

# Urban flash flood forecast using support vector machine and numerical simulation

Jun Yan, Jiaming Jin, Furong Chen, Guo Yu, Hailong Yin and Wenjia Wang

## ABSTRACT

In order to provide urban flood early warning effectively, two support vector machine (SVM) models, using a numerical model as data producer, were developed to forecast the flood alert and the maximum flood depth, respectively. An application in the urban area of Jinlong River Basin, Hangzhou, China, showed the superiority of the proposed models. Statistical results based on the comparison between the results from SVM models and numerical model, proved that the SVM models could provide accurate forecasts for estimating the urban flood. For all the rainfall events tested with an identical desktop, the SVM models only took 2.1 milliseconds while the numerical model took 25 hours. Therefore, the SVM model demonstrates its potential as a valuable tool to improve emergency responses to alleviate the loss of lives and property due to urban flood.

**Key words** | forecast, numerical simulation, support vector machine, urban flood

**Jun Yan** (corresponding author)

**Furong Chen**

**Guo Yu**

**Wenjia Wang**

DHI-China,

Room 307, Guyi Road 181-A, Shanghai,

China

E-mail: [cookie\\_yan@live.cn](mailto:cookie_yan@live.cn)

**Jiaming Jin**

Hangzhou City Comprehensive Transportation

Research Center,

Zhonghezong Road 275-1, Shangcheng District,

Hangzhou city, Zhejiang Province,

China

**Hailong Yin**

College of Environmental Science and Engineering,

Tongji University,

1239 Siping Road, Shanghai 200092,

China

## INTRODUCTION

Urban flood is a serious problem in cities worldwide, especially with increasing climate change and urbanization. Urban storm sewer systems are often incapable of withstanding and satisfying the current and future demands of urban development. Therefore, when sudden heavy rain arrives, serious inundations often occur due to the incoming rainfall intensity exceeding the design intensity of the storm sewer system. Infrastructure damage and property losses caused by extreme rain event could be alleviated with the aid of urban flood forecast systems.

Numerical models based on physical process are commonly used as hydrodynamic simulation tools to forecast urban flood (Bach *et al.* 2014; Leskens *et al.* 2014; Chen *et al.* 2015). However, they usually require sizable computing time (Casulli 2009; Stelling 2012). In order to achieve urban flood forecast over short temporal scale, support vector machine (SVM) supported with available data could be an alternative solution.

Support vector machine (SVM), a kind of machine learning, has been exploited for hydrological forecasting

recently. For example, Li *et al.* (2014) used the SVM and data assimilation method to simulate the rainfall–runoff process in real time flood forecast. Nikam & Gupta (2014) developed a SVM-based model for short-term rainfall forecasts. Yu *et al.* (2004) adopted chaos theory and SVM to hydrologic forecasting. The application to two daily runoff time series showed a slight accuracy improvement and high forecast efficiency. Huang *et al.* (2017) applied the chaotic PSO-SVM to predict the daily groundwater levels in the Huayuan landslide and the weekly, monthly groundwater levels in the Baijiabao landslide. Lin *et al.* (2013) developed a forecasting model to yield 1- to 3-h lead time inundation maps; in their study, the SVM was used to develop the control point inundation forecasting module. However, like all data-driven models, without enough data to support the SVM model, the predictions usually substantially and unsteadily deviate from observations (Sivapragasam *et al.* 2001; Babovic 2005; Sivapragasam & Liong 2005). The current circumstance is that few cities have installed enough urban flood depth monitors. The insufficiency of measured

data limits the application of SVM in urban flood forecasting. It will take a great deal of time and money to develop the SVM model to forecast urban flood maps just using monitoring.

The aim of the present study was to develop an urban flood forecast framework combining a numerical model based on MIKE FLOOD (DHI 2016c) with SVM models. The numerical model was the data source for the SVM model, and the SVM model provided fast forecast. The comparison between the results from SVM models and the numerical model is presented, and the performance of the SVM models evaluated. With this approach, the real-time urban flood forecast system can be developed with limited monitoring data and cost. For many cities that have suffered from urban flood, this approach is one in which government officials and disaster management officers are interested. Drainage system operators can make more effective decisions in advance with the support of a real-time flood forecast. Also, citizens can avoid some risk with the early warning. From this perspective, the combination of a numerical model with SVM used in urban flood forecast should be highlighted.

### Hydraulic models for urban flood

Hydraulic models have been used in urban flood simulation, drainage system planning, and disaster decision support systems for many years. The drainage systems are commonly modeled using a one-dimensional (1D) modeling approach which lacks information about flooding on the urban surface. One approach for urban flood modeling, where the urban surface is treated as open channels and connected to the drainage system, is usually called a coupled 1D/1D model. However, because of the complex urban topography, the flows on the urban surface usually are very different from the flows in channels.

In recent years, the coupled 1D/2D methodologies, in which the urban surface is modeled through two-dimensional (2D) flow approaches and coupled with the 1D pipe network model, have been studied and applied to many cases. There are many features of urban surface, such as roads, buildings, walls, and so on. These features, especially buildings, will change the direction and velocity of the flood water, and generate different kinds of complex flow paths.

Coarse grid resolutions may distort or lose the information on the buildings. Models with finer grid resolutions could provide higher accuracy and a better description of physical processes. However, the 2D models usually require considerable computing resources. When using very small grids, the computational time will increase rapidly. This is the major challenge for the applications to urban flood forecast systems, especially in large areas. Various approaches have been developed to improve the efficiency of the 1D/2D modeling. Some of them are as follows.

#### Local fine grid

Both structured and unstructured mesh can be used in a fine grid in urban areas (or important areas), and coarse grid in rural areas (or unimportant areas) (DHI 2016b). However, the modelers have to do more pre-processing to the mesh. If the study area is an urbanized area, the percentage of fine grid would be very high, and the efficiency will not increase significantly.

#### A method based on sub-grid scale porosity treatment

In McMillan & Brasington (2007), the porosity information of fine grid is represented within a coarse grid through sub-grid scale parameterization. The depth-dependent porosity function, which can reflect the difference between assumed planform storage volume in a coarse cell and the actual volume in sub-grid features, is developed to adjust the continuity equation. For further examples and information, the reader is referred to Yu & Lane (2006) and McMillan & Brasington (2007).

#### A method based on adjusted conveyance and storage characteristics

Vojinovic *et al.* (2013) adopted this approach for a hypothetical and real-life case study. This method uses coarse grid resolutions in 2D non-inertial models. The volume–depth relationship and area–depth relationship are calculated for transition between fine and coarse cells. Meanwhile, the friction values of the coarse-resolution model are adjusted to replicate the results of the fine grid model. The work of Vojinovic *et al.* (2013) showed that this method achieved

results close to a fine-resolution model with little increased computing time.

### Multi-cell overland solver from pre-simulations

The modified equations are solved on a coarse grid taking the variation of the bathymetry within each grid cell into account. Then, the coarse grid results are determined based on the fine scale bathymetry (DHI 2016b).

### Multilayered approach

This approach adopted by Chen et al. (2012) improves coarse-resolution through bringing the building coverage ratio and conveyance reduction factor into the 2D model to represent building features within a coarse grid. Instead of using a single layer which fails to match the flow phenomena when a coarse cell is bisected by a building, a multilayered approach is adopted. The parameters of the cell in each layer are specified. With the multiple layers, a coarse grid model can achieve a high-resolution result with little increase in computing time.

Most of these approaches increase the efficiency through using a coarse grid instead of fine grid, while representing key features within the coarse grid. The efficiency and accuracy depends on how coarse the grid is. The modelers have to make a considerable effort to balance the accuracy and computational speed. Usually, a 10 m to 20 m grid has been selected as the coarse grid (McMillan & Brasington 2007; Abdullah et al. 2012; Chen et al. 2012; Vojinovic et al. 2013). For a large area, the efficiency of this grid size may be not enough for real-time forecast.

### Overview of SVM model

SVM, which is based on the basis of the structural risk minimization principle, has been verified to be a robust and efficient algorithm for equation fitting, data analysis, hydrological forecasting, and so on (Collobert & Bengio 2001; Yu et al. 2004; Deris et al. 2011; Wang et al. 2013; Atiquzzaman & Kandasamy 2016). In solving small sample, non-linear, and high-dimensional pattern recognition, SVM has special advantages. For more detailed information on this subject,

the reader is referred to Cortes & Vapnik (1995). SVM can be used in both classification and regression problems.

In practical application, the SVMs models are trained through the training data set (Equation (1)). For classification problems, it means finding the hyper plane (Equation (2)) with the largest geometrical margin, and for regression problems, it means finding linear regression function (Equation (3)) which can best approximate the output vector with an error tolerance  $\varepsilon$ .

$$S = \{(x_1, y_1), \dots, (x_l, y_l)\}, x_i \in R^n, y_i \in \begin{cases} \{0, 1\}, & \text{for classification problems} \\ R, & \text{for regression problems} \end{cases} \quad (1)$$

where  $x_i$  is the input vector processed by the data preprocessor,  $y_i$  is the output data.

$$\omega \cdot x_i + b = 0 \quad (2)$$

$$f(x) = \omega \cdot x_i + b \quad (3)$$

where  $\omega$  is the weight vector and  $b \in R$  is the threshold value.

The primal problem can be described as Equation (4) (for classification problems) and Equation (5) (for regression problems):

$$\min \frac{1}{2} \|\omega\|^2 \text{ s.t. } y_i(\omega \cdot x_i + b) \geq 1, i = 1, \dots, l \quad (4)$$

where  $l$  is the number of training vectors.

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t. } f(x_i) - y_i \leq \xi_i^* + \varepsilon, i = 1, \dots, l \\ \text{and } y_i - f(x_i) \leq \xi_i + \varepsilon, i = 1, \dots, l \\ \text{and } \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \quad (5)$$

where  $C$  indicates the capacity parameter cost,  $\xi_i$  and  $\xi_i^*$  determine the degree to which sample points are penalized if the error is larger than  $\varepsilon$ .

The principal concepts behind SVM methodology are mapping the input data set into a high-dimensional feature space to deal with the nonlinear classification or regression. In detail, through a Lagrange duality (Equation (6)) and

mapping the input space into a higher dimension with the function  $K(\cdot, \cdot)$ , which is called kernel function, the dual problem can be formed (Equation (7) for classification problems, Equation (8) for regression problems). The common kernel functions include linear kernel, polynomial kernel, and radial basis function.

$$L(\omega, b, \alpha) = \frac{1}{2} \|\omega^2\| - \sum_{i=1}^n \alpha_i (y_i (\omega \cdot x_i + b) - 1) \quad (6)$$

$$\begin{cases} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.}, \sum_{i=1}^l \alpha_i y_i = 0 \\ \text{and } \alpha_i \geq 0, i = 1, \dots, l \end{cases} \quad (7)$$

where  $\alpha_i$  is the vector of non-negative Lagrange multipliers.

$$\begin{cases} \min \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) \\ + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t.}, \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ \text{and } 0 \leq \alpha_i \leq C, i = 1, \dots, l \\ \text{and } 0 \leq \alpha_i^* \leq C, i = 1, \dots, l \end{cases} \quad (8)$$

The optimal Lagrange multipliers are solved by the standard quadratic programming algorithm. The reader can find more details of the quadratic programming algorithm from [Chang & Lin \(2011\)](#). The training vectors for which the Lagrange multiplier is non-zero are called support vectors. The decision function (Equation (9) for classification problems, Equation (10) for regression problems) can be used to calculate new input vectors:

$$f(x) = \sum_{i=1}^{l_k} \alpha_i y_i K(x_i, x) + b \quad (9)$$

where  $f(x) = 0$  or  $1$ , means different classes;  $l_k (l_k < l)$  is the number of support vectors.

$$f(x) = \sum_{i=1}^{l_k} (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (10)$$

## METHODS

### Study area and data set description

The study area is the urban area of Jinlong River Basin, Hangzhou, China, with an area of 4.5 km<sup>2</sup>. In the study site, the Hupao Road is a main traffic road. Whenever the Hupao Road suffers a flood, the road is blocked. For example, during the 'Feite' typhoon, October 2013, this road was blocked for several days. This area has been reserved as a historic scenic spot and new constructions are strictly restricted. Potential measures, such as building a new storm drainage system, are limited. Therefore, a flood forecast system is a helpful tool for emergency response to alleviate the flood damage, e.g., arranging emergency pumps before heavy rainfall occurrence.

In the study area, three rainfall gauge stations and two water level gauge stations were installed ([Figure 1](#)). The monitoring data during the year of 2013, the terrain data, Jinlong River data, and local rainfall statistic data were employed in this study. Annual accumulated rainfall during 2013 was about 1,431.4 mm, which included two storm events (June 26th–29th and October 6th–8th).

### Framework of urban flood forecast

A framework of the urban flood forecast in this study is shown in [Figure 2](#). It includes two components: i.e., MIKE FLOOD and SVM. In this framework, the MIKE FLOOD model provides surrogate data for training SVM models. Combination of the two models is able to provide a good way to deliver their advantages and achieve high accurate and real-time urban flood forecast.

MIKE FLOOD is a physical process-based modeling suite developed by the Danish Hydraulic Institute (DHI), and has shown great accuracy in simulating urban flood events ([Patro et al. 2009](#); [Hlodversdottir et al. 2015](#)). The core hydrodynamic functions of MIKE FLOOD are the Saint-Venant equations and the vertically integrated shallow water equations ([DHI 2016a](#); [2016b](#)). Any numerical model based on similar hydrodynamic functions can be used to combine with SVM.

In this study, LIBSVM, which is an open source software for support vector classification (SVC), regression

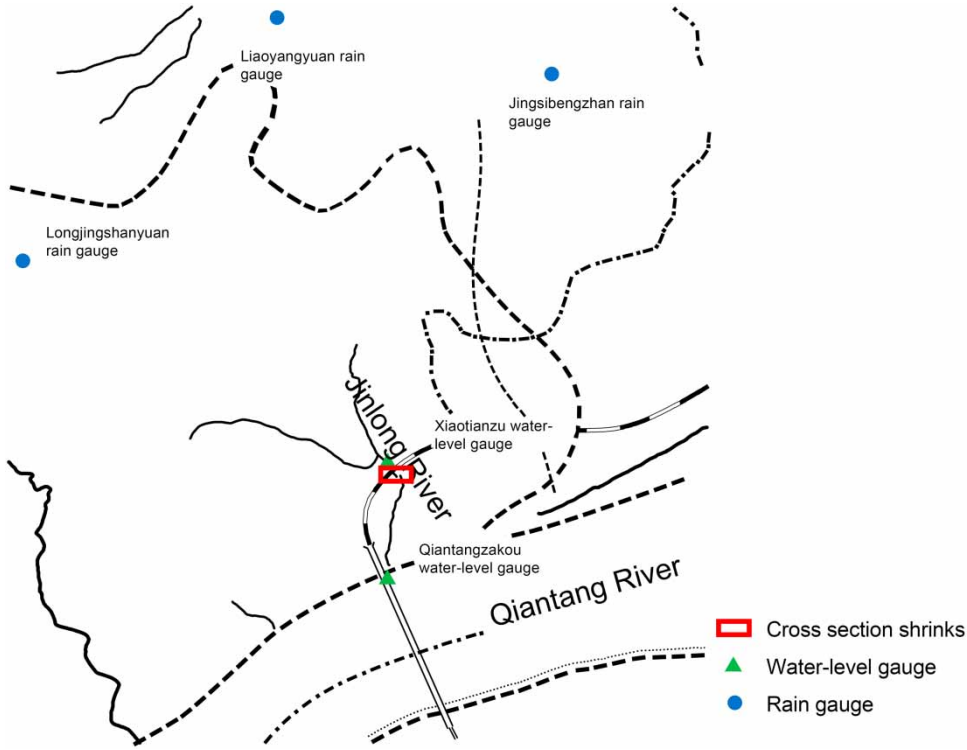


Figure 1 | Description of the study site.

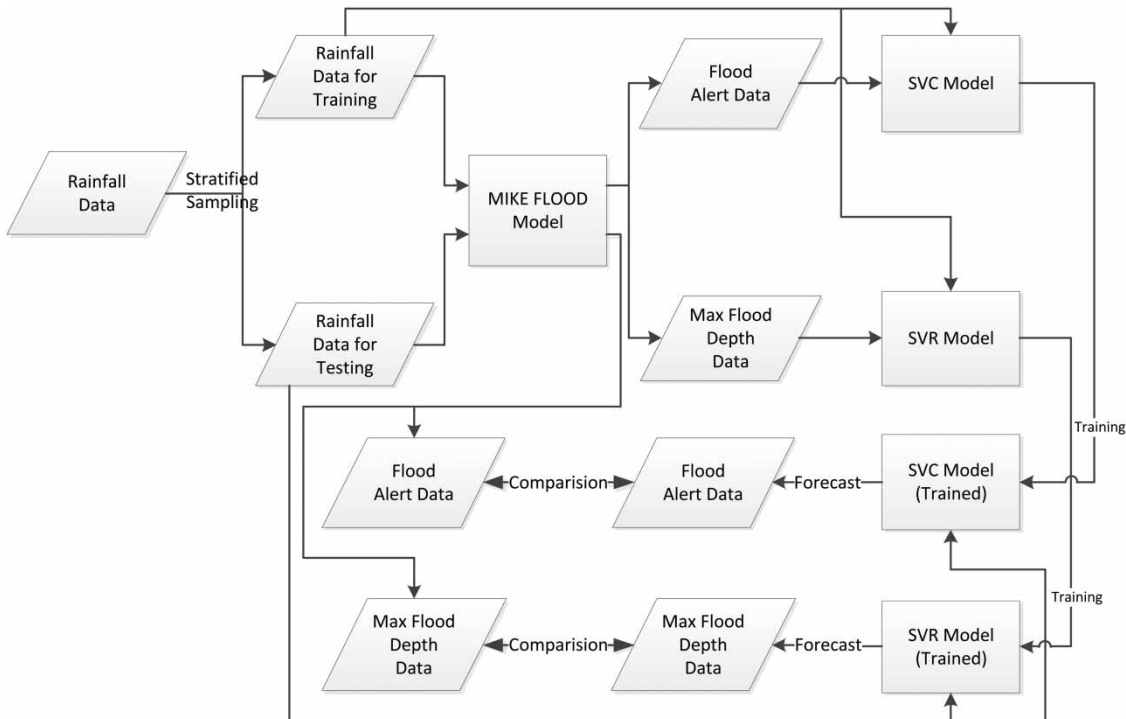


Figure 2 | Framework of urban flood forecast combining SVM and numerical model.

and distribution estimation, is selected to establish the SVM models (Chang & Lin 2011). Two SVM models were established in this paper: the maximum flood depth was forecast by a support vector regress (SVR) machine, and the urban flood alert was forecast by a SVC machine.

Both SVM models included training and testing steps. In the training step, the flood data extracted from MIKE FLOOD were used for training the SVM models. In the testing step, the MIKE FLOOD model results were used to assess the performance of the trained SVM models. Then, the trained SVM models could be used in urban flood forecast.

### The data set pre-processor

The data pre-processor includes rainfall events sampling, rainfall data dimension reduction, the max flood depth, and urban flood alert labels prepared.

The rainstorm intensity formula based on local rainfall data of the past 30 years was used to evaluate the return period of each rainfall event. The frequency distribution of the rainfall events by rain intensity throughout the year of 2013 is presented in Figure 3. An extremely uneven distribution may be observed. As seen in Figure 3, the return period of 69 events in a total of 72 rainfall events is less than one year. The low probability of the high return period event may result in the bias of the SVM model trained by these data. Therefore, a series of new storm

events was generated through selecting actual rainfall events randomly, and making random amplifications. The generated rainfall events and the actual rainfall events were mixed. The frequency distribution of these events is presented in Figure 4.

Then, through stratified sampling, the rainfall events were selected for training and testing the SVM models. In principle, the samples should cover as wide a range as possible. In practice, when an extreme heavy storm occurs, the emergency management officers will execute the highest level emergency plan. The sample data set should include this extreme condition. In that case, even if heavier storms occur, the forecast biases of extrapolating SVM models beyond the range of the training data set usually will not cause any problem for emergency decisions. Therefore, the stratified sampling should be based on the local conditions and cover as many data as possible.

Another issue is that the rainfall duration is different among each event. Thus, the rainfall data cannot be used as the input vector directly. In order to unify and decrease the dimensions of the input vector, an unsupervised four-order polynomial fitting algorithm was applied. The polynomial curves can be described as follows:

$$R(t) = a_0 + a_1t + a_2t^2 + a_3t^3 + a_4t^4 \quad (11)$$

where  $R(t)$  is the rainfall depth, and  $t$  is the time since the rain started.

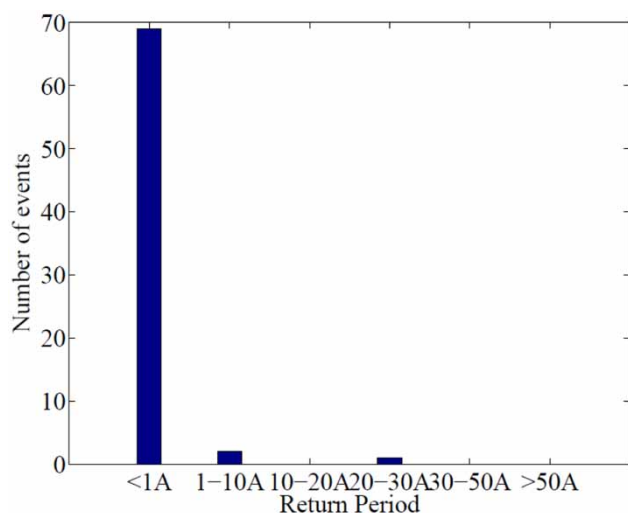


Figure 3 | Frequency distribution of actual rainfall events based on return period.

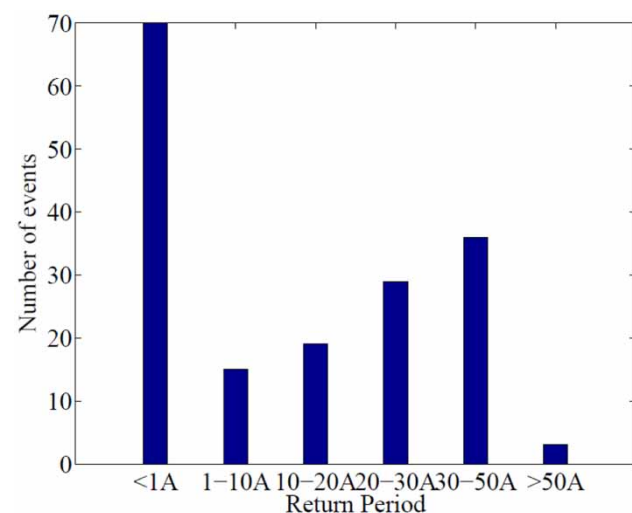


Figure 4 | Frequency distribution of merged rainfall events based on return period.



Then, the coefficients of the rainfall data quartic polynomial curve and three key features (the maximum and accumulated value of rainfall depth, the rainfall duration) were merged together as the input vector.

The urban flood alert and max flood depth which are two factors that concern the decision-makers most were forecast for each rainfall event. They were calculated from the MIKE FLOOD simulation results, and provided as the training output data of SVM models. During a rainfall event, if a flood depth above 0.15 meter is sustained for more than 0.5 hour, the alert should be triggered. The alert is labeled as 1 when the urban flood alert is triggered and 0 when no alert is triggered.

### Statistical analysis of SVM models

Six performance measures comparing the flood results simulated by numerical model with the results simulated by SVM models were used to evaluate the performance of SVM models. Four parameters, including root mean square error (RMSE), mean bias error (MBE), coefficient of efficiency (CE), and coefficient of correlation (CC), were used to evaluate the SVR model. Precise rate (PR) and true positive rate (TPR) were used to evaluate the flood alert forecasting accuracy forecasted by the SVC model. Specifically, these parameters are described as follows.

#### 1. The root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (H_{SVM}(k) - H_{MIKE}(k))^2} \quad (12)$$

where  $H_{SVM}(k)$  and  $H_{MIKE}(k)$  are the simulated flood depth by the SVR model and numerical model at event  $k$ , and  $n$  is the total number of rainfall events. RMSE represents the standard deviation of the differences between the values predicted by numerical model and those by SVR model.

#### 2. The mean bias error (MBE)

$$MBE = \frac{1}{n} \sum_{k=1}^n (H_{SVM}(k) - H_{MIKE}(k)) \quad (13)$$

MBE shows the mean value of error between the SVR model and the MIKE FLOOD model.

#### 3. Coefficient of efficiency (CE)

$$CE = 1 - \frac{\sum_{k=1}^n (H_{MIKE}(k) - H_{SVM}(k))^2}{\sum_{k=1}^n (H_{MIKE}(k) - \overline{H_{MIKE}})^2} \quad (14)$$

where  $\overline{H_{MIKE}}$  is the average of simulated flood depth by numerical model. CE is widely used to evaluate the forecasting performance of hydrological models. It is the ratio of the mean square error to the variance in the max flood depth data simulated by MIKE FLOOD subtracted from unity. The more perfect the forecast by the SVR model, the more close to 1 is the CE value.

#### 4. Coefficient of correlation (CC)

$$CC = \frac{\sum_{k=1}^n (H_{MIKE}(k) - \overline{H_{MIKE}})(H_{SVM}(k) - \overline{H_{SVM}})}{\sqrt{\sum_{k=1}^n (H_{MIKE}(k) - \overline{H_{MIKE}})^2} \sqrt{\sum_{k=1}^n (H_{SVM}(k) - \overline{H_{SVM}})^2}} \quad (15)$$

where  $\overline{H_{SVM}}$  is the average of simulated flood depth by the SVR model. The CE and CC are used to measure the similarity between the numerical model and the SVM model forecast result of each rainfall event. The higher the CE (or CC), the better agreement between the flood depth forecasted by the numerical model and the SVR model.

#### 5. Precise rate (PR)

$$PR = (TP + TN)/TSN \quad (16)$$

where  $TP$  is the number of true positives (flood alerts forecasted correctly by the SVC model);  $TN$  is number of true negatives (no alerts forecasted correctly by the SVC model);  $TSN$  is the total sample number. The PR shows the correction of the forecasting by the SVC model.

#### 6. True positive rate (TPR)

$$TPR = TP/(TP + FN) \quad (17)$$

where  $FN$  is number of false negatives (no alerts forecasted incorrectly by the SVC model). TPR shows how many flood events were forecasted correctly by the SVC model.

## RESULTS AND DISCUSSION

### The data set processing

Using the stratified sampling, 77 rainfall events were selected for training the SVM models and 34 rainfall events were selected for testing. Figures 5 and 6 show that the samples are well-distributed using the stratified sampling. In this study site, when the rainfall intensity is bigger than 1-in-50-year rainfall, local officers block the roads and execute the highest level emergency procedures

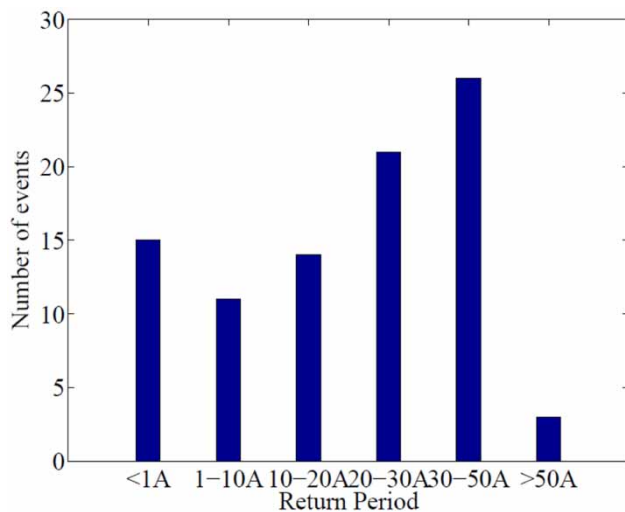


Figure 5 | Frequency distribution of training rainfall events based on return period.

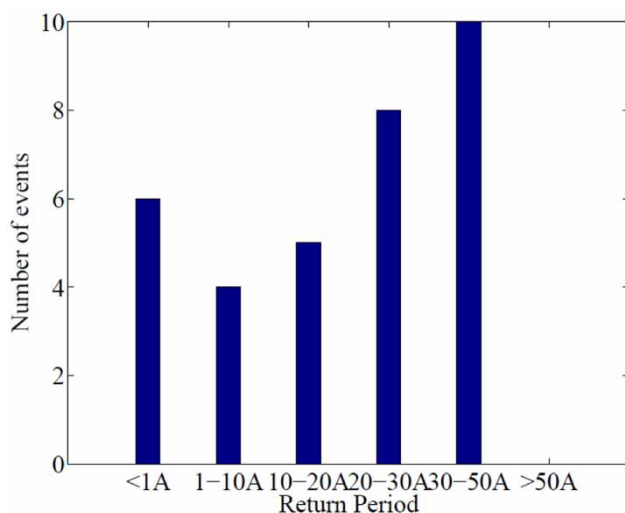


Figure 6 | Frequency distribution of testing rainfall events based on return period.

to avoid other secondary disasters. In terms of the accuracy of urban flood forecasts, the local officers pay most attention to those rainfall events with more frequency than 1-in-50-year incidents. Therefore, the return periods of the rainfall events in the training data set were mainly in the range of 0 to 50 years.

Through the proposed fourth-order polynomial fitting algorithm, the data of each rainfall event was converted into the input vector for the SVM models. Each rainfall event was an input vector, i.e., there were 77 input vectors in the training data set and 34 input vectors in the testing data set.

### MIKE FLOOD model calibration

The water level data from Xiaotianzu gauging station of two storm events were used for model calibration (October 6th–8th, 2013) and verification (June 26th–29th, 2013). This station was installed beside the Hupao Road. When the water level is above the road, the data of this station can show the flood level on the road.

The model calibration result is presented in Figure 7. The absolute biases were below 0.15 m; the maximum water level at the rainfall peak was 8.1 m, which matched well with the observed data. The model verification result (Figure 8) also showed a good similarity between the observed and MIKE FLOOD simulated result. At the

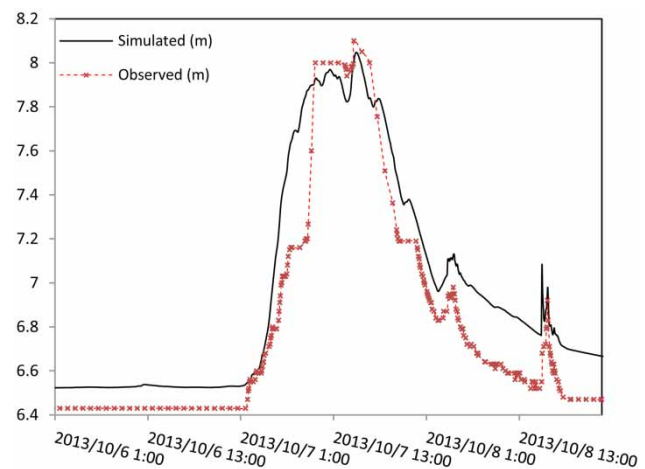
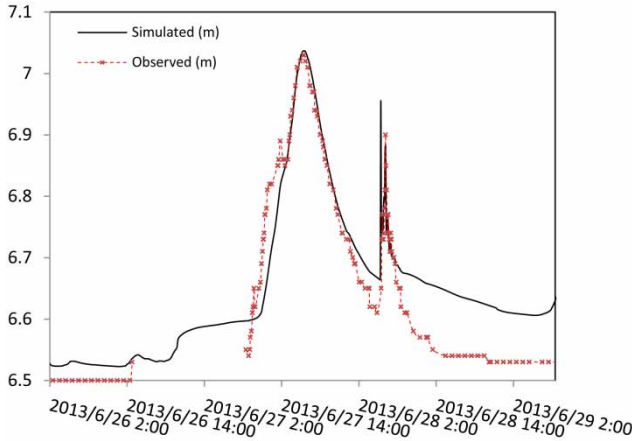


Figure 7 | Calibration results for the MIKE FLOOD model: comparison between observed (corrected by manual record) and simulated river water level (October 6th–8th, 2013).





**Figure 8** | Comparison between observed and MIKE FLOOD simulated river water level (June 26th–29th, 2013).

rainfall peak, the absolute biases were below 0.05 m. Therefore, it is reliable to employ the urban flood depth data produced by the developed deterministic numerical model for further training and testing the SVM model.

**Performance of the SVM models**

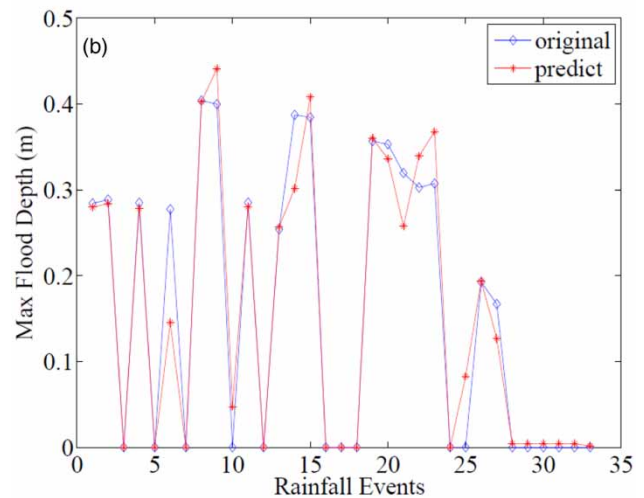
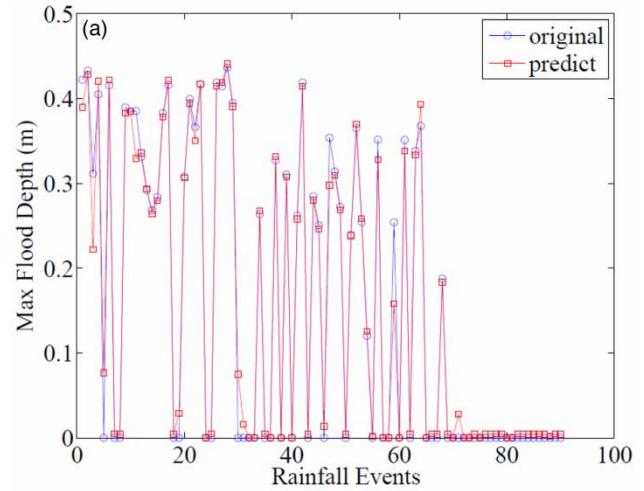
The performance indicators are summarized in Tables 1 and 2. Considering the testing events, PR was 96.9%, TPR was 92.9%, and the RMSE, MBE, CE, and CC was 0.046, -0.011, 0.919, and 0.964, respectively, indicating the SVM models provided similar forecast results as the MIKE FLOOD model. Figure 9 also showed a good agreement between results forecasted by the SVM model and those by the MIKE FLOOD model.

**Table 1** | PR and TPR of 77 training events and 34 testing events for the SVC model

	PR (%)	TPR (%)
Training events	100	100
Testing events	96.9	92.9

**Table 2** | RMSE, MBE, CE, and CC of 77 training events and 34 testing events for the SVR model

	RMSE	MBE	CE	CC
Training events	0.021	0.000	0.985	0.993
Testing events	0.038	-0.001	0.945	0.972



**Figure 9** | Comparison of the forecasted max flood depth using MIKE FLOOD with that using SVM: (a) training events, (b) testing events.

Meanwhile, the SVM models were also evaluated with real rainfall events. From all the rainfall events in 2013, except the two events for calibrating and verifying the MIKE FLOOD model, we selected three of them: one with the biggest rainfall intensity, a very light one, and one with medium intensity. The details of the three events and the evaluation of the SVM models are summarized in Table 3. The absolute errors of the SVR model were all below 0.05 m, and the flood alert judgments of the SVC model were all correct.

The comparisons with the surrogate data and measured data showed the SVM models could achieve high accuracy close to the MIKE FLOOD model. Although the SVM

**Table 3** | SVM models tested by three real rainfall events

Date		2013/3/19	2013/6/19	2013/8/1
Observed	Amount of rainfall (mm)	13.1	42.5	35.0
	Duration (hour)	13.5	1.5	2.0
	Max flood depth (m)	0.00	0.19	0.00
	Flood alert	No	Yes	No
MIKE FLOOD	Max flood depth (m)	0.00	0.17	0.00
	Absolute error (m)	0.00	0.02	0.00
	Flood alert	No	Yes	No
SVM	Max flood depth (m)	0.00	0.21	0.03
	Absolute error (m)	0.00	0.02	0.03
	Flood alert	No	Yes	No

**Table 4** | Computation time analysis of the SVM models and the MIKE FLOOD model

Cost time	MIKE FLOOD model	SVM models
All training events	53 hours	3.7 ms
All testing events	25 hours	2.1 ms

models only forecasted the maximum flood and the flood alert, this information was sufficient for the drainage system operators to execute specific emergency procedures in advance and for the local citizen to avoid losses.

### Efficiency of the SVM models in forecasting urban flood

The MIKE FLOOD model and SVM models were run on the same desktop (Intel® Xeon® CPU E5-2687 W v2, 32G RAM) in order to compare their computation time (Table 4). In detail, for all the training events, the total computation time based on MIKE FLOOD was 53 hours, while the SVM models only took 3.7 milliseconds. The comparison of computation time for test events represents a similar manner: MIKE FLOOD took 25 hours, while the SVM models only took 2.1 milliseconds.

## CONCLUSION

To provide highly effective real-time urban flood forecasts, an urban flood forecast framework combined SVM model with a well-calibrated numerical model based on MIKE FLOOD was developed. In this framework, MIKE FLOOD was used to provide surrogate data which were

used to train SVM models. The trained SVM models were applied to real-time urban flood forecast. This combination provided a new data source for SVM models and high efficiency urban flood forecast.

An application in Jinglong River Basin, Hangzhou, China demonstrated the superiority of the developed models. Through stratified sampling, 77 synthetic rainfall events were selected for training the SVM models, and 34 events were selected for testing. Among performance indicators used to check the performance of the SVM models, the RMSE and MBE were all below 0.05, the PR and TPRH were all above 90%, and the CE and CC were all above 0.9. Then, further evaluation was undertaken with three real rainfall events. The absolute biases were all below 0.05 m. It showed that the SVM models could provide high accurate urban flood forecasts, while offering 15 million times faster calculating speed compared with the numerical model.

In conclusion, a combination of numerical model and SVM model will achieve high solution accuracy and save significant computational time. The engineers can focus on the accuracy and utility of the hydraulic models without considering the forecast efficiency. With a well-established MIKE FLOOD model, the SVM models can be trained and applied to forecast systems. Using this technique, real-time urban flood forecast systems can be developed with limited monitoring costs, which is an advantage of this approach. That means it is easier for government officials to accept this technique. Therefore, the presented methodology demonstrates its potential as a valuable tool to provide a highly effective flood forecast and improve emergency responses to alleviate loss of lives and property due to urban flooding. However, there are some restrictions in the application and further improvement is still needed. As the SVM model is a black-box model, it is hard to set hydraulic structure controls directly in the trained SVM model. When there are a great many structures and the operators want to compare flood results by setting different controls, the application is limited. The approach presented here will be further developed to solve this limitation. In addition, although stratified sampling showed good performance in this study, how to estimate the minimum sample set needs further study.

## REFERENCES

- Abdullah, A. F., Vojinovic, Z., Price, R. K. & Aziz, N. A. A. 2012 A methodology for processing raw LiDAR data to support urban flood modelling framework. *Journal of Hydroinformatics* **14** (1), 75–92.
- Atiquzzaman, M. & Kandasamy, J. 2016 Prediction of hydrological time-series using extreme learning machine. *Journal of Hydroinformatics* **18** (2), 345–353.
- Babovic, V. 2005 Data mining in hydrology. *Hydrological Processes* **19** (7), 1511–1515.
- Bach, P. M., Rauch, W., Mikkelsen, P. S., McCarthy, D. T. & Deletic, A. 2014 A critical review of integrated urban water modeling-urban drainage and beyond. *Environmental Modeling and Software* **54**, 88–107.
- Casulli, V. 2009 A high-resolution wetting and drying algorithm for free-surface hydrodynamics. *International Journal for Numerical Methods in Fluids* **60**, 391–408.
- Chang, C. C. & Lin, C. J. 2011 LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2** (3). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chen, A. S., Evans, B., Djordjević, S. & Savić, D. A. 2012 Multi-layered coarse grid modelling in 2D urban flood simulations. *Journal of Hydrology* **470**, 1–11.
- Chen, Y. B., Zhou, H. L., Zhang, H., Du, G. M. & Zhou, J. H. 2015 Urban flood risk warning under rapid urbanization. *Environmental Research* **139**, 3–10.
- Collobert, R. & Bengio, S. 2001 SVM-Torch: support vector machines for large-scale regression problems. *The Journal of Machine Learning Research* **1**, 143–160.
- Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Machine Learning* **20**, 273–297.
- Deris, A. M., Zain, A. M. & Sallehuddin, R. 2011 Overview of support vector machine in modeling machining performances. *Procedia Engineering* **24**, 308–312.
- DHI 2016a MIKE11, A Modelling System for Rivers and Channels. Reference Manual. DHI Group, Horsholme, Denmark.
- DHI 2016b MIKE21 Tidal Analysis and Prediction Module. Scientific Documentation. DHI Group, Horsholme, Denmark.
- DHI 2016c MIKE FLOOD 1D-2D Modeling. User Manual. DHI Group, Horsholme, Denmark.
- Hlodversdottir, A. O., Bjornsson, B., Andradottir, H. O., Eliasson, J. & Crochet, P. 2015 Assessment of flood hazard in a combined sewer system in Reykjavik city centre. *Water Science and Technology* **71** (10), 1471–1477.
- Huang, F., Huang, J., Jiang, S. H. & Zhou, C. 2017 Prediction of groundwater levels using evidence of chaos and support vector machine. *Journal of Hydroinformatics* **19** (4), 586–606.
- Leskens, J. G., Brugnach, M., Hoekstra, A. Y. & Schuurmans, W. 2014 Why are decisions in flood disaster management so poorly supported by information from flood models? *Environmental Modelling and Software* **53**, 53–61.
- Li, X. L., Lü, H., Horton, R., An, T. & Yu, Z. 2014 Real-time flood forecast using the coupling support vector machine and data assimilation method. *Journal of Hydroinformatics* **16**, 973–988.
- Lin, G. F., Lin, H. Y. & Chou, Y. C. 2013 Development of a real-time regional-inundation forecasting model for the inundation warning system. *Journal of Hydroinformatics* **15** (4), 1391–1407.
- McMillan, H. K. & Brasington, J. 2007 Reduced complexity strategies for modelling urban floodplain inundation. *Geomorphology* **90** (3), 226–243.
- Nikam, V. & Gupta, K. 2014 SVM-based model for short-term rainfall forecasts at a local scale in the Mumbai urban area, India. *Journal of Hydrologic Engineering* **19**, 1048–1052.
- Patro, S., Chatterjee, C., Mohanty, S., Singh, R. & Raghuvanshi, N. S. 2009 Flood inundation modeling using MIKE FLOOD and remote sensing data. *Journal of the Indian Society of Remote Sensing* **37**, 107–118.
- Sivapragasam, C. & Liang, S. 2005 Flow categorization model for improving forecasting. *Nordic Hydrology* **36**, 37–48.
- Sivapragasam, C., Liang, S. Y. & Pasha, M. 2001 Rainfall and runoff forecasting with SSA-SVM approach. *Journal of Hydroinformatics* **3**, 141–152.
- Stelling, G. S. 2012 Quadtree flood simulations with sub-grid digital elevation models. *Proceedings of the Institution of Civil Engineers: Water Management* **165**, 567–580.
- Vojinovic, Z., Seyoum, S., Salum, M. H., Price, R. K., Fikri, A. K. & Abebe, Y. 2013 Modelling floods in urban areas and representation of buildings with a method based on adjusted conveyance and storage characteristics. *Journal of Hydroinformatics* **15** (4), 1150–1168.
- Wang, W. C., Xu, D. M., Chau, K. W. & Chen, S. 2013 Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *Journal of Hydroinformatics* **15** (4), 1377–1390.
- Yu, D. & Lane, S. N. 2006 Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 1: mesh resolution effects. *Hydrological Processes* **20** (7), 1541–1565.
- Yu, X., Liang, S. Y. & Babovic, V. 2004 EC-SVM approach for real-time hydrologic forecasting. *Journal of Hydroinformatics* **6** (3), 209–223.

First received 20 May 2017; accepted in revised form 9 November 2017. Available online 29 November 2017