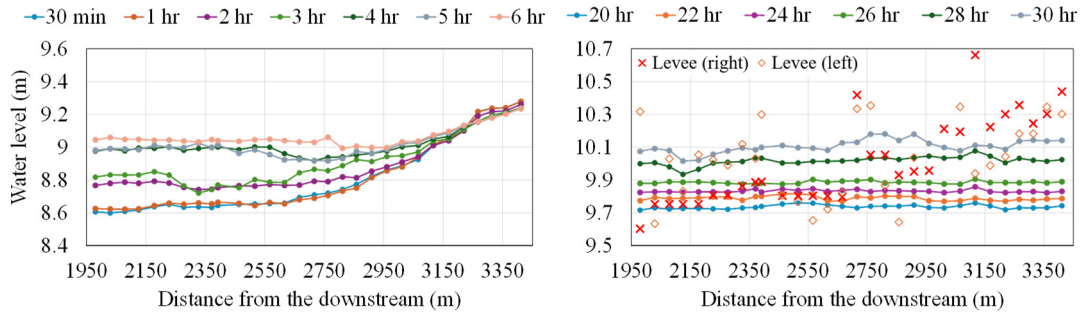


Fig. 6-14 Longitudinal water level of the Kaube river predicted by CMB4_EXT (2021 rainfall).

the first hour. The gradient consistency is also not fully reproduced for both predictions as indicated by the significant local variation at 3 hr and the retained gradient at the upstream when the simulated water surface flattens. Two reasons considered to contribute to this inconsistency are the abrupt incremental change in longitudinal water level and the independency in RF prediction. As noted from the simulation results, the water level rises significantly from the downstream and propagates backwards during the first few hours and the longitudinal gradient vanishes for the majority of the rainfall event. Since the prediction interval was 10 min, the prior state alone is not sufficient enough when abrupt variations occur. The local inconsistency at each specific time is mainly the result of the inherited nature of RF models in terms of making independent predictions based purely on data-driven statistics instead of the governing equations in HD models. In comparison, the performance of the two models is significantly improved as time progresses when the overall gradient diminishes. For example, the predicted longitudinal water level is more consistent compared to simulation around the period when overflow occurs as indicated by the figures on the right of **Fig. 6-12** through **Fig. 6-14**. The overall longitudinal characteristic is captured with high consistency from 20 hr to 26 hr for both models, suggesting that the time when overflow occurs at the downstream region is predicted with high accuracy. A slight overestimation is observed for the water level prediction afterwards, especially within the 2.55 km to 2.95 km region. Since the error is mostly restrained within 0.1 m, the performance of both models is considered acceptable in reproducing the longitudinal water level characteristics during real rainfall events and highlighted the potential feasibility for practical consequence relating applications.

As for the prediction of the maximum water level (**Fig. 6-15**), CMB4 resulted in a significant underestimation by approximately 0.2 m. In comparison, the prediction by CMB4_EXT coincides highly with simulation results with observable variations at some locations. The results again confirmed the compromised capability of RF model for extrapolation and the importance of considering greater range of training data for model construction. The ability of the CMB4_EXT to reproduce peak behavior of the entire river basin is then evaluated by

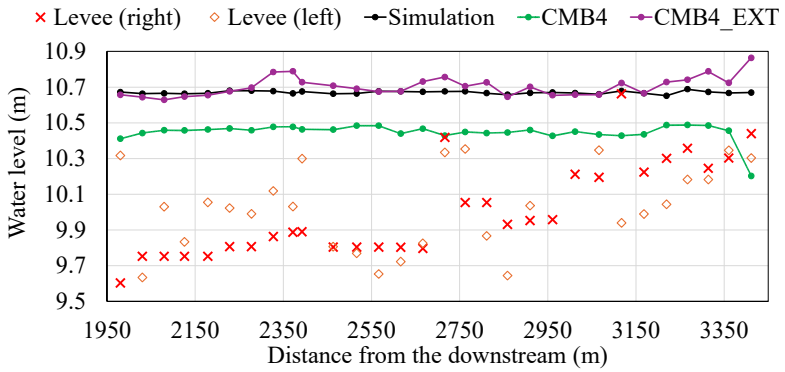


Fig. 6-15 Comparison of maximum longitudinal water level (2021 rainfall).

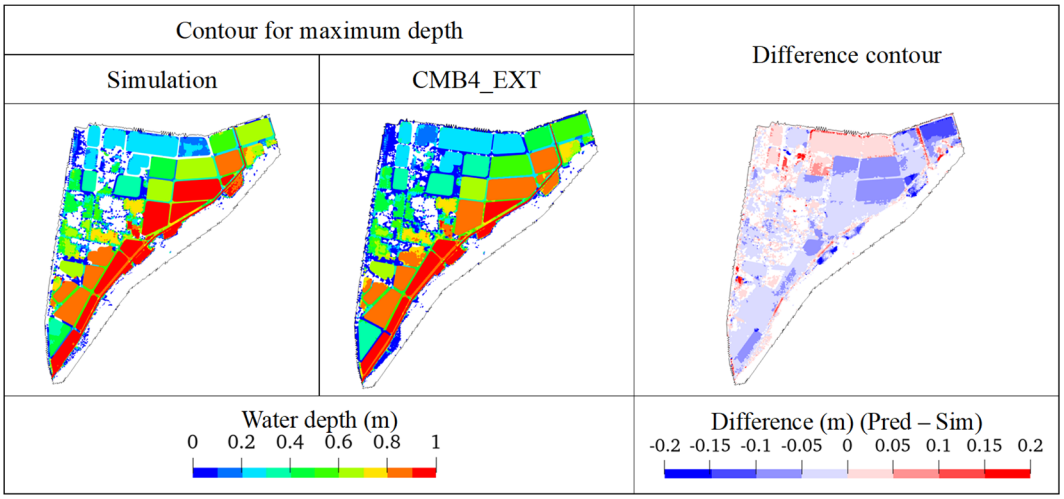


Fig. 6-16 Comparison of maximum depth contour.

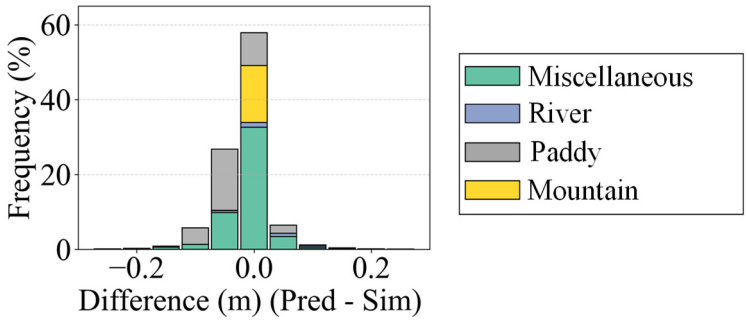


Fig. 6-17 Error histogram for maximum depth (CMB4_EXT).

comparing the maximum depth contour (**Fig. 6-16**) and error histogram with respect to land use with frequency evaluated against the total number of meshes (**Fig. 6-17**). The majority of prediction errors center in the range of -0.075 m to 0.075 m and the frequency significantly decreases as the error diverges from 0. On the other hand, a slight bias towards underestimation is still observable, especially for miscellaneous land use and paddies. Together with the evaluations regarding flood arrival time, it can be concluded that while CMB4_EXT is capable of making spatiotemporal water depth predictions with relatively high performance, this model may bring risks in practical flood warning systems under medium-scale precipitations such as the 2021 event.

6.3.3 Model performance against extreme-scale rainfall

The models were then tested against 2018 rainfall, which is an extreme event with approximate scale of 150 $yr^{2)}$. **Fig. 6-18** illustrates the simulation results of this event. Precipitation of this extreme scale induces severe inundation over the entire area and

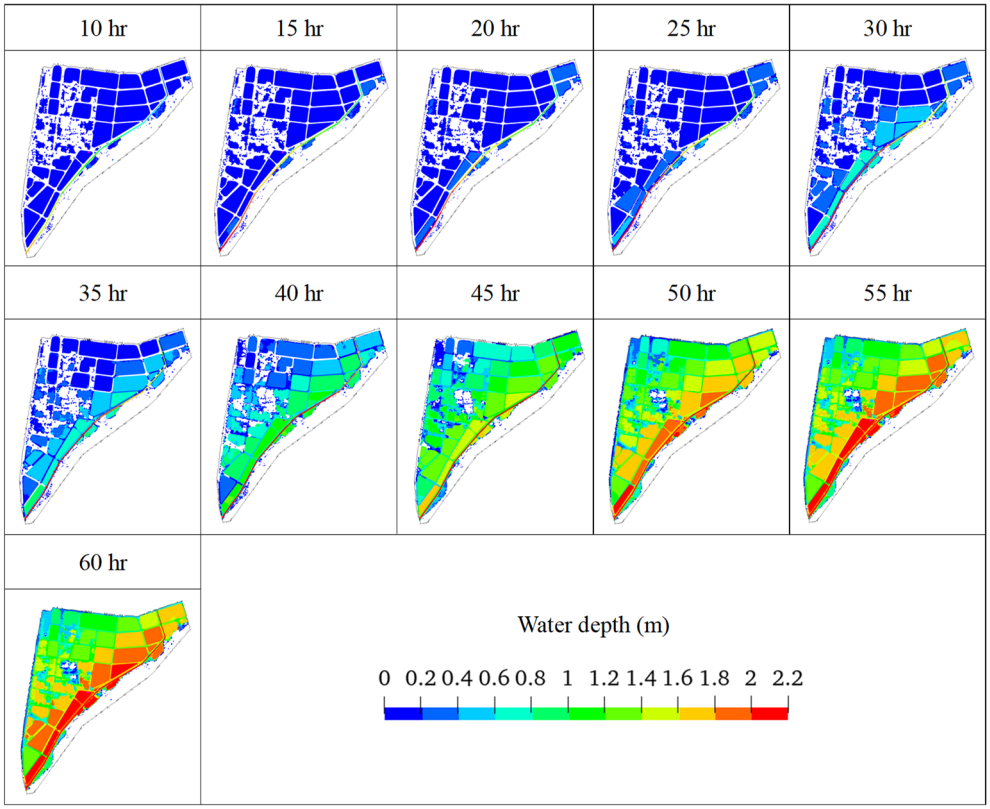


Fig. 6-18 Water depth contour from HD simulation (2018 rainfall).

inundation depth exceeds 2.2 m under the no-drainage situations.

First shown in **Fig. 6-19** is the temporal variation of RMSE of the two RF models. Similar to the results from the previous section, the performance of the two models first aligns closely with overall RMSE gradually approaching 0.05 m. A significant increase in RMSE is observed for CMB4 at approximately 38 hr while not for CMB4_EXT. Again this time when the performance of the two model diverges corresponds to the time when cumulative rainfall approaches the range of training data in CMB4, and the transition from interpolation to extrapolation induced the significantly degraded prediction accuracy. The RMSE for CMB4_EXT, on the other hand, retains an overall value around 0.05 m up until approximately 47 hr, and a slight increasing trend is observed afterwards with a final RMSE around 0.15 m. Since this increase in RMSE is noted for all land uses, it is important to identify whether the model induces significant bias and the reasons leading to this observation. **Fig. 6-20** illustrates the scatter plot for all predictions compared with simulation results of the two models. The majority of CMB4 predictions reveals substantial biases towards underestimation. The predictions result in a relatively parallel distribution to the perfect prediction reference line where predictions remain nearly constant as simulated depth increases, suggesting that CMB4 fails to respond proportionally to inundation depths. In comparison, the distributions of CMB4_EXT predictions clusters more closely relative to the reference line. For predictions

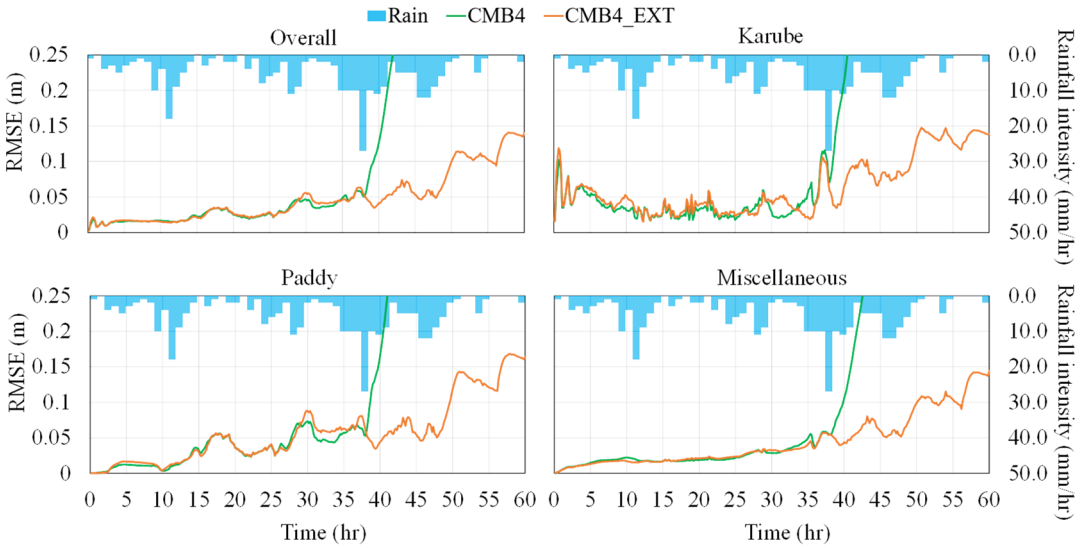


Fig. 6-19 Temporal variation of RMSE with respect to land use (2018 rainfall).

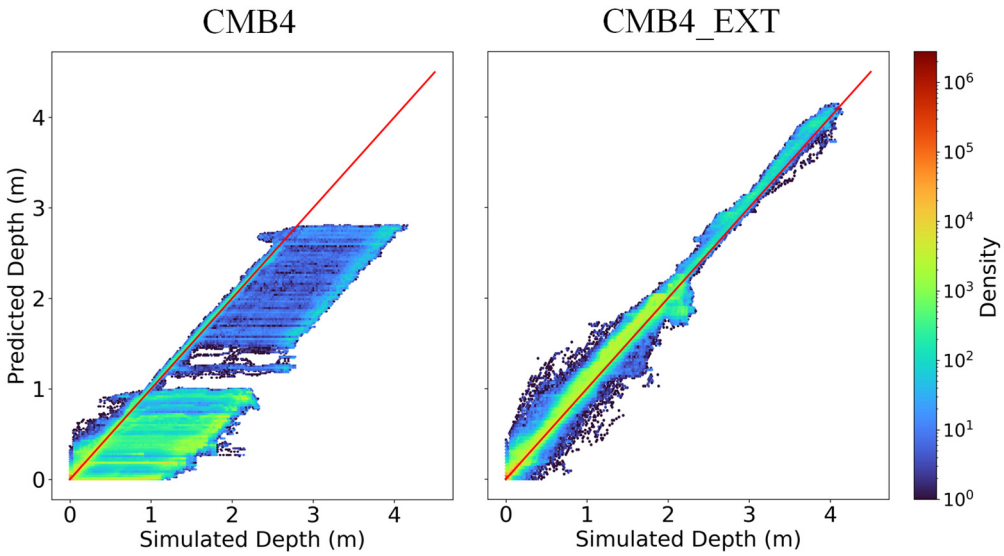


Fig. 6-20 Comparison of simulated and predicted water depth with density (2018 rainfall).

with inundation depths less than 2.5 m, the dense point clouds align symmetrically along the reference line. However, bias towards overestimation is observed for the prediction of deeper water depths, suggesting an inconsistency in the prediction performance with respect to inundation depths. The contributing factor to this observation is likely to be induced by the imbalance in training data and detailed investigations are carried out with other evaluation metrics and visualization of results.

Fig. 6-21 through **Fig. 6-24** depicts the predicted water depth contour and difference contour referring to simulation of the two models respectively. The prediction of CMB4 up until 35 hr indicates a relatively high coincidence with simulation results where the errors fall in the range of -0.05 m to 0.05 m. However, the water depth contours afterwards suggest limited response

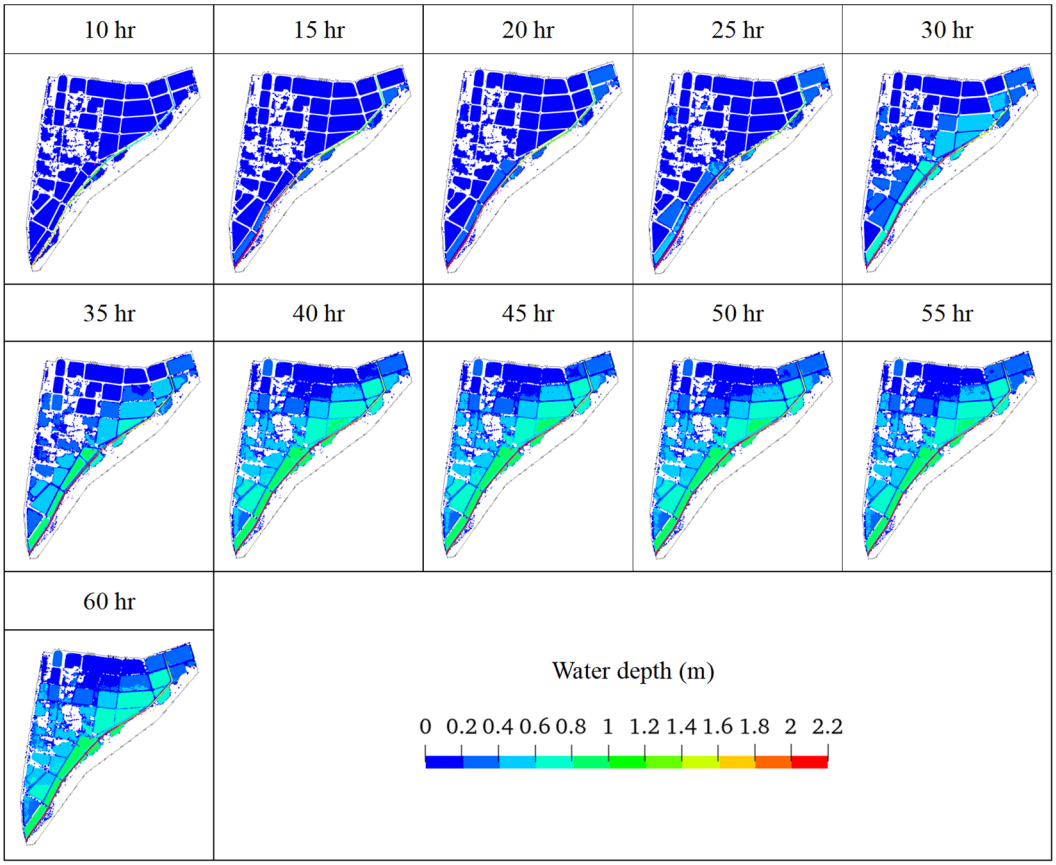


Fig. 6-21 Water depth contour predicted with iteration using CMB4 (2018 rainfall).

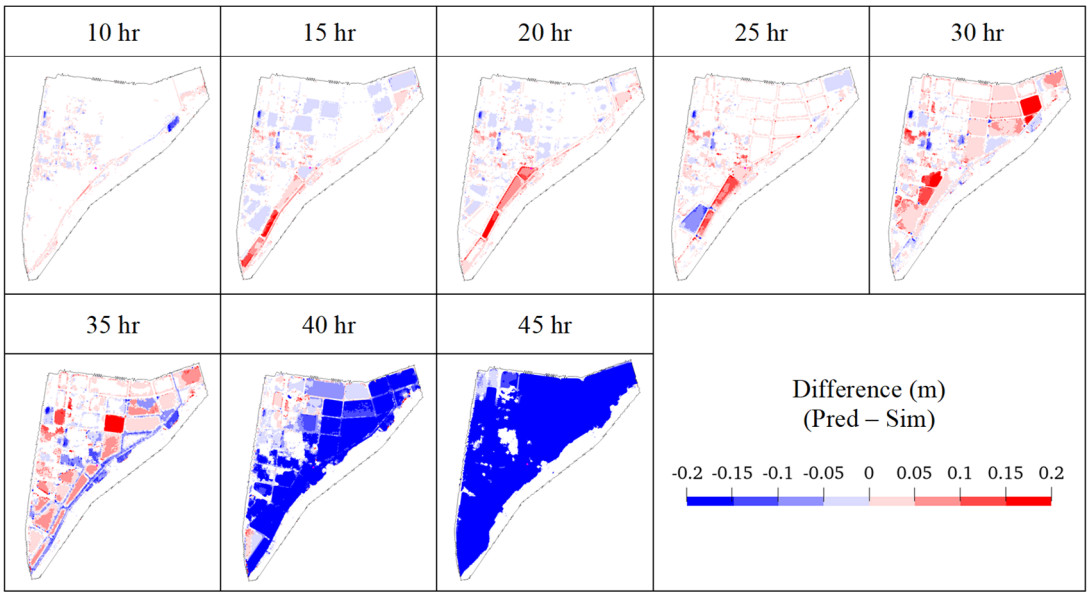


Fig. 6-22 Difference contour of CMB4 prediction compared to simulation (2018 rainfall) *.

* figures for 50 hr and after omitted.

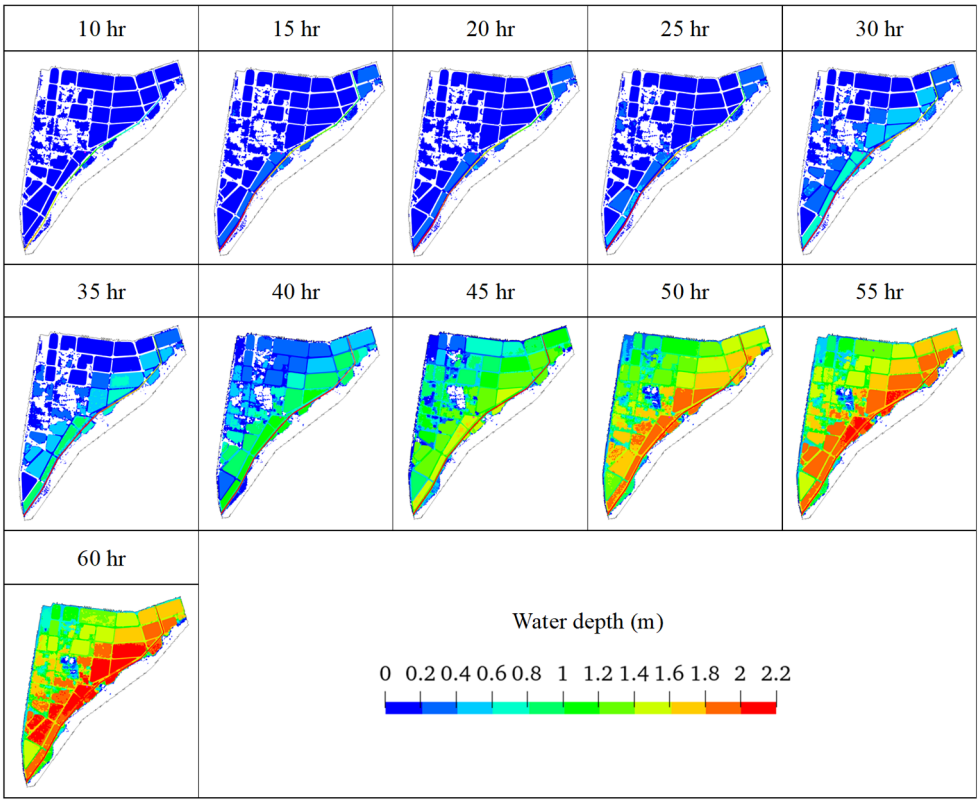


Fig. 6-23 Water depth contour predicted with iteration using CMB4_EXT (2018 rainfall).

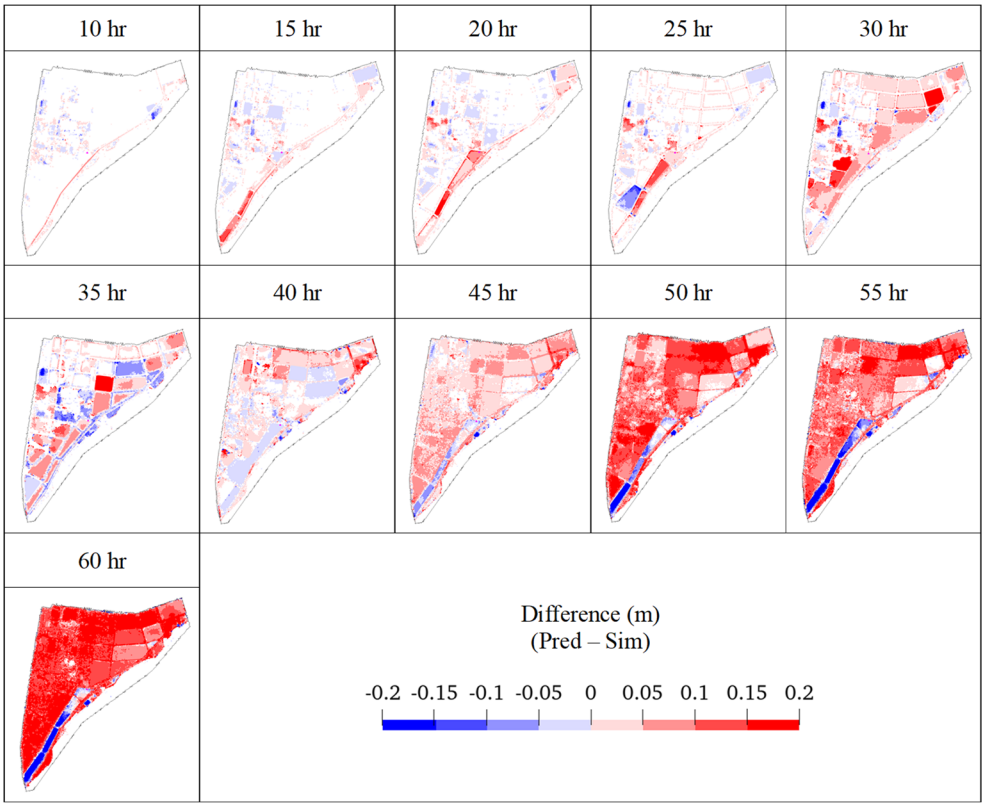


Fig. 6-24 Difference contour of CMB4_EXT prediction compared to simulation (2018 rainfall).

to the succeeding precipitation where flood continues to propagate and inundation depth increases. And this lack of temporal variation eventually leads to the gradually diverged prediction relative to simulation, which produces the significant increase in RMSE and consistently underestimations over the entire area. This results further provide evidence that RF models respond poorly to extrapolation and prediction accuracy significantly degrades once the target event exceeds the range of training data. Comparing **Fig. 6-23** to **Fig. 6-18**, it can be concluded that the predictions of CMB4_EXT overall capture the trend of temporal propagation of flood during the entire prediction period. The difference contour suggests high performance up until 40 hr with errors within the range of – 0.05 m to 0.05 m. Yet a gradual temporal error accumulation towards overestimation is observed over the majority of area after 40 hr, and this results are consistent with the increasing RMSE as discussed previously. As stated in Chapter 4, the boundary for the target area of RF model construction is delineated by the elevated topography of railway, prefecture road and mountain areas that cover the most frequently inundated areas. However, from the simulation results of the entire Karube river basin (**Fig. 4-10**), it can be observed that flood gradually propagates upstream of prefecture road under extreme precipitation conditions starting approximately 40 hr to 45 hr, resulting in a broader inundation area outside the target area used for RF model construction (**Fig. 4-12**). In other words, while flux exchange occurs at the boundary of downstream target area during the actual hydrodynamic computation, this phenomenon is not reflected by the RF model due to the limited area selection and imbalanced training data. Although the maximum cumulative rainfall used in the construction of CMB4_EXT was 347 mm, only three of the hypothetical rainfall patterns out of the total 42 patterns have cumulative rainfall exceeding 200 mm, which is the approximate magnitude at 40 hr for the 2018 rainfall event. With such limited training data on extreme conditions, the ensembled nature of the RF model resulted in predictions under the same logistics as the predictions for shallower water depth. The omitted outflow across the target boundary eventually induced the degradation of performance of CMB4_EXT during the later period of prediction.

The flood extent evaluation in terms of CSI and time of inundation contour was then performed investigating the models' capability to capture spatial and temporal behavior of flood. Summarized in **Table 6-4** is the CSI metrics for CMB4. The results suggest that the performance significantly degrades as the threshold increases for all land uses, and the overall evaluation with threshold of 0.3 m indicates that only two thirds of the meshes are correctly classified in the validation of flood extent. The evaluations for CMB4_EXT, in comparison, show a relatively consistent performance for all land uses, with an average value greater than 0.9 for all threshold selections (**Table 6-5**). The contour comparison for the start time of inundation, depicted in **Fig. 6-25**, also indicates that CMB4 fails to identify the temporal flood

Table 6-4 CSI metrics for different land use of CMB4 (2018 rainfall).

Threshold	0.01 m	0.05 m	0.1 m	0.3 m
Overall	0.872	0.854	0.822	0.673
Miscellaneous	0.733	0.676	0.646	0.541
Mountain	0.363	0266	0.160	0.134
Paddy	0.979	0.986	0.960	0.738

Table 6-5 CSI metrics for different land use of CMB4_EXT (2018 rainfall).

Threshold	0.01 m	0.05 m	0.1 m	0.3 m
Overall	0.930	0.963	0.955	0.949
Miscellaneous	0.870	0.934	0.944	0.952
Mountain	0.548	0.702	0.797	0.896
Paddy	0.978	0.983	0.960	0.939

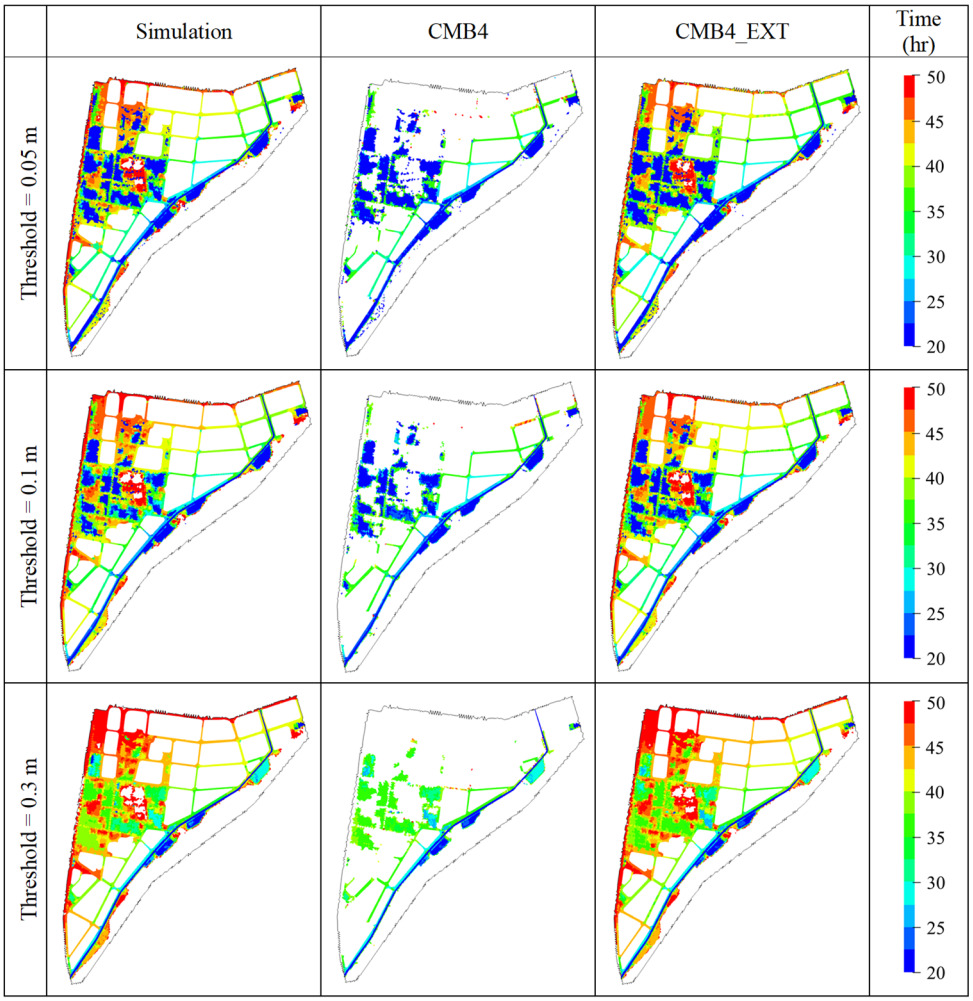


Fig. 6-25 Comparison of start time of inundation contours with different thresholds.

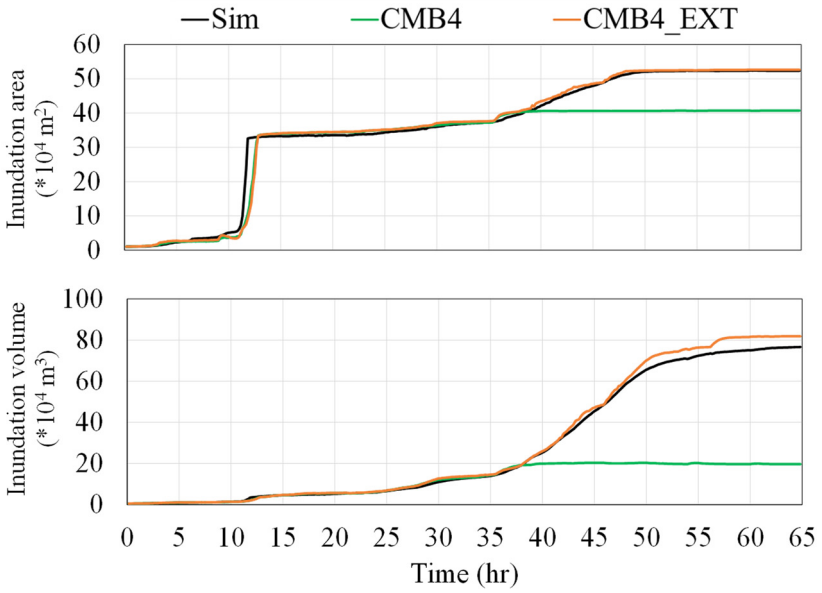


Fig. 6-26 Comparison of simulated and predicted inundation area and volume.

propagation by missing large inundation areas over miscellaneous land use. The results for CMB4_EXT are more consistent with simulation results, and no significant mismatch of flood arrival time is observed for all thresholds, which further confirms the high performance of this model under shallow inundation conditions.

Illustrated in **Fig. 6-26** is the quantitative comparison of inundation area, with 0.05 m threshold, and predicted inundation volume of the two models compared with simulation results. During the early stage of rainfall events, the differences in predicted inundation area and volume between the two models are not significant, both having high alignments with simulated values. However, the values for predicted area and volume of CMB4 flatten when the true flood propagates over larger areas with greater depth, which contributed to the consistent underestimations over the entire area. The ability of CMB4_EXT to identify inundation area is relatively robust throughout the entire period while the prediction for inundation volume slightly diverges with simulation after 45 hr. The results for longitudinal water depth prediction also draw similar conclusions. Illustrated from **Fig. 6-27** to **Fig. 6-29** are the temporal water level of simulation, prediction from CMB4 and prediction from CMB4_EXT, respectively. Both models struggled to provide accurate predictions during the first few hours, especially the acuate increase of more than 0.2 m at the downstream from 1 hr to 2 hr, when significant longitudinal gradient in water level is present. As the overall gradient gradually diminishes, the performance of both models increases with slight overestimation at the upstream most end. The consistency of prediction compared to simulation results is further increases as the gradient becomes negligible and the time when overflow occurrences starting approximately 13 hr at the downstream region are correctly captured by both models. With a

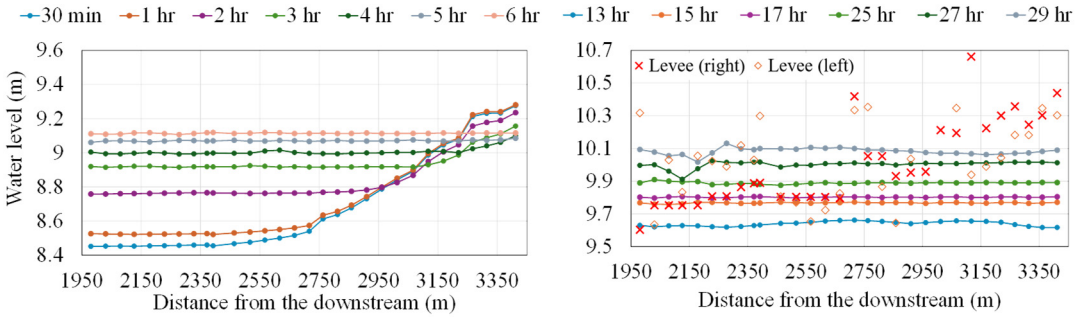


Fig. 6-27 Longitudinal water level of the Kaube river simulated from HD model (2018 rainfall).

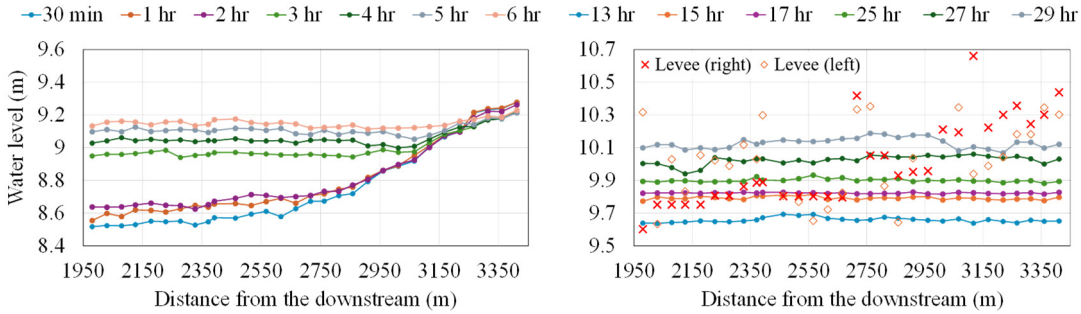


Fig. 6-28 Longitudinal water level of the Kaube river predicted by CMB4 (2018 rainfall).

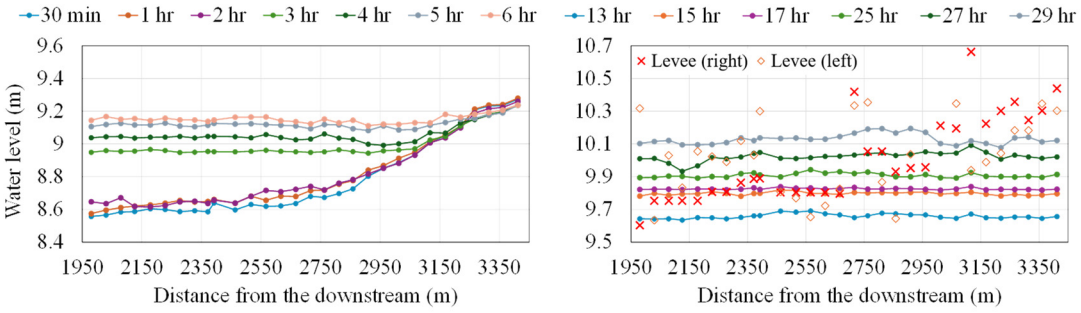


Fig. 6-29 Longitudinal water level of the Kaube river predicted by CMB4_EXT (2018 rainfall).

temporal cease in precipitation, the water level increases again at 25 hr and significant overflow from multiple locations are noted longitudinally. The predictions from both models resulted in slight overestimation, but the overall water level prediction is considered highly consistent with simulation results. The capability in predicting for maximum water level, on the other hand, differs significantly for the two models. As also noted from the evaluation regarding inundation area and inundation volume, the prediction of CMB4 levels when the cumulative rainfall of the 2018 event exceeded the range of training data. The corresponding prediction for water level also resulted in significant underestimation as shown in **Fig. 6-30**. In contrast, the prediction of CMB4_EXT is overall highly consistent with simulation with a slight trend for overestimation.

As observed from previous contour maps, flood starts to propagate outside the downstream

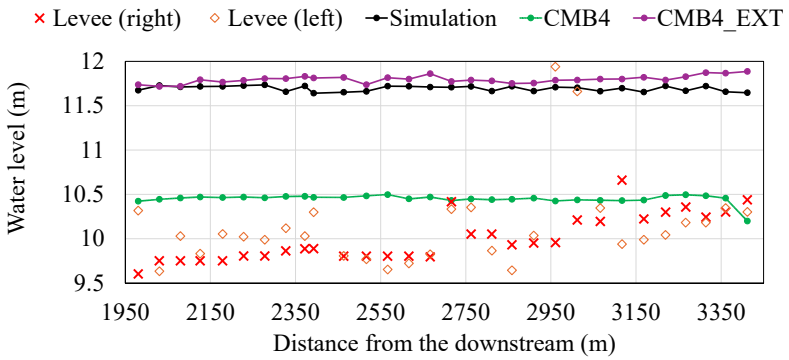


Fig. 6-30 Comparison of maximum longitudinal water level (2018 rainfall).

area at this time and the outflow from boundary reduced the rate at which inundation volume increases. On the other hand, CMB4_EXT fails to capture the flux exchange at boundaries due to limited training data for extreme events, and the predicted control volume continues to increase, resulting in a deviation from true dynamics. The corresponding prediction for the maximum depth, as a result, also results in significant overestimation. Shown in **Fig. 6-31** and **Fig. 6-32** are the comparison of predicted and simulated maximum depth, corresponding

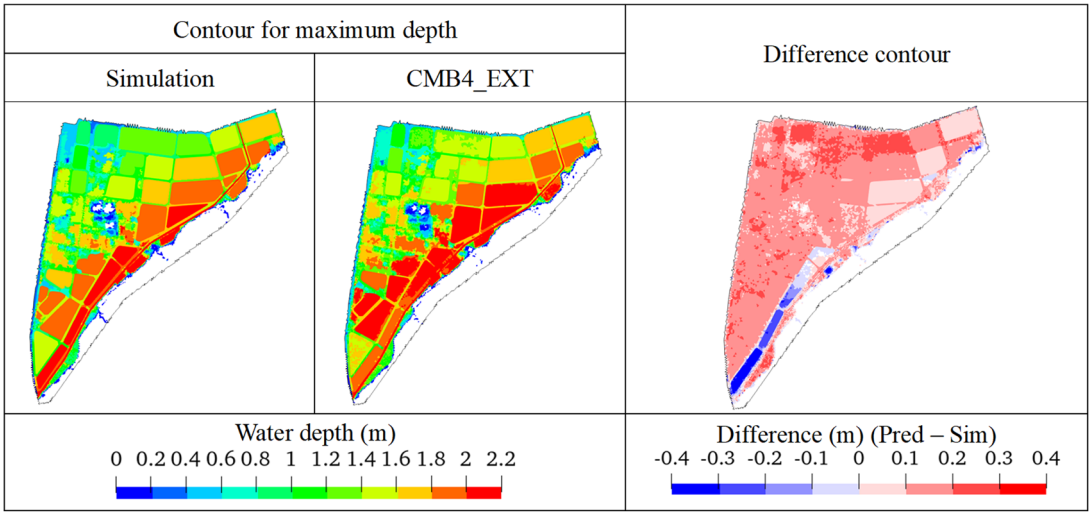


Fig. 6-31 Comparison of maximum depth contour.

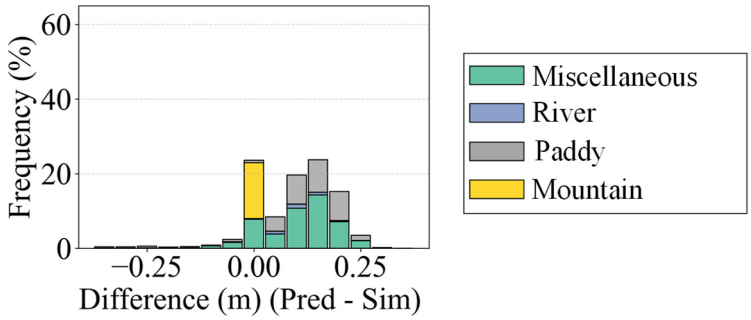


Fig. 6-32 Error histogram for maximum depth prediction (CMB4_EXT).

difference contour and error histogram summarized with respect to land use. The contour maps, again, provide evidence that CMB4_EXT successfully identifies inundation extent within the target area while overestimating the inundation depth over a large area. It is also noted that the larger overestimations distribute closely to the boundaries where flux exchanges occur in actual simulation. Certain downstream areas result in underestimations and a slight gradient in the predicted water depth within paddies is observed, suggesting a compromised robustness under extreme conditions. The histogram also indicates a significant bias towards overestimation as much as 0.25 m.

Although overestimation is more desirable in practical application compared to underestimation in terms of hazard warning, it is important to identify whether other factors also contribute to the biased trend. For example, previous chapters concluded that the iteration process induces error accumulation over time and the prediction performance is subjective to initial error. Considering the significant number of iteration under long-lasting practical rainfall events, error accumulation is also likely to contribute to the compromised performance of the model. Therefore, a prediction process without iteration (i.e. using simulation results as the input for each step of prediction) was performed to evaluate incremental error versus error accumulation. **Fig. 6-33** illustrates the RMSE predicted without iteration using CMB_EXT. Despite a high RMSE in the Karube river, the performance of the model predicted with a constant 1-step lead time (10 minutes) is promising during the entire prediction period with values suppressed below 0.04 m. The scatter plot comparison of prediction to simulation, shown in **Fig. 6-34**, also indicates a more densely concentrated prediction along the perfect prediction reference line. The bias towards overestimation under greater inundation depths observed in the iterated prediction results is greatly resolved. On the other hand, an increasing trend during the later period of prediction is still observed in the temporal variation of RMSE even eliminating the influence of error accumulation. **Fig. 6-35** and **Fig. 6-36** depict the comparison of maximum depth and error distribution of the results predicted without iteration respectively. The contour maps match more closely to simulation results compared to results

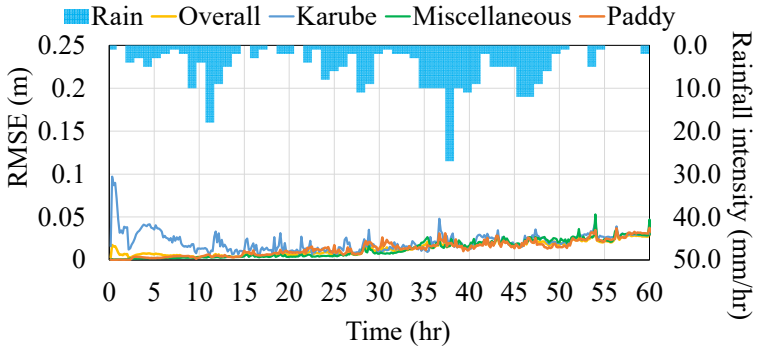


Fig. 6-33 RMSE with respect to land use predicted without iteration using CMB4_EXT.

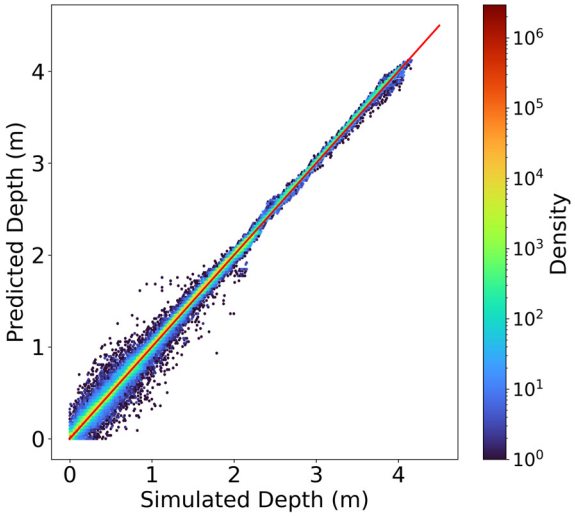


Fig. 6-34 Comparison of simulated water depth and predicted water depth (predicted without iteration using CMB4_EXT).

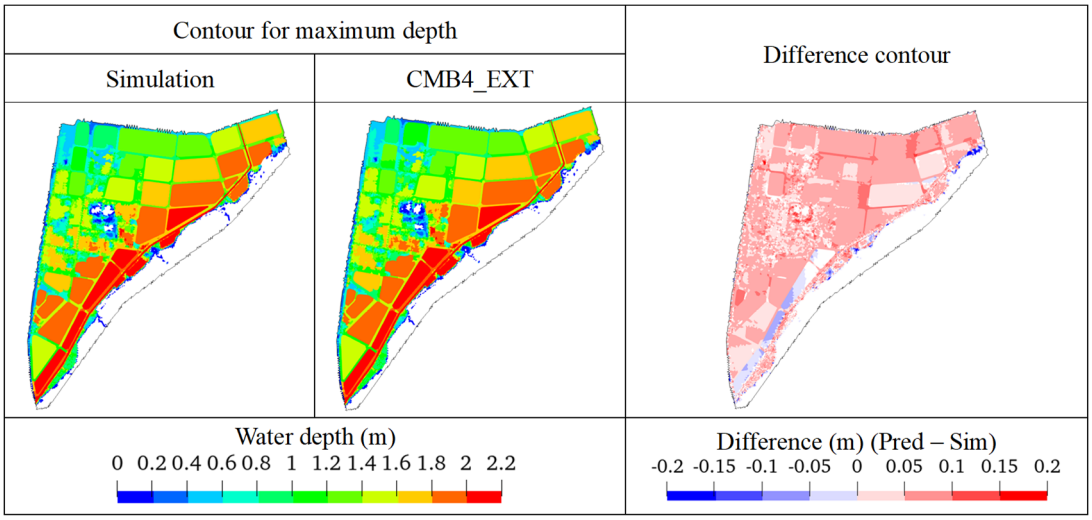


Fig. 6-35 Comparison of maximum depth contour predicted without iteration.

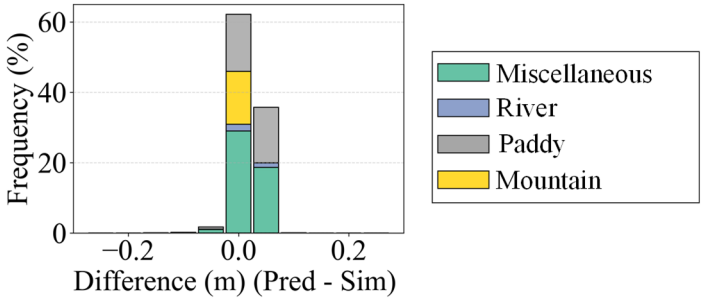


Fig. 6-36 Error histogram for maximum depth predicted without iteration using CMB4_EXT.

from iterated prediction process. The error histogram suggests a highly concentrated distribution around 0 and the spread of prediction error is reduced from an absolute range of 0.25 m to less than 0.1 m. However, the trend for overestimation still exists where a majority of area is overestimated by 0.05 m to 0.1 m. Since this result represents an incremental error, it provided evidence that this bias is induced mostly by the imbalanced training data rather than error accumulation due to the iteration process. While the results suggest that further investigation regarding the selection of training data is necessary, the overall performance of the model is confirmed with acceptable errors. The results also conclude the potential for the parallel usage of HD model and augmented RF model in practical applications. Instead of iterated predicting for an entire rainfall event, the predictions could be calibrated over time by updating the input for prediction with simulation results. In this case, the temporal error accumulation of the iteration process resets after a certain period of time, and more accurate predictions could be carried out with high efficiency for succeeding periods.

6.3.4 Discussion regarding generalization capability and efficiency

From the above-mentioned evaluation against two practical rainfall events, it is concluded that the model performance is highly dependent on the selection of training data. The accuracy of model is significantly compromised if the cumulative rainfall of the target event exceeds the range used in the training data as proven by the results of CMB4 under both rainfall conditions. The relatively high performance of CMB4_EXT in comparison suggests that including extra training data of extreme events improves the overall generalization capability of the RF model. On the other hand, it is also notable that this large dataset also introduces the potential for conservative prediction under floods with shallow depths. Both the RMSE and contour maps prove evidence that CMB4 outperforms CMB4_EXT during a certain period of time, for example during 30 hr to 45 hr of 2021 rainfall and during 27 hr and 35 hr of 2018 rainfall.

Besides further investigation regarding the selection of training data, another factor that affects the generalization capacity is the settings of hyperparameter. In general, a larger value for the number of tree configures the model with diverse tree and stronger variance reduction[3)]. Models with more trees, in other words, are likely to perform better under rare and extreme events while not for moderate patterns. Large trees also result in longer training time and increased size of output model, both of which could compromise the efficiency in application. To evaluate whether significant difference in performance and efficiency is introduced by the settings of number of trees, another RF model was constructed with all conditions the same as the original CMB4_EXT model except for increasing number of trees from 100 to 200

(CMB4_EXT200). The new model was then evaluated against the same two rainfall events for comparison.

Fig. 6-37 and **Fig. 6-38** illustrate the RMSE for the 2021 and 2018 rainfall events predicted with CMB4_EXT200 respectively. The overall trend of the temporal variation for each land use, for example the high initial error in the Karube river and gradual error accumulation in paddy and miscellaneous land use, is similar to the trend observed from the results predicted by the original model. The prediction for 2021 rainfall resulted in a similar overall RMSE of 0.05 m at the end of prediction, while a slightly lower RMSE for 2018 rainfall. The comparison of water depth contour and difference maps for the two events, illustrated from **Fig. 6-39** to **Fig. 6-42**, also indicates the same conclusions. For example, CMB4_EXT200 results in significant error in prediction with local overestimations of approaching 0.2 m at 25 hr of 2021 rainfall (**Fig. 6-40**), while the model with 100 trees has significantly better coincides with simulation results (**Fig. 6-9**). The predictions over other time steps also suggest greater variance in error. On the other hand, CMB4_EXT200 outperforms the original model when evaluated under extreme rainfall of 2018. Although the contour maps still show tendency for overestimation at the end of prediction, it is observed from the comparison between **Fig. 6-42**

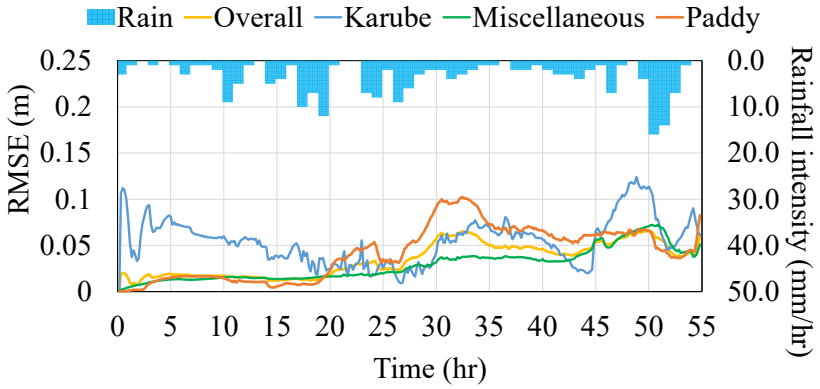


Fig. 6-37 Temporal variation of RMSE for 2021 rainfall predicted with CMB4_EXT200.

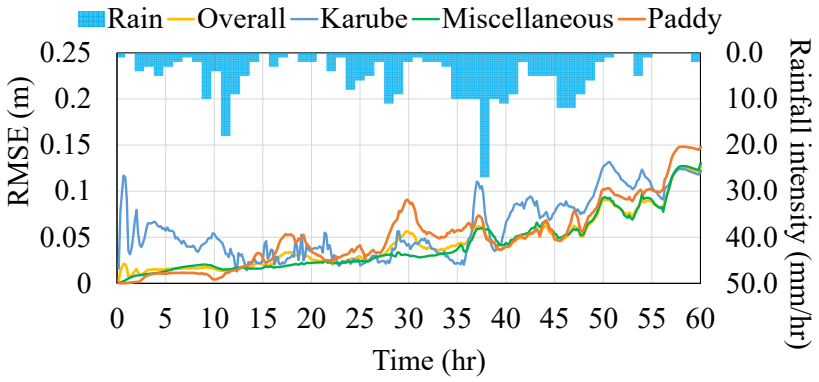


Fig. 6-38 Temporal variation of RMSE for 2018 rainfall predicted with CMB4_EXT200.

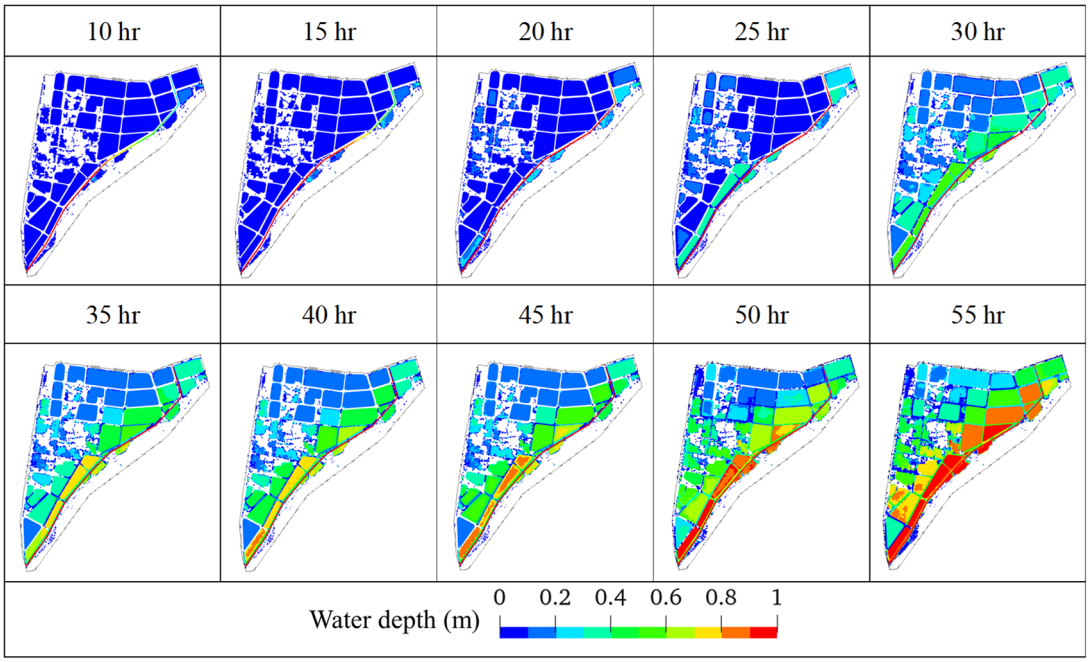


Fig. 6-39 Water depth contour predicted with iteration using CMB4_EXT200 (2021 rainfall).

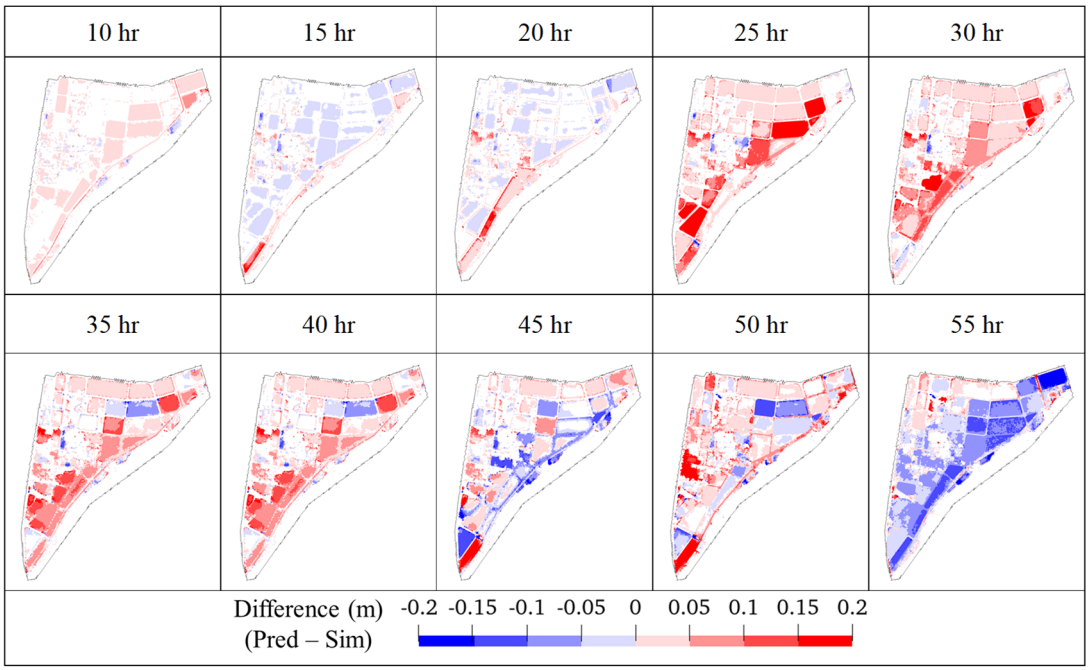


Fig. 6-40 Difference contour comparison of CMB4_EXT200 (2021 rainfall).

and **Fig. 6-24** that CMB4_EXT200 predicts for water depth with less error for almost all time steps being evaluated. The corresponding prediction for maximum depth also suggests that large values of trees generate larger variance for medium-scaled precipitations (**Fig. 6-43**) and slightly more consistent results for extreme events (**Fig. 6-44**) compared to simulation. This difference in prediction performance under different rainfall conditions is consistent with how

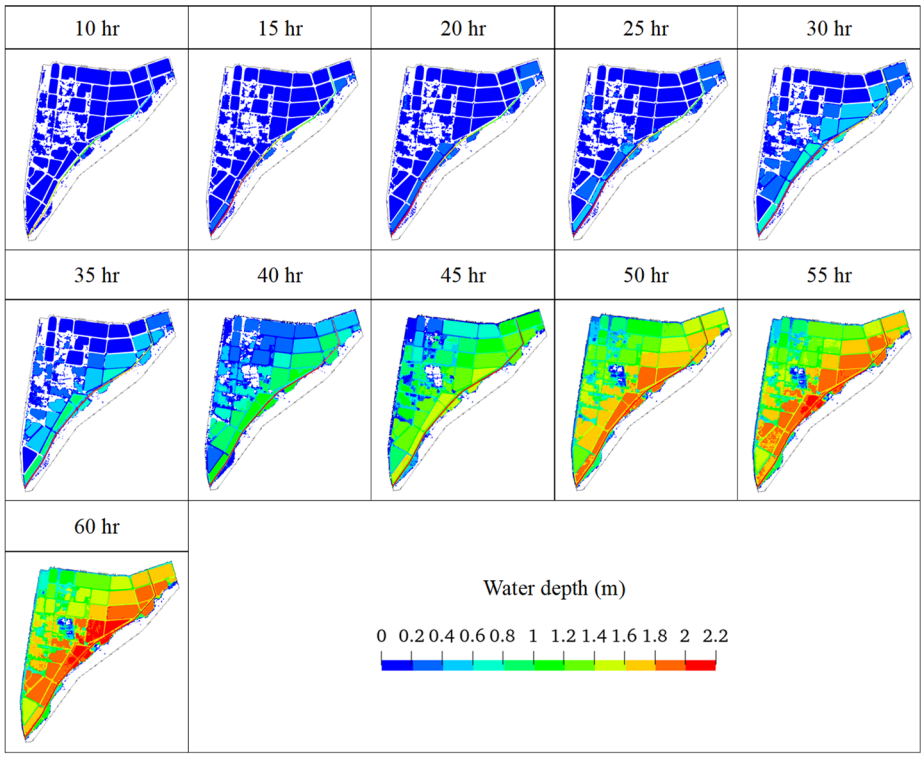


Fig. 6-41 Water depth contour predicted with iteration using CMB4_EXT200 (2018 rainfall).

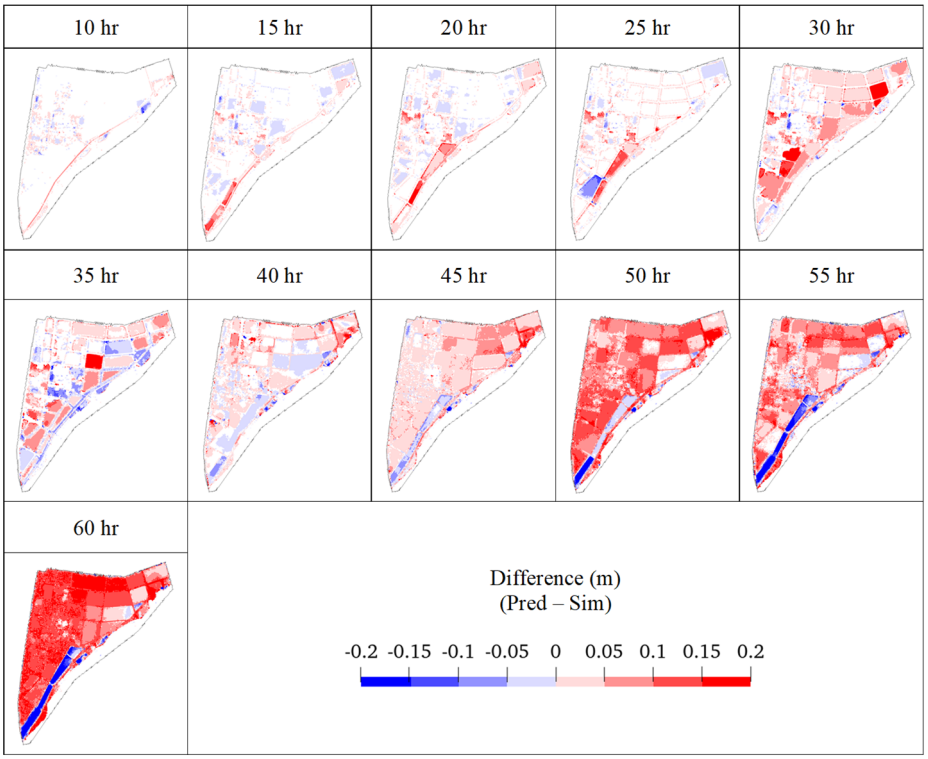


Fig. 6-42 Difference contour comparison of CMB4_EXT200 (2018 rainfall).

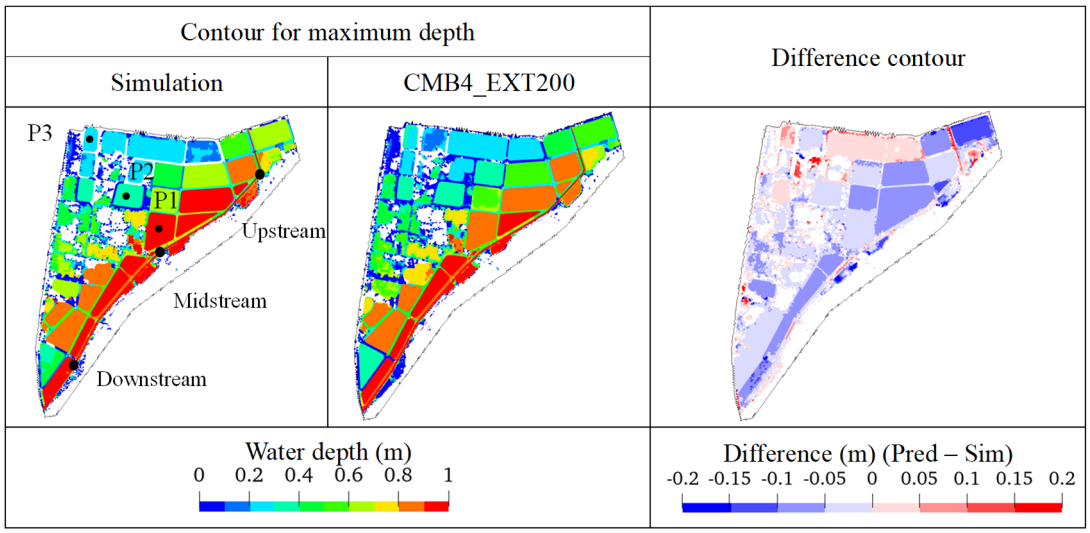


Fig. 6-43 Comparison of maximum depth contour (2021 rainfall).

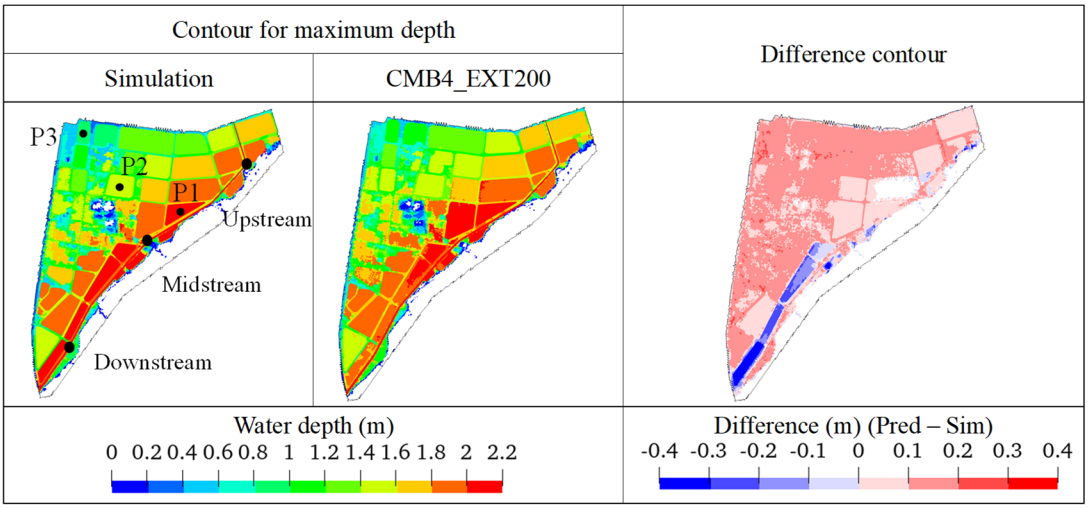


Fig. 6-44 Comparison of maximum depth contour (2018 rainfall).

RF models ensembles the diversity of trees. Even trained with the exact same training data, more trees in the model result in stronger averaging process. While such models are more robust to extreme events, local variances would be significantly generalized. As a result, CMB4_EXT200 performs slightly better when evaluated against extreme conditions.

Furthermore, the temporal variation of water level simulated by HD model at locations indicated in both **Fig. 6-43** and **Fig. 6-44** is summarized as scatterplots. **Fig. 6-45** illustrates the three locations at the upstream, midstream and downstream of the Karube river of both rainfall events with horizontal axis representing time t and vertical axis being time $t+1$ (10 min later). A strong temporal correlation is observed for both events, indicating a smooth evolution of water level between consecutive 10-min time steps. Similar evaluation is performed for

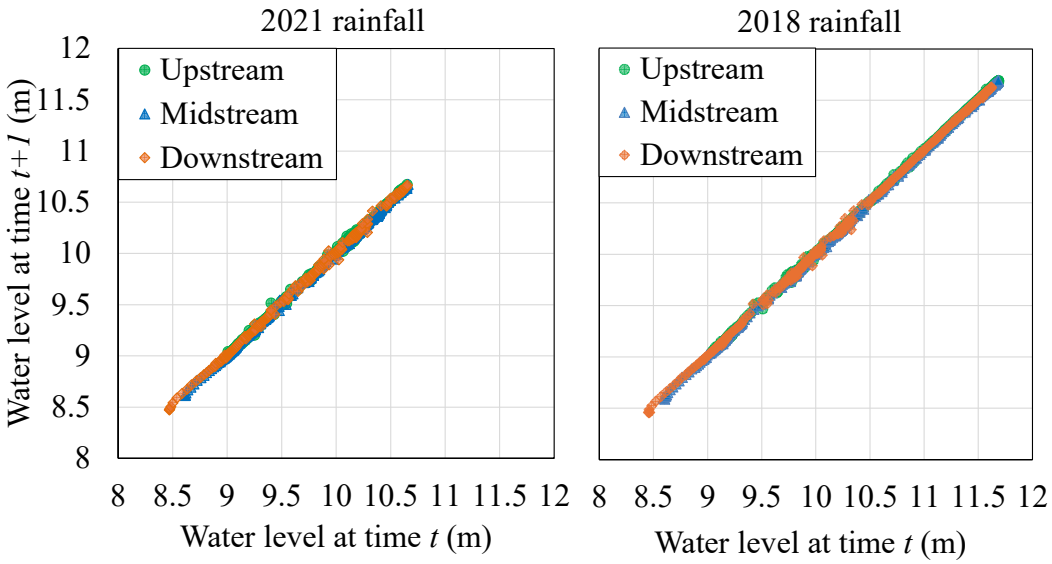


Fig. 6-45 Temporal correlation of water level in the Karube river (HD simulation).

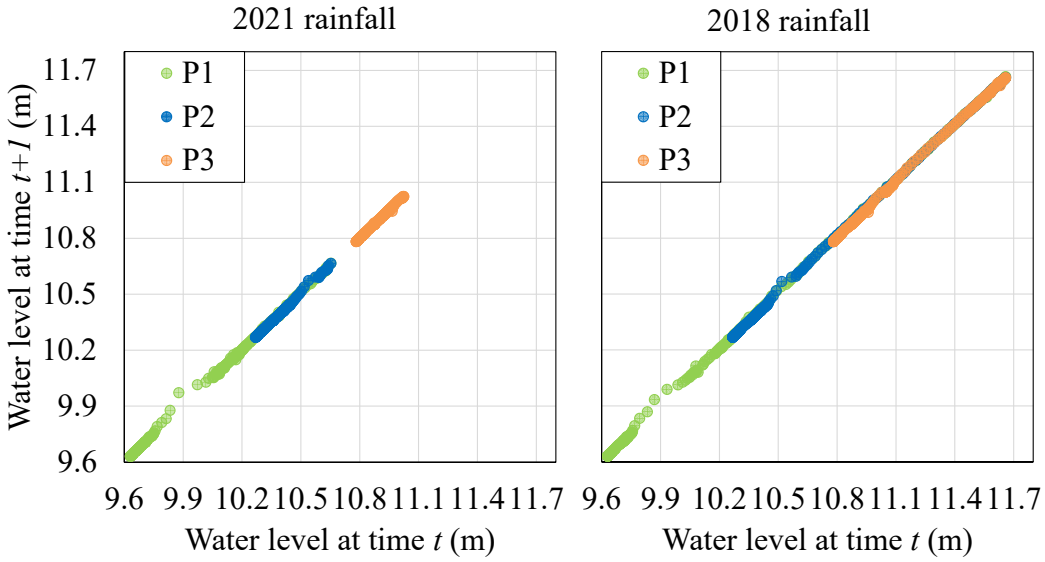


Fig. 6-46 Temporal correlation of water level of different paddies (HD simulation).

paddies with three specific paddies located gradually distant to the Karube river (**Fig. 6-46**), where strong linear correlation is also observed. A slight abrupt incremental change is noted for P1 and P2 for both rainfall and such local changes is indicated to be the time when overflow from paddies occur. These results confirm that the short-term water level dynamics are strongly dominated by local storage effects rather than abrupt fluctuations, and it is considered that RF models captured this statistically strong correlation which contributed to the overall high accuracy even with error accumulation over time.

Regardless of the minor difference in prediction performance, the other more important consideration in terms of practical application of the RF model is the computational cost.

Table 6-6 Detailed comparison of CMB4_EXT configured with different number of trees.

Number of trees	Model size (GB)	Training time (s)	Load time (s)
100	3.93	9367.7	360.5
200	7.86	24901.4	727.0

Table 6-7 Computational cost comparison.

Rainfall event	HD model (GPGPU)	RF model (CPU)	
		CMB4_EXT	CMB4_EXT200
2018	140 min	8.8 min	9.1 min
2021	114 min	7.3 min	7.7 min

Faster predictions are favored since prediction latency, or the sufficient lead time for authorities and individuals to prepare for non-structural measures such as evacuation, becomes critical in emergency situations. **Table 6-6** summarizes the model size, training time and load time of the two models. The size of output RF model doubles as the number of trees increases from 100 to 200. Training time significantly increases by more than twice. Even after model configuration, the computational cost involved in the prediction process of practical applications consists of the time to load model, make predictions, output as visualizable format, and most importantly, the data processing between iterations. The time required to load the pre-trained model approximately doubles as the number of trees double. The computational time needed for the actual prediction process also varies slightly even if the model is loaded ahead of time. Summarized in **Table 6-7** is the computational cost comparison of different models. HD model required 140 min and 114 min for the 2018 and 2021 rainfall event respectively. In comparison, the predictions performed with RF models reduced the calculation time within 10 min for both events. Considering that this calculation time also included the time consumed in data reconstruction between iterations, the RF models are significantly efficient compared to HD models. Currently the RF models are constructed and predicted using CPU, and the prediction time is expected to further shorten with the assistance of GPU in practice. As for the difference in computational cost between two RF models, that model with a greater number of trees requires more time for prediction. This difference again introduces the trade-off between prediction accuracy and computational cost. For the target rainfall events investigated in this study, doubling the number of trees only induces a minor difference in accuracy while resulting in significant increase in memory requirement and computational time in the training phase. And this increase in computational cost is expected to become notable for future operational purposes when predictions over larger regions or longer periods are necessary.

Overall, the difference in the predictive performance of the two models highlights a balance between generalization capability to common patterns and stability under extreme events. Together with a variation in the requirements of computational cost, the results emphasize a subtle influence of hyperparameters across different rainfall scenarios and the necessity for careful tuning in consideration with target rainfall events.

6.4 Chapter summary

This chapter further investigated the potential of applying RF models, trained with simulation results of HD model under hypothetical rainfall pattern, to predict the spatiotemporal flood propagation of real rainfall events with efficiency. The main findings of this chapter are summarized below.

1. The most important feature used in model configuration was the prior state of depth and the least was rainfall information. The observations were consistent with the logistics of HD model that the current water depth is dominated mostly by the prior state of the system rather than direct precipitation.
2. When evaluated against 2018 and 2021 rainfall events, error accumulation over time becomes more notable as induced by the increasing number of iterations. The CMB4 model, trained with simulation results of only 5-hr, resulted in a significant increase in RMSE at the time when cumulative rainfall of the target event exceeded the maximum range of training data. In comparison, CMB4_EXT significantly improved the prediction performance with a final overall RMSE of 0.15 m and 0.05 m for the two rainfall events respectively since the training data further included scenarios under extreme events with 10-hr durations. The difference in model performance provided evidence for the limited capability of RF models to extrapolation, which highlighted the importance of careful selection of training data in correspondence to the target rainfall event.
3. The capability of the RF model to spatially predict for the occurrence of flood was evaluated through CSI and contour map of flood arrival time. Both of the evaluations indicated that CMB4_EXT succeeded in capturing the flood propagation of the two rainfall events. The quantitative evaluation of inundation area and inundation volume also resulted in high coincidence with simulation results. The evaluation of longitudinal water level further suggested that RF models struggled to provide accurate prediction during the early stage of rainfall event due to significant temporal longitudinal gradient. The performance was relatively high as longitudinal gradient diminishes and the predicted results were highly consistent around the time when overflow occurs for both rainfall

events. The performance of CMB4_EXT in predicting for maximum water level was also high, which highlighted the potential of this method to be applied for future practical purposes.

4. The performance of RF model is significantly sensitive to the selection of training data. The inclusion of training data on extreme events overall improved the generalization ability of the model since extrapolations were avoided. On the other hand, the ensemble algorithm of RF model also introduced potential for predictions biasing towards averaging. Doubling the number of trees in the model also resulted in slightly different performance. Yet the difference was considered subtle across the different scenarios evaluated in this research.
5. The computational cost during the training phase was highly dependent on the parameter settings of the model. Once being trained, the RF models were able to significantly reduced the computational cost compared to HD models, even taking consideration of the time needed for dataset processing between iterations.

References

1) Ministry of Land, Infrastructure, Transport and Tourism: Water Information System. http://www1.river.go.jp/ (Accessed November 11, 2024).
2) Crisis Management Department of Soja City: Memorial Journal of Heavy Rain Disaster Response of July 2018. https://www.city.soja.okayama.jp/s/kikikanri/kurashi/bousai/h30gouusaigai_kiroku/h30gouusaigai_kiroku_top.html (Accessed May 22, 2025).
3) Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3), e1301. https://doi.org/10.1002/widm.1301

CHAPTER 7 CONCLUSIONS AND FUTURE PERSPECTIVES

7.1 Conclusions

This research proposed an iterative hydrodynamic-simulation-informed random forest (RF) model targeting for efficient spatiotemporal flood predictions under practical rainfall conditions. Taking advantage of the high-accuracy aspect of conventional hydrodynamic (HD) models and efficiency aspect of machine learning (ML) models, this augmentation approach aims to solve for data scarcity concerns regarding observable extreme events and provide detailed flood prediction results with sufficient lead time for practical applications. With Chapter 1 and Chapter 2 introducing the motivations of this study and methodology of the augmentation approach respectively, Chapter 3 through Chapter 6 each investigated different aspects from method validation to practical applications of this proposed approach. The main conclusions for each chapter are as follows.

Chapter 3 carried out numerical experiments under constant rainfall conditions as a validation for the proposed augmentation approach. Performed with a simple bowl-shaped hypothetical terrain setting, the first numerical experiment indicated that including prior state information of adjacent meshes significantly improved model performance by providing more accurate classification on inundation areas and predictions of water depths. The construction of RF model targeting more complex terrain settings with different land use conditions further required more features such as elevation of neighbor meshes. As another dominant factor for control volume within computational domain of prior state, feature indicating river boundary conditions was also important in improving model performance. The representation of land use with one hot encoding method did not significantly affect model accuracy. The iterative prediction process suggested that this approach is sensitive to initial error and error accumulation during long-term predictions. Under the settings of the constant rainfall conditions over hypothetical terrain settings, an iteration interval of 2.0-hr was able to provide high accuracy predictions with overall RMSE within 0.02 m.

Chapter 4 applied the augmentation approach of HD model and RF model to the Karube river basin. Targeted on the downstream area where severe inundation occurred frequently in history, the RF model was configured with simulation results of HD models under constant rainfall conditions. The feature importance suggested a more significance in features indicating inflow and outflow boundary conditions due to the complexity of flux exchange in practical

terrains compared to the conditions carried out in the previous chapter. Using an iterative prediction interval of 2.0-hr, the accuracy of prediction decreases significantly around the time when overflow occurs, even provided with accurate prior information. While reducing computational cost, such interval failed to properly capture intense changes in hydraulic characteristics, especially before and after overflow occurs. Evaluations performed with a shortened iteration interval of 10-min allowed for better reflection of immediate dynamic changes and increased local performance. Yet a shortened interval required a greater robustness of RF model due to increased number of iterations, and the model when tested against real historical rainfall events of 2021 suggested a significant degradation in performance. The results of this chapter highlighted the importance of incorporating more diverse hypothetical rainfall patterns in the construction of training data, and refining feature selections in order to improve model generalization capability for practical rainfall conditions.

Chapter 5 further investigated hypothetical rainfall patterns of temporal varying intensities together with refinements of features in the configuration of RF models. Using the same features as the previous chapter, the evaluations carried out under rainfall conditions with temporal intensity variations resulted in a compromised performance in the predictions of both inundation area and volume, suggesting that the features were not sufficient enough to represent complex hydrodynamic properties. Based on the results of feature importance and observations from contour maps, investigations were then performed by converting the less representative features into prior state of water level of adjacent meshes and land use indicators of dynamic features of runoff coefficients. The indication of river boundary conditions was also revised into water level at the inflow and outflow mesh considering the fact that water levels are more measurable than discharges in practical applications. With such features refinements, the performance of the model was significantly improved. Evaluations were then carried out investigating how the selection of rainfall patterns in the training data affects the generalization capability and overall performance of the RF model. Including diverse patterns of training data enhanced the generalization capability of the model. The results suggested that the performance of RF models was significantly dependent on the selection of training data due to the limited capability of extrapolation. On the other hand, such large database also led to the potential for conservative prediction due to the ensembled nature of RF regression algorithms such that predicted results tend to average towards mean values when targeting extreme or rara events.

Chapter 6 investigated the potential of applying RF models, trained with simulation results of HD model under all hypothetical rainfall patterns, to predict the spatiotemporal flood propagation of real rainfall events with efficiency. Evaluated against 2018 and 2021 rainfall events, error accumulation over time became more notable as induced by the increasing

number of iterations. The CMB4 model, trained with simulation results of only 5-hr, resulted in a significant increase in RMSE at the time when cumulative rainfall of the target event exceeded the maximum range of training data. CMB4_EXT, the RF model incorporating rainfall patterns of extreme conditions, significantly improved prediction performance. This latter model also successfully captured the overall flood propagation of the two rainfall events. The quantitative evaluation of inundation area and inundation volume also resulted in high coincidence with simulation results. While the evaluation of longitudinal water level of the Karube river resulted in a slightly compromised performance during early stage of event, the high performance around the time when overflow occurs and for the prediction of maximum water level during both events suggested that the model holds the potential for future practical applications. As for the computational cost aspect, the time required for the training phase was highly dependent on the parameter settings of the model. Once being trained, the RF models were able to significantly reduced the computational cost compared to HD models, even taking consideration of the time needed for dataset processing between iterations.

7.2 Future perspectives

This research targeted for spatiotemporal flood prediction using RF model trained with simulation results from HD models under hypothetical design rainfall events. Tested over the Karube river basin and against two historical rainfall events, the augmentation approach was proved to reproduce the flood propagation process with high coincidence compared to simulation results while significantly reducing computational costs. Limitations also arise from this investigation and are summarized as following.

1. The applicability of the augmentation approach was evaluated against the Karube river basin and the results suggested overall high predictive performance. On the other hand, this region inherits relatively simple topographical characteristics where inundation naturally accumulates over the downstream area. The computational conditions evaluate in this research were also simplified where boundary conditions at upstream weir and downstream sluiceway were both closed throughout calculation period and no pump drainage outside the river basin was considered. Therefore, the generalization capability of this augmentation approach over other areas such as urban areas where terrain settings and inundation characteristics are more complex are still to be investigated. The feasibility of the approach under other rainfall conditions, such as events with complicated spatial distributions, rainfalls with multiple severe peaks, and situations where both pluvial and fluvial inundations are induced, should also be evaluated in future studies.

2. The selection of features was carried out under assumptions that precipitation and river boundary conditions are known inputs. Further studies are needed to cooperate with time-changing aspect of rainfall forecast in real-life applications and the availability of real-time observations of river conditions. As also concluded in this research, the iterative performance was still subjective to long-term error accumulations, where non-iterative predictions resulted in relatively higher accuracy. Therefore, the feasibility of calibrating RF models from time to time over long-term predictions using parallelly simulated results from HD models should be investigated in practical applications.
3. The ensemble prediction process of RF model induced an overall high generalization capability when the training data covers rainfall patterns with greater diversity in terms of shape, peak distribution and cumulative rainfall. Yet this property also results in relatively less sensitivity for local variations. As a result, the selection of training data should take consideration of target rainfall events, for example constructing multiple RF models to target for medium scales and extreme events separately, instead of incorporating all scales within one single model. Feature selection could also be further refined, especially regarding the meshes at the boundary line of target area where information for adjacent meshes is absent and regarding situations when the no-flux exchange assumption is no longer valid. Since the evaluations suggested that cumulative rainfall an important factor dominating the generalization capability of the RF model, direct indicators for such information such as time steps and cumulative rainfall could be included in model construction. Furthermore, other machine learning models, especially deep learning models should be investigated since such models are more suitable for capturing the non-linear relationships between features and for extrapolations of untrained events.
4. While the computational cost of constructed RF models was highly reduced compared to HD models, greater efficiency is expected by compiling the current process over GPU. The data processing procedure between iterations could also be further optimized by removing unnecessary information.
5. Evaluations regarding whether this augmentation approach could provide information accurate and efficient enough for decision-making under emergency situations such as releasing early flood warning and provide evacuation routes should be carried out with more details. While this research provided a short investigation regarding critical success index and comparison of flood arrival time contour maps, detailed quantitative analysis is still necessary regarding flood damage reduction aspect for future applications.