



Urban pluvial flooding prediction by machine learning approaches – a case study of Shenzhen city, China

Qian Ke^a, Xin Tian^{b,c}, Jeremy Bricker^a, Zhan Tian^{d,*}, Guanghua Guan^e, Huayang Cai^{f,g},
Xinxing Huang^h, Honglong Yangⁱ, Junguo Liu^d

^a Department of Hydraulic Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628CN Delft, the Netherlands

^b Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628CN Delft, the Netherlands

^c KWR Water Research Institute, Groningenhaven 7, 3433PE Nieuwegein, the Netherlands

^d School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

^e State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 430072 Wuhan, China

^f Institute of Estuarine and Coastal Research/ Guangdong Provincial Engineering Research Center of Coasts, Islands and Reefs, School of Marine Engineering and Technology, Sun Yat-sen University, Guangzhou, China

^g Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai)/ State and Local Joint Engineering Laboratory of Estuarine Hydraulic Technology, Guangzhou, China

^h Shanghai Institute of Technology, Shanghai, China

ⁱ Shenzhen National Climate Observatory of Meteorological Bureau of Shenzhen Municipality, Shenzhen, China

ARTICLE INFO

Keywords:

Urban pluvial flooding
Rainfall threshold
Machine learning
Shenzhen city

ABSTRACT

Urban pluvial flooding is a threatening natural hazard in urban areas all over the world, especially in recent years given its increasing frequency of occurrence. In order to prevent flood occurrence and mitigate the subsequent aftermath, urban water managers aim to predict precipitation characteristics, including peak intensity, arrival time and duration, so that they can further warn inhabitants in risky areas and take emergency actions when forecasting a pluvial flood. Previous studies that dealt with the prediction of urban pluvial flooding are mainly based on hydrological or hydraulic models, requiring a large volume of data for simulation accuracy. These methods are computationally expensive. Using a rainfall threshold to predict flooding based on a data-driven approach can decrease the computational complexity to a great extent. In order to prepare cities for frequent pluvial flood events – especially in the future climate – this paper uses a rainfall threshold for classifying flood vs. non-flood events, based on machine learning (ML) approaches, applied to a case study of Shenzhen city in China. In doing so, ML models can determine several rainfall threshold lines projected in a plane spanned by two principal components, which provides a binary result (flood or no flood). Compared to the conventional critical rainfall curve, the proposed models, especially the subspace discriminant analysis, can classify flooding and non-flooding by different combinations of multiple-resolution rainfall intensities, greatly raising the accuracy to 96.5% and lowering the false alert rate to 25%. Compared to the conventional model, the critical indices of accuracy and true positive rate (TPR) were 5%-15% higher in ML models. Such models are applicable to other urban catchments as well. The results are expected to be used to assist early warning systems and provide rational information for contingency and emergency planning.

1. Introduction

1.1. Background

Urban pluvial flooding is a threat to a great number of cities worldwide, especially given its increasing frequency of occurrence in recent years (Martina et al., 2006; Atta-ur-Rahman et al., 2016; Ziegler, 2012). Its impact, including loss of life and damages to both public and private

properties, can be further deepened by climate change and accelerated urbanization (Falconer et al., 2009). This type of flooding usually occurs when intense rainfall exceeds the capacity of an urban drainage system. Recent extreme precipitation events have raised awareness from both authorities and citizens to the challenges of predicting and managing urban pluvial floods. In July 2019, heavy rain caused at least 18 deaths and triggered massive transport disruptions in Mumbai, India (BBC, 2019). In the UK, about 40% of damages and associated

* Corresponding author.

E-mail addresses: q.ke@tudelft.nl (Q. Ke), x.tian@tudelft.nl (X. Tian), j.d.bricker@tudelft.nl (J. Bricker), tianz@sustech.edu.cn (Z. Tian), ggh@whu.edu.cn (G. Guan), caihy7@mail.sysu.edu.cn (H. Cai), xxhuang1031@outlook.com (X. Huang), yangl01@163.com (H. Yang), liujg@sustech.edu.cn, j.d.bricker@tudelft.nl (J. Liu).

<https://doi.org/10.1016/j.advwatres.2020.103719>

Received 30 November 2019; Received in revised form 18 July 2020; Accepted 6 August 2020

Available online 7 August 2020

0309-1708/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

economic losses in cities are estimated to result from pluvial flooding (Douglas et al., 2010). In China, 98% of cities are exposed or vulnerable to frequent floods (Jiang et al., 2017). A survey, conducted between 2008 and 2010, showed that 218 Chinese cities suffered severe urban pluvial flooding at least once, and more than 100 cities experienced it more than three times (Jiang et al., 2018). Therefore, urban pluvial flood prediction and management is a critical topic in the context of urban water management.

In order to prevent pluvial flooding and its consequences, city authorities (e.g. meteorological offices, emergency agency offices or water authorities) usually need to make predictions of pluvial floods. This is based on good prediction of precipitation characteristics, such as peak intensity, arrival time and duration. Many studies investigate the prediction of pluvial flooding by hydraulic models (Li, 2020; Li and Willems, 2019), by simulating the inundated area and depth given certain historical or predicted rainfall scenarios (Babaei et al., 2018; Thorndahl et al., 2016; Xing et al., 2019). However, hydraulic models need a large volume of data and computational resources. As a result, the output of a hydraulic model is usually case-specific. In other words, we have to run the model to make predictions for flooding during each separate rainfall scenario. On the other hand, using a rainfall threshold based on data-driven models can provide an easy and intuitive solution. By comparing the current/predicted rainfall with the threshold, one can straightforwardly estimate the likelihood of the city being flooded (Martina et al., 2006; Montesarchio et al., 2011; Tian et al., 2019; Yang et al., 2016). Specifically, a rainfall threshold specifies one or several rainfall depth(s) over certain time windows, above which a pluvial flood is likely to occur. Moreover, rainfall-threshold-based hazard prediction is widely applied to landslides (Garcia-urquia and Axelsson, 2015; Giannecchini et al., 2012; Golian et al., 2015; Hong et al., 2018; Martelloni et al., 2012), debris flow (Nikolopoulos et al., 2014; Pan et al., 2018; van Asch et al., 2014) and flash floods (Montesarchio et al., 2011; Norbiato et al., 2008; Zhai et al., 2018). To determine a cumulative rainfall threshold, a physically-based model is usually needed to compute critical rainfall thresholds over time (Norbiato et al., 2008; Yang et al., 2016), or a statistical, data-driven analysis can be applied (Carpenter et al., 1999; Golian et al., 2010; Martina et al., 2006; Montesarchio et al., 2011). However, on the one hand, there is a gap in short-term prediction capability of physical models (Costabile and Macchione, 2015). Short lead time flood prediction is of crucial importance for highly urbanized areas in order to provide timely warnings to residents (Zhang et al., 2018); on the other hand, statistical models have a limitation on the accuracy of prediction (Fawcett and Stone, 2010). Furthermore, urban catchments often lack sufficient data on both the drainage network and topography, complicating the estimate of rainfall threshold (Yang et al., 2016). Machine learning (ML) models can deal with data scarcity based on an ensemble method (Breiman, 2001). Therefore, in this paper, we use ML approaches to derive the flooding thresholds for different rainfall duration periods.

ML is a family of algorithms derived from statistics and computer science, which aims to train mathematical models to make predictions or decisions based on observed samples. ML is suggested as an effective tool to explore the connectedness between human and water systems (Shen et al., 2018). The latter is anticipated to be a key interdisciplinary issue to deal with in future hydrological studies (Vogel et al., 2015). Moreover, ML models can numerically reproduce flood nonlinearity, solely based on historical data, without requiring knowledge about the underlying physical processes (Mosavi et al., 2018). Therefore, this study utilizes ML algorithms to attempt to classify the presence or absence of flooding based on rainfall characteristics. Although ML algorithms have shown powerful applicability to flood prediction and forecasting (Liu et al., 2017; Mosavi et al., 2018; Noymanee et al., 2017; Tayfur et al., 2018), there are still very few studies that utilize ML to classify or predict urban pluvial flooding, which is a challenge due to lack of flood inundation data, drainage system data, and fine resolution topography data (Yang et al., 2016). Therefore, we aim to test

ML algorithms for classifying urban flooding in the city of Shenzhen, which is frequently flooded. A sudden rainstorm event claimed 11 lives in April 2019 in Shenzhen (Hua, 2019), attracting great attention for the local authorities to reconsider the early warning system for pluvial flooding in the city. Moreover, Shenzhen is a pioneer city in terms of high-technology development, socio-cultural development and disaster emergency management. This experience can be shared with other cities in China and abroad.

The paper is organized as follows. Section 2 describes the study area and data used for this study, and introduces the conventional and ML methods for flood prediction. Section 3 shows the results of the models and proposes the rainfall threshold for Shenzhen. Section 4 compares the ML results for rainfall thresholds to the current rainfall threshold and cumulative rainfall threshold in Shenzhen. Section 5 presents the conclusions and recommendations.

2. Materials and methodology

2.1. Study area

In the past decades, Shenzhen has grown rapidly from a rural area to a prosperous economic zone and an important industrial city in Southern China. It is located on the central coast of Guangdong Province, which is the passageway from mainland China to Hong Kong (See Fig. 1). It is also an important city in the Pearl River Delta (PRD). It has a total land area of 1,948 km². The average elevation is 3–4 m above mean sea level. Rainstorm-induced catastrophes in Shenzhen city are mostly caused by persistent short-duration heavy rainfall in the summer (Zhou et al., 2017). Pluvial flooding is one of the primary natural hazards in Shenzhen. In recent years, urbanization has increased the surface runoff and intensified the flood frequency (Shi et al., 2007; Yan et al., 2019).

Shenzhen is identified as an area under a high flood risk, since many properties are built in flood-prone areas, such as the harbour-front area (Chan et al., 2014). The total average annual precipitation is ~1,900 mm/y, of which rainstorms caused by typhoons (July – September) make up 36% (i.e., 689 mm/y) and approximately 85% of precipitation occurs from April to September (See Fig. 2) (Data source: Meteorological Bureau of Shenzhen (SMB)). Convective (March – June) and typhoon rainstorms (July – October) are the two main rainfall sources in this region.

As of 2019, Shenzhen has a population of 13 million, with a population density of 6,234 people/km². Most of the city is drained by a separated storm sewer system (4,883.92 km) whereas the remaining area (1,693 km) is drained by a combined sewer system (i.e. wastewater combined with rainwater sewer system), with a drainage pipe density of 12.5 km/km² (SSB, 2019). In total, 126 municipal pumps with a capacity of 671 m³/s are used to drain stormwater out of the city (SSB, 2019).

Short duration, high intensity rainfall is the main driver of pluvial flooding in Shenzhen. Due to the rapid pace of urbanization, the impervious area has significantly increased while the water storage area such as rivers, lakes and wetlands has decreased. With climate change (increasing frequency of typhoon occurrence and intensity of torrential rainfall) (Tracy et al., 2007), pluvial flooding has a high likelihood of occurrence in the paved area. On May 11, 2014, for instance, the daily rainfall volume reached 233 mm, and some districts experienced a peak rainfall intensity of 310 mm in 6 hours (Cai, 2014). Currently, SMB uses a rainfall threshold for predicting urban pluvial flooding, only based on 30-min rainfall depth (i.e., 20 mm) or 3-h rainfall depth (i.e. 80 mm) (SMB, 2019). In the subsequent sections, we will further testify and compare this threshold with that from the proposed ML models.

2.2. Records of flood events

Records of historical flood events from 1 June 2014 to 14 June 2017, consisting of 1,110 days and 663 records in total, were retrieved

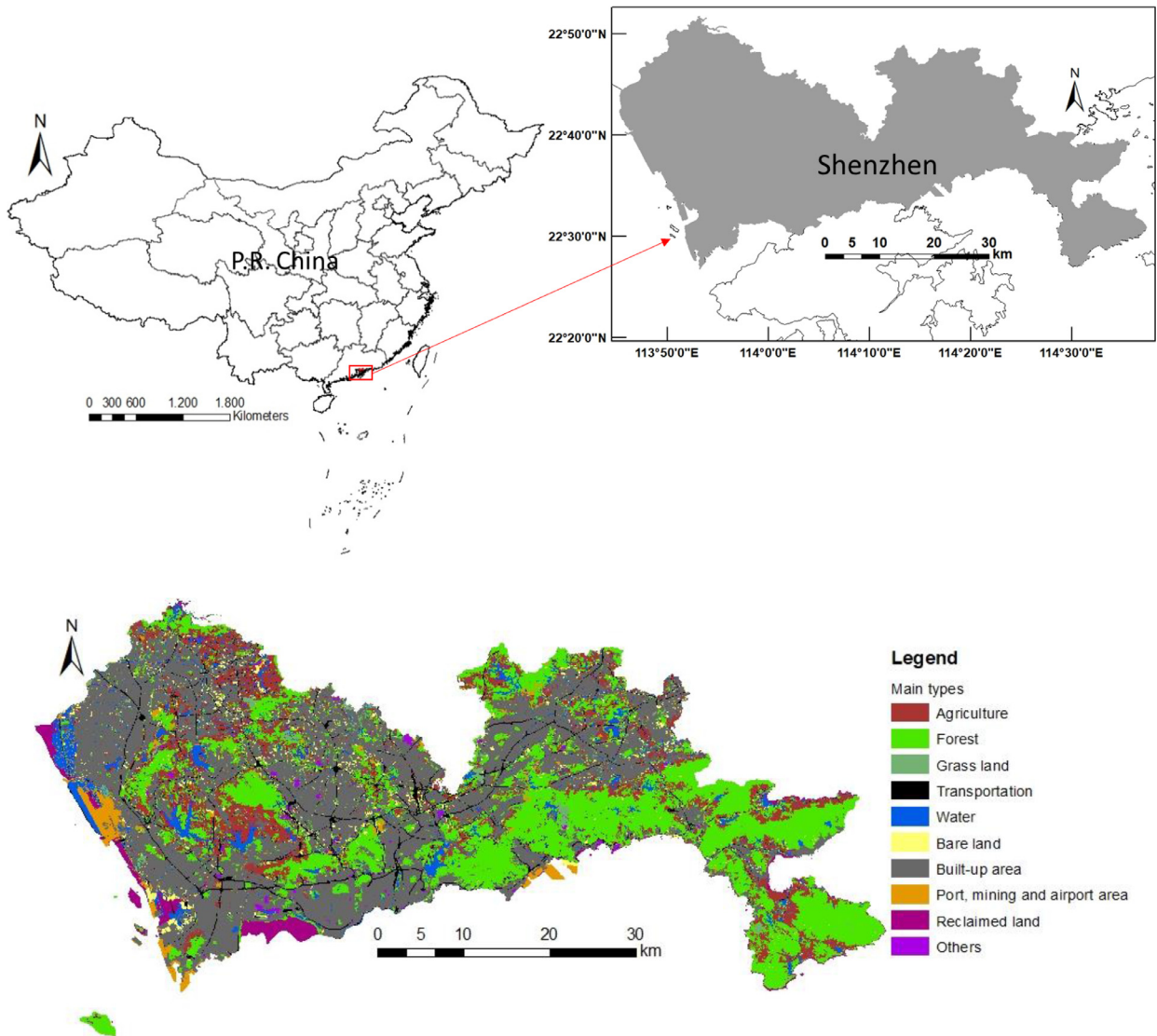


Fig. 1. Location and land use maps of the city of Shenzhen.

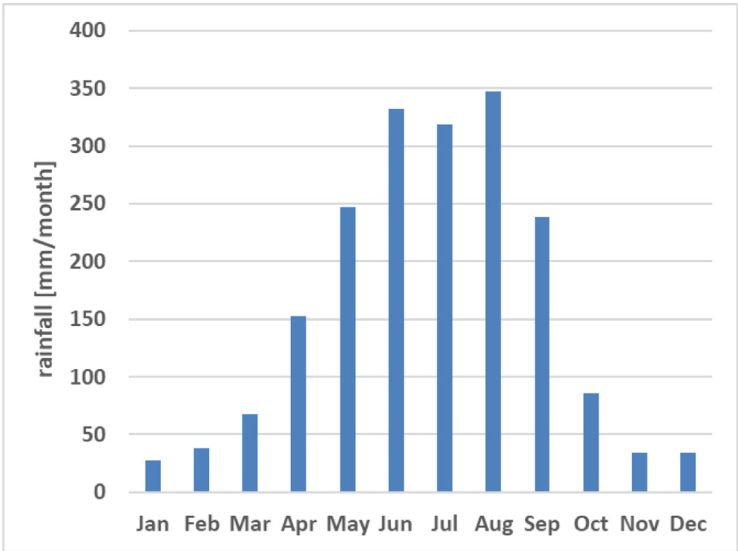


Fig. 2. Monthly average rainfall volumes for the city of Shenzhen, based on historical data from 1970 to 2015 (Data source: SMB).

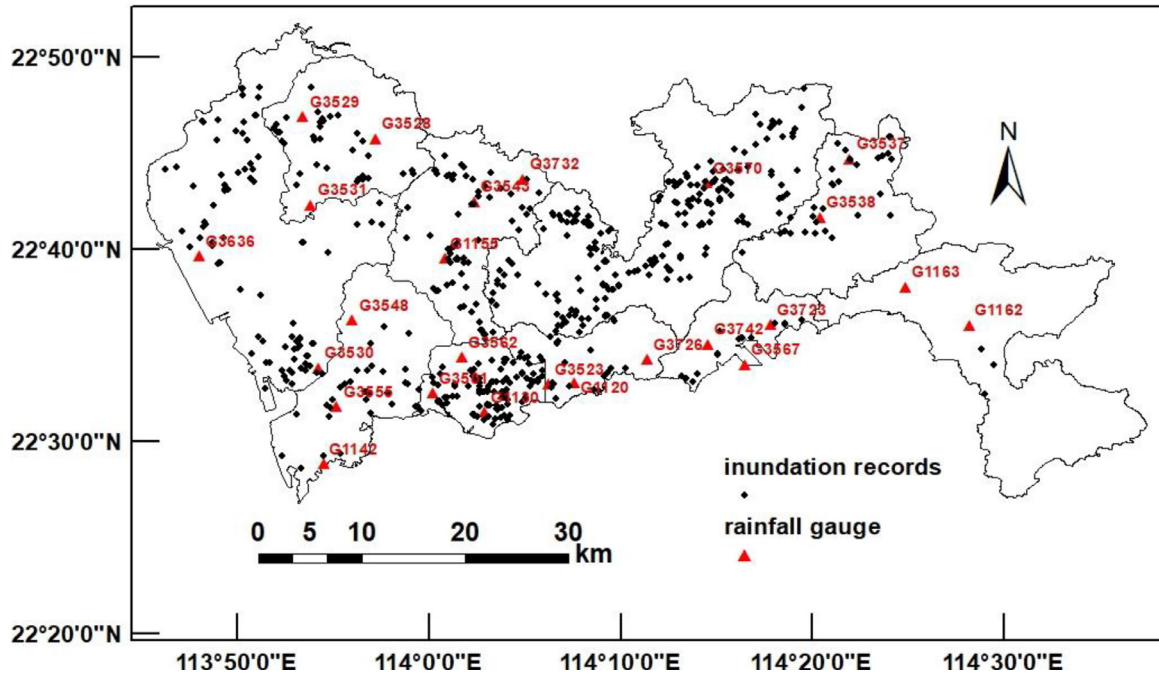


Fig. 3. Spatial distribution of 640 inundation records between 2014 – 2017 and 25 rainfall gauges in Shenzhen, China.

from the water sector of Shenzhen municipality (<http://swj.sz.gov.cn/>), which has developed and implemented a disaster reporting system (i.e. a flood report APP named 'shenzhensanfeng') since 2014. Citizens of Shenzhen can report flood events via this system at any time. These records register the date, the location (geotagging), and a description. As most of the records indicating pluvial flood events fall in the period between June and September (640 records, i.e., 96.5%), we only consider data points in the summer of each year, namely, 413 days in total over the 3-year study period. In doing so, we can exclude hundreds of non-flooding events to lower the imbalance of the dataset (too many non-flooding events and too few flooding events). Note that the high frequency of the flooding record corresponds to the precipitation characteristics in Shenzhen. The 640 records were registered over 24 days (c.a. 27 records/d), which are regarded as **days with floods**. The remaining 389 days of the study period are regarded as **days without floods**. These records are spatially distributed throughout the whole city (see Fig. 3). It should be noted that as the inundation records were submitted by citizens, socioeconomic background (such as age, education level and experience with previous pluvial flooding) may affect the recording. This may cause false alerts or missed alerts.

2.3. Rainfall observations

The rainfall intensity each minute at 25 rainfall gauges (see Fig. 3) from 1 June 2014 to 14 June 2017 was retrieved from SMB (<http://weather.sz.gov.cn/>). We used areal average rainfall intensity to represent the study area, which stands for the mean value of rainfall intensities of all study sub-areas (districts). The original database consisted of 1-min rainfall intensity. These 1-min rainfall intensities were aggregated to rainfall volumes of longer temporal scale, namely, 5, 10, 15, 30, 60, 120, 360, 720, and 1440 mins. Each day, the maximum rainfall volume at each temporal scale, denoted as Rd_x (in mm), is calculated by Eq. (1) (Tian et al., 2019).

$$Rd_x = \max_j \left\{ \sum_{k=1}^x R_1(k), \dots, \sum_{k=1+(j-1)*x}^{(j-1)*x} R_1(k), \dots, \sum_{k=1441-x}^{1440} R_1(k) \right\} \quad (1)$$

where $j = 0, 1, \dots, 1440/x$ [min]; $x = 1, 5, 10, 15, 30, 60, 120, 360, 720, 1440$ [min]. Note that each item in the bracket of Eq. (1) stands for

x-min rainfall volume accumulated from 1-min rainfall intensity in the interval $[1+j*x, (j+1)*x]$

2.4. Flood classification models

In this study, we first apply a conventional rainfall curve method as a benchmark. Then we further develop multiple parametric and non-parametric ML models to classify flooding and non-flooding events based on rainfall intensities. With respect to a binary classification problem, four possible predicted outcomes are expected (See Table 1), namely, true positives (TP or correctly classified flooding events), false positives (FP or falsely classified flooding events), true negatives (TN or correctly classified non-flooding events), and false negatives (FN or missed flooding events). Ideally, an urban flood classification model should achieve a high true positive rate (TPR), a high true negative rate (TNR) and high overall accuracy (ACC). On the other hand, a prediction model with a low positive predictive rate (PPR) or a low TPR implies that a number of actual flood events are wrongly labeled or unexpectedly missed. ACC is also called the proportion of correct forecasts (Wilks, 2005).

2.5. Conventional model with cumulative rainfall volume thresholds

The cumulative rainfall volume threshold is a reference curve, representing a cumulative amount of rainfall over a certain time window (see Fig. 4). When the observed cumulative rainfall exceeds the threshold at a given moment, flooding is expected to occur.

We propose a way to determine a threshold curve via the following steps:

- 1) Calculating the cumulative rainfall (max. in 24 hours), based on the 1-min rainfall intensity, for all flooding and non-flooding events.
- 2) Computing the lower α percentile of the 1-min rainfall for all flooding events, denoted as T_α . Note α is to be determined in step (5). In doing so, T_α depicts a curve that a certain number of cumulative rainfall curves for flooding events stay above. For instance, all curves of flooding events are above the curve $T_\alpha | \alpha=0$.
- 3) Computing the upper β percentile of the 1-min rainfall for all non-flooding events, denoted as T_β . Note β is also to be determined in

Table 1
Confusion matrix for quantifying the performance of a classification model.

	True classes			
		Flooding	Non-flooding	
Predicted classes	Flooding	True positives (TP)	False positives (FP)	Positive predictive rate (PPR) = TP/(TP+FP)
	Non-flooding	False negatives (FN) True positive rate (TPR) = TP/(TP+FN)	True negatives (TN) True negative rate (TNR) = TN/(FP+TN)	Negative predictive rate (NPR) = TN/(TN+FN) Accuracy (ACC) = (TN+TP)/(TP+TN+FN+FP)

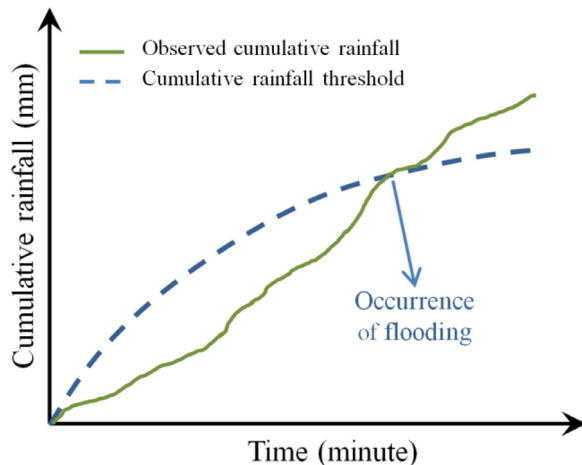


Fig. 4. An example of a cumulative rainfall threshold curve (when actual rainfall exceeds the rainfall threshold, pluvial flooding is expected to occur), modified from Martina et al. (2006).

step (5). The definition of T_β is analogous to that of T_α . T_β depicts a curve that a certain number of cumulative rainfall curves for flooding events stay below.

- Constituting a linear combination of T_α and T_β , based on a weight μ , namely, $T(\alpha, \beta, \mu) = \mu * T_\alpha + (1 - \mu) * T_\beta$. As a result, we obtain a rainfall threshold based on three variables: α , β , and μ . Any assigned values can result in a given cumulative rainfall threshold curve and its corresponding model performance.
- Solving an optimization problem that maximizes the model performance by tuning α , β , and μ . Three optimal combinations for α , β , and μ were pursued, aiming for the maximum TPR, the highest TNR, and the highest ACC:

$$\begin{aligned} \max_{\alpha, \beta, \mu} \quad & \text{Performance determined by } T(\alpha, \beta, \mu) \\ \text{Performance} = & \text{TPR, TNR, or Accuracy} \end{aligned} \quad (2)$$

2.5.1. Machine learning (ML) algorithms

Machine learning (ML) algorithms are a collection of computational data-driven methods. Without utilizing a pre-defined equation as the basic model, ML algorithms train a model, using a certain type of algorithms, fully based on known data whereas the trained model can be applied to new data. As the number of training data sets increases, the performance of ML algorithms can improve. ML consists of two families, namely, supervised learning and unsupervised learning.

Specifically, supervised learning algorithms aim to find functions that are able to map inputs to labeled outputs, also including two categories, classification and regression. Flooding prediction is commonly an application of classification (Jhong et al., 2018; Tayfur et al., 2018; Zhou et al., 2018), which aims to distinguish flood events vs. no-flood events based on hydrological variables, i.e., a binary classification problem.

Given the size of the database available, we adopt a collection of models in this study that usually show good performance for small- to

medium-sized data sets. 14 classification algorithms from 5 major ML families are considered to classify urban pluvial flooding based on rainfall intensities of multiple temporal scales (Table 2). Brief introductions of these algorithms follow:

- Decision trees:** Decision trees build a tree-shaped top-down structure from the root (at the top) to leaf nodes (at the bottom) (Breiman et al., 2017). Each leaf node represents a predicted response. Given the fact that we focus on a binary classification problem, the bifurcation starts from one parent node of a given layer to two child nodes of a subsequent lower layer, relying on different values of variables. Specifically, to find the optimal bifurcation, we maximize Gini's diversity index but stop maximization when (i) a node only contains a single-class of data, (ii) a child node to be generated contains fewer than five data points, or (iii) the number of layers exceeds a pre-defined criterion (five for a coarse decision tree and twenty for a medium decision tree). In general, decision tree learning is one of the fastest algorithms. Its results are also easy to interpret. We built the decision tree model in Matlab by using the function *fitctree*.
- Discriminant analysis:** discriminant analysis (DA) classifiers assumes a Gaussian distribution for data of each class. The Gaussian distribution is determined by the sample mean of each class and the identical covariance matrix for linear DA or different class-based covariance matrices for quadratic DA. Under this assumption, linear or quadratic DAs make predictions by minimizing prediction costs based on Bayes' theorem. Note that the prediction costs are the sum of the multiplication of the posterior probability of a given class k for a data sample and the cost of classifying a sample as y but its actual class is k (0 for accurate classification and 1 for misclassification). Readers can refer to (Ledoit and Wolf, 2004; T. Hastie, R. Tibshirani, 2008) for more details. Note that this study considers both linear and quadratic discriminant analyses. As their names suggest, linear discriminant analysis can only learn linear boundaries, while quadratic discriminant analysis can learn quadratic boundaries, both of which are fast to run and easy to interpret. We build the discriminant analysis model in Matlab by using the function *fitcdiscr*.
- Support vector machine:** linear support vector machine (SVM) applied to binary classification aims to find an optimal hyperplane that separates two classes with a margin of the maximal width. In other words, we look for the maximum margin width while keeping the data of two classes on each side of the margin. Samples that are misclassified are penalized. Using kernel functions, such as quadratic and cubic kernels, can turn a linear SVM into a non-linear SVM. The latter is more flexible but also requires more computational resources and becomes less straightforward to explain. Readers can refer to (Ng, 2000) for more details. We build the support vector machine model in Matlab using the function *fitcsvm*.
- K-nearest neighbor:** K-nearest neighbor is a distance-based learning technique that determines the predicted response of a given point by checking the major class of the k closest points (Cover and Hart, 1967). Note that we use the Euclidean and cosine distance as the metric to measure the closeness between points. The KNN algorithm is one of the easiest and most intuitive learning techniques

Table 2
14 machine learning models applied in this study.

No.	Algorithm categories	Models	Settings
1	Decision trees	Coarse tree	Maximum number of layers = 5
2		Medium tree	Maximum number of layers = 20
3	Discriminant analysis	Linear discriminant	Based on a linear classifier
4		Quadratic discriminant	Based on a quadratic classifier
5	Support vector machine (SVM)	Linear SVM	Based on a linear kernel
6		Quadratic SVM	Based on a quadratic kernel
7		Cubic SVM	Based on a cubic kernel
8	K nearest neighbor (KNN)	Fine KNN	Based on Euclidean distance, number of NN=1
9		Medium KNN	Based on Euclidean distance, number of NN=10
10		Cos KNN	Based on cosine distance, number of NN=100
11	Ensemble algorithms	Cubic KNN	Based on Euclidean distance, number of NN=10
12		Bagged trees	Number of learners =50 Learning rate = 0.1
13		Subspace discriminant	Number of learners =30 Subspace dimension = (number of features)/2
14		Subspace KNN	Number of learners =30 Subspace dimension = (number of features)/2

widely used in many applications (Cheng et al., 2014; Zhang, 2016). However, it is also very sensitive to outliers (Ramaswamy et al., 2000), which we may encounter frequently when predicting urban flooding based on rainfall intensities. We build the K-NN model in Matlab using the function *fitcknn*.

- Bagged trees (Breiman et al., 1984; Breiman, 2001): Bagging stands for a type of ensemble learning, which is used to reduce the variance of a single decision tree. To build a bagged tree model, we create multiple subsets of new data from original samples, which are chosen randomly with replacement. As a result, we obtain an ensemble of decision trees, also referred to as weak learners, and they are proven to be more robust than a single decision tree. We build the bagged trees model in Matlab using the function *fitcensemble*.
- Subspace ensembles: the random subspace method is also an ensemble technique to increase the accuracy of the discriminant classifier and KNN classifier. The subspace ensemble aims to train random sample features, rather than the entire feature set. It is proven to be an effective method to deal with the issue of high-dimensional feature sets and small training sets. As the name suggests, classifiers are constructed in a random subspace of data feature space and then combined by simple majority voting. Readers can refer to (Tin Kam Ho, 1998) for details. (García-Pedrajas and Ortiz-Boyer, 2009; Skurichina and Duin, 2002) also prove that the random subspace method can be further used for DA and KNN, which are applied in our study. We build the subspace-DA and subspace-KNN models in Matlab using the function *fitcensemble*.

2.5.2. Feature selection and model validation

All the models listed in Table 2 are first tested on ten features, which are the ten multi-temporal rainfall accumulations Rd_x (see Eq. (1)). Later, we also run a principal component analysis (PCA), based on the singular value decomposition method (Madsen et al., 2004), to reduce the number of dimensions and find the most meaningful components for predicting flooding events.

As we only have a small dataset with 413 data points, it is difficult to divide the whole dataset into several subsets for building, calibrating and validating the model. Instead, we use the 10-fold cross-validation technique (Bengio and Grandvalet, 2004) to deal with this issue. We randomly partition the dataset into 10 subsets of an equal size. Then we compute the mean value of the model performance for each subset. If the 10-fold cross-validation error is close to the error using the entire dataset, it means the model built from the entire dataset is unlikely to be over-fitted. In doing so, we are able to examine the performances of all

models. In the subsequent section, the accuracy of the model indicates the mean value of the accuracies of 10 models based on all data subsets.

3. Results

3.1. Conventional model

The conventional method is based on a linear combination of the lower percentile of the cumulative rainfall volumes of flooding events and the upper percentile of the cumulative rainfall volumes of non-flooding events. Fig. 5-(a) shows all the cumulative rainfall curves of wet days (daily rainfall depth ≥ 0.1 mm) from 1 June 2014 to 14 June 2017. More than 60% of flooding events occur with intensive rainfall of short temporal scale, e.g., 60 min to 360 min, but also with larger accumulation (blue dashed lines). More than 90% of non-flooding events have small rainfall volumes, for instance, daily accumulation being less than 20 mm. However, there are also exceptions, where events with large rainfall volumes were reported as non-flooding and vice versa.

We conducted an exhaustive search for all possible values of α , β , and μ between 0 and 1 and derived 112 Pareto optimal threshold curves, shown in Fig. 5-(b). Four representative rainfall threshold curves are selected, which have the highest rate for at least one of the five model quality metrics. Threshold 1 has the highest values in terms of TNR (0.98), PPR (0.73) and ACC (0.91), but also the lowest value of TPR (0.46); Threshold 2 has the highest ACC (0.91) but medium TPR (0.5); Threshold 3 has the highest NPR (0.98) and Threshold 4 has the highest TPR (0.96), which are presented in Fig. 5-(c) and Table 3. We can see that thresholds 1 and 2 ensure more non-flooding events are correctly classified, but also miss many flooding events. Thresholds 3 and 4 are more inclined to correctly classify flooding events, which implies that many non-flooding events can be labeled as flooding events based on these two thresholds. Threshold curves #1 and #2 are based on the lower 0-percentile ($\alpha = 0$) of the rainfall depth for all the flooding events and the upper 100-percentile ($\beta = 1$) of the rainfall depth for all the non-flooding events. These curves use a coefficient of 0 and 0.2 to make the linear combination. Both curves have a low TPR, meaning many actual flooding events are missed, and a high TNR, meaning non-flooding is well captured. ACC is thus relatively high, at 0.91. Curves #3 and #4 have the highest NPR and TPR, respectively, but very low ACC. This is because the threshold is low in Fig. 6-(b), ensuring flooding events are correctly classified, but missing non-flooding events. In general, it is difficult to find a threshold curve that can robustly indicate both flooding and non-flooding events, based on only the cumulative rainfall depth. Therefore,

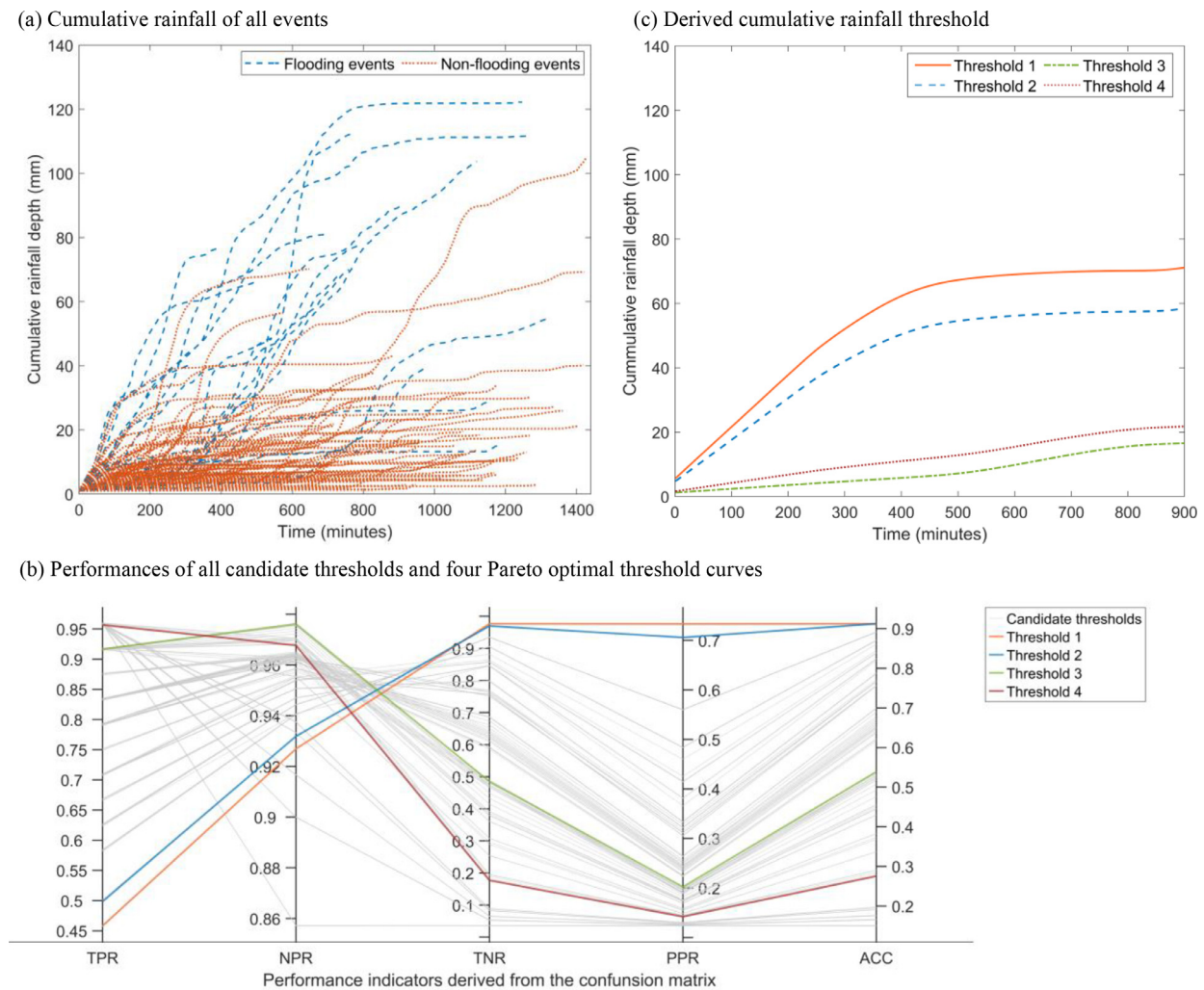


Fig. 5. (a) Cumulative rainfall volumes over time for all events on wet days; (b) performance indicators for all the candidate thresholds with the four selected ones highlighted; (c) four selected rainfall thresholds.

Table 3

Four cumulative rainfall threshold curves for the conventional method each target different performance metrics.

Threshold	A	β	μ	TPR	NPR	TNR	PPR	ACC
1	0	1	0	0.46	0.93	0.98	0.73	0.91
2	0	1	0.2	0.5	0.93	0.97	0.71	0.91
3	0.05	0.85	0.6	0.92	0.98	0.49	0.2	0.54
4	0.08	0.90	0.8	0.96	0.97	0.18	0.15	0.27
Range	[0, 0.2]	[0.8, 1]	[0, 1]	[0.46, 0.96]	[0.86, 0.98]	[0.04, 0.98]	[0.12, 0.73]	[0.15, 0.91]

we need other variables, rather than only the cumulative rainfall depth, to make a better classification.

3.2. Machine learning (ML)

3.2.1. Prediction results with 10 features

The first collection of ML models was trained based on ten rainfall volumes at 1, 5, 10, 15, 30, 60, 120, 360, 720, and 1440-min temporal resolutions, with definitions given in Eq. (1). All of the ML models have an ACC between 0.94 and 0.96 (Fig. 6), except for one model with an ACC of 0.92. This implies that only 16 to 25 events, out of 413 events, were misclassified in thirteen of the ML models used. This shows a slightly better performance than that of the conventional model. On the other hand, the TPR has a larger variation, ranging from 0.29 to 0.75. In other words, the miss rate ranges from 0.25 to 0.71. Among all the fourteen models, the DA family shows the most satisfactory perfor-

mance. Specifically, the Quadratic DA (Model 4) has the highest TPR (0.75), implying that 18 out of 24 actual flooding events can be well predicted while the ensemble DA (Model 13) has the highest ACC of 0.96 (See Fig. 6). All the performance metrics are listed in Appendix X1.

Although each ML model is easy to run with the complete set of all ten features, the result cannot be visualized in a ten-dimensional space, resulting in difficulty interpreting results. Therefore, we need to further reduce the number of dimensions to three or even fewer, as shown in the subsequent section.

3.2.2. Prediction results with 2 features

The second collection of ML models were trained by using two principal components, which were derived from ten rainfall accumulations by running a principal component analysis. The new features are linear combinations of the ten daily peak rainfall intensities at different temporal resolutions, with a set of coefficients given in Table 4. The first

Table 4
Coefficients of ten temporal resolutions of rainfall (Eq. (1)) for two principal components.

Coefficients	Rd_1	Rd_5	Rd_{10}	Rd_{15}	Rd_{30}	Rd_{60}	Rd_{120}	Rd_{360}	Rd_{720}	Rd_{1440}	explained
Feature 1	0.02	0.05	0.08	0.11	0.15	0.21	0.28	0.43	0.53	0.61	97.5%
Feature 2	0.05	0.16	0.25	0.31	0.40	0.44	0.42	0.15	-0.23	-0.46	2%

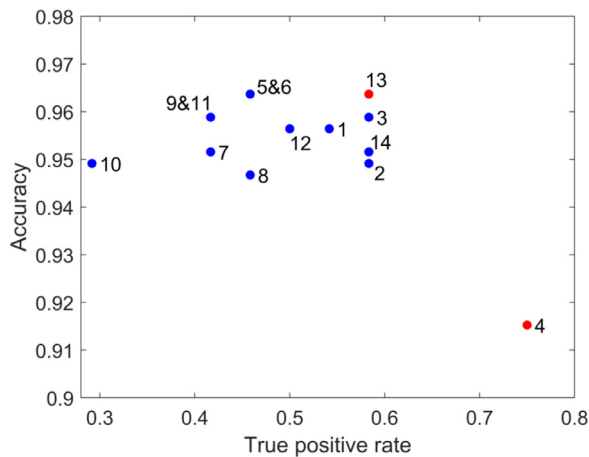


Fig. 6. True positive rates (TPR) and model accuracies (ACC) of 14 trained ML models based on rainfall accumulations at 10 temporal resolutions. Models 4 and 13 are marked in red as they have the best performance in terms of TPR and ACC, respectively. Model numbers correspond to Table 2.

feature is a weighted sum with larger temporal scales receiving more weight, explaining 97.5% of the total variance, while the second feature has more weight at time scales between 30 and 120 min, explaining 2% of the total variance. Therefore, using these features can explain 99.5% of the original dataset when classifying the labeled events. In principle, one can easily compute the values of the two features for present or future events based on the combination of rainfall accumulation volumes. If a study area has a coarser temporal resolution of rainfall measurement than that used here, principal component analysis can be run based on historical data of coarser resolution to generate two new sets of weights.

All models using two features have a performance that is slightly worse than that of the ten-feature models presented in Section 3.2.1. The ACC only drops by 0.01 to 0.02 for some models such as the medium decision trees (Model 2) and the fine/subspace KNN (Models 8 and 14), while other models (Model 3, 4, 5, 7 and 13) do not see reduced ACC. In terms of the TPR, fewer models reach 0.5 or higher, compared to the models in Section 3.2.1. However, as seen from Fig. 7, the subspace DA (Model 13) is still one of the best performing models. Two linear models, namely, the linear DA (Model 3) and the linear SVM (Model 5) also show a Pareto optimal performance in terms of ACC and TPR. It should be noted that Pareto optimality is a situation that cannot be modified so as to make any one individual or preference criterion better off without making at least one individual or preference criterion worse off. Models 3, 5, and 13 are adopted for further discussion because they have the best performance regarding either TPR or ACC. As shown in Fig. 7, these three models (in red) perform better than other models (in blue) for both performance indicators. The performance metrics of other models are listed in Appendix X2.

With two decision variables (i.e., features), we are able to visualize the outcome of the models in a two-dimensional plane. As shown in Fig. 8, Models 3, 5, and 13 determine rainfall threshold lines based on combinations of principle component feature 1 and feature 2. Flooding and non-flooding events occur to the right-hand and left-hand sides of each line, respectively. Among these three models, the threshold line from the linear DA model is furthest left, so classifies more events as flooding, while the linear SVM is the furthest right, so classifies fewer

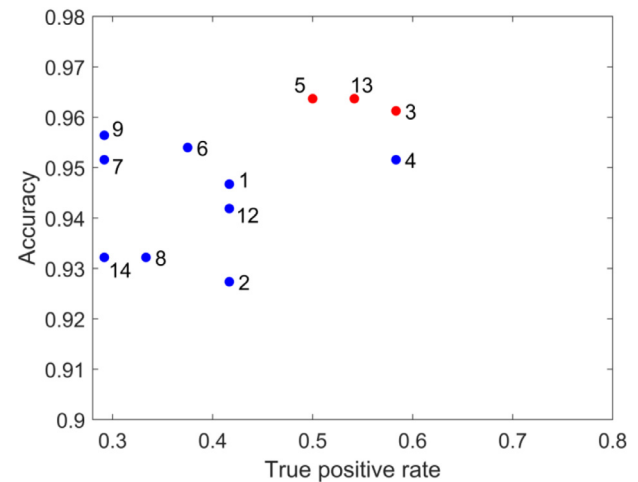


Fig. 7. True positive rates (TPR) and model accuracies (ACC) of 14 trained ML models based on two principal components. Models 3, 5, and 13 are marked in red as they have the best performance for either TPR or ACC. Model numbers correspond to Table 2.

events as flooding. The subspace DA provides a threshold line in between the other two. Note that Fig. 8 offers an intuitive look-up graph that one can easily tell whether an event is flooding or not based on the values of two features. For instance, a combination of feature 1 of 60 mm and feature 2 of 10 mm is predicted to not be a flooding event, but a combination of feature 1 of 100 mm and feature 2 of 10 mm is predicted to be a flooding event according to all models. However, further effort is still required to classify an event falling in the area between the lines of Model 3 and Model 5, as the three models may give different answers. We further elaborate on the fact that the nature of the data can lead to different thresholds from each of the three ML models in the Discussion section below.

4. Discussion

4.1. ML model compared to current rainfall threshold and cumulative rainfall threshold

We first elaborate on how the proposed ML model estimates the rainfall threshold better than the current empirical threshold provided by the local authority (SMB, 2019). The threshold suggests any event is regarded as a pluvial flood if either 30-min rainfall depth is over 20 mm or 3-h rainfall depth is over 80 mm. This threshold, and the historical data points, are shown in Fig. 9. As the 3-h rainfall threshold is placed too high, many flooding events are missed, resulting in a bad result for the TPR (only 0.25) although the overall ACC is good (0.95) as a large number of non-flooding events are correctly predicted. In other words, the miss rate for flooding events is very high, i.e., 0.75.

Even if the ML model is built based on a single feature, namely 30-min or 3-h rainfall depth, the ML model is still able to explore the dataset and find thresholds. We used one of the proposed ML models, specifically the subspace DA model, as one of the models with the best performance, to test the performance when using the same feature(s) of 30-min rainfall depth, 3-h rainfall depth, or their combination. The DA model suggests that the threshold should be either 30-min rainfall depth

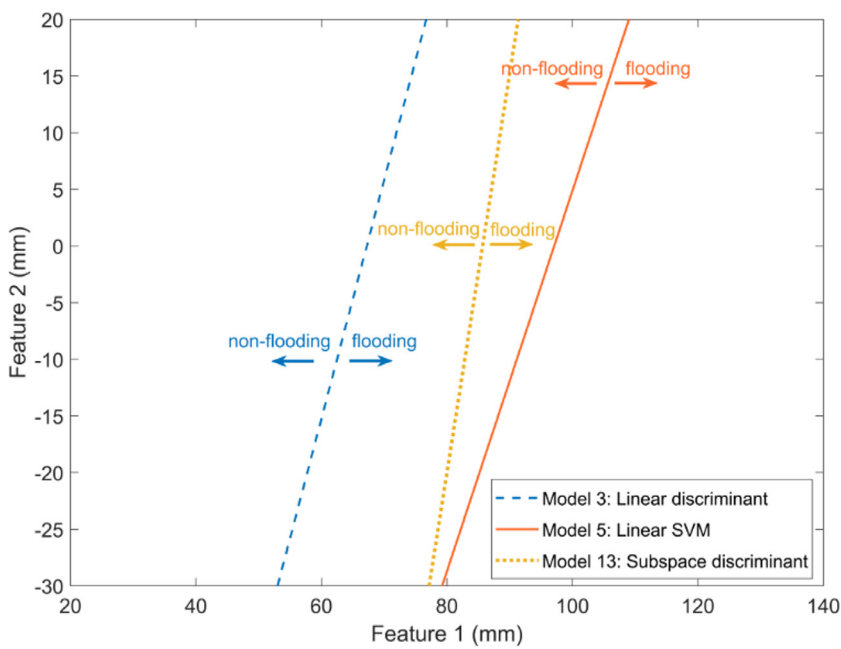


Fig. 8. Rainfall thresholds from three selected ML models which have the best performance with respect to the TPR or ACC. Note that a PCA was run in advance to derive two features which represent 99.5% of the original dataset.

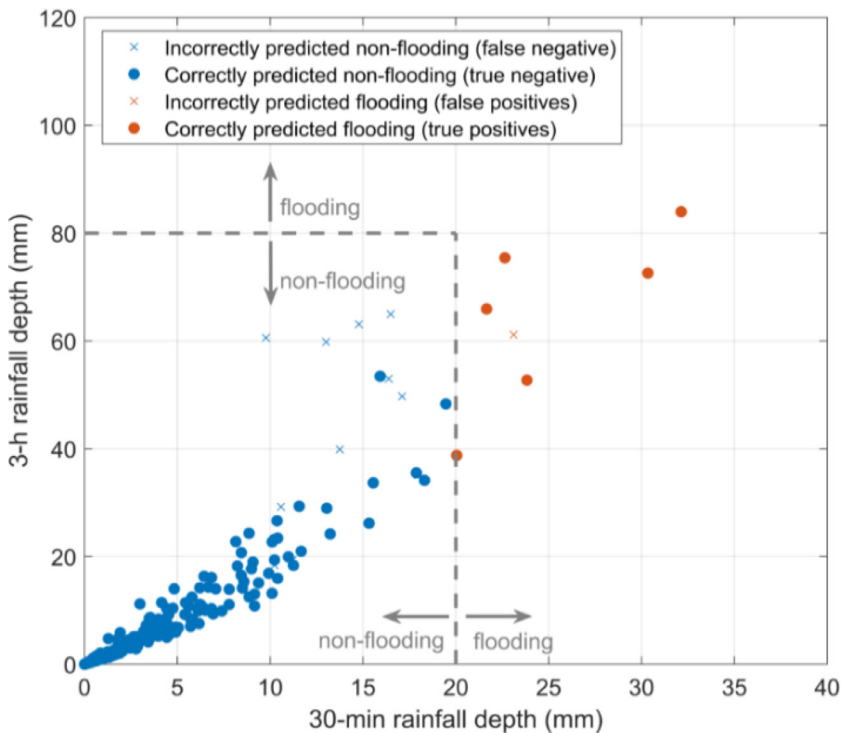


Fig. 9. The current empirical rainfall threshold (dashed lines) for urban pluvial flooding in Shenzhen, based on 30-min and 3-h rainfall depths, compared with historical data points.

of 12.5 mm (Fig. 10-(a)), 3-h rainfall depth of 29.1 mm (Fig. 10-(b)), or a combination of these (Fig. 10-(c)). Performance TPR's are all higher than 0.54, which is more than twice the TPR using the empirical rainfall threshold. Detailed metrics are shown in Appendix X3. This means that the machine learning models can improve the current empirical rainfall threshold to a great extent.

Next, we compare the performance of conventional cumulative critical rainfall curves to those derived from the ML models. The results show that ML models, especially linear discriminant analysis, can classify flooding and non-flooding by two principle components, raising the ACC and TPR to 96% and 58%, respectively; and lowering the false alert

rate to 25%. Compared to the conventional model, the critical indices of ACC and TPR were 5%-15% higher in ML models. Therefore, in general, ML models can better classify flooding and non-flooding events than the conventional empirical method, based on different temporal resolutions of rainfall measurements.

The minimum temporal resolution for the input of our ML models is 1 minute. However, the method is generic. The minimum temporal resolution can also be 5-min or 10-min to re-train the model. To train the ML model, the user needs reports or observations of flooding and non-flooding events. These inputs (rainfall and flood reports) are identical to the inputs needed by the conventional method.

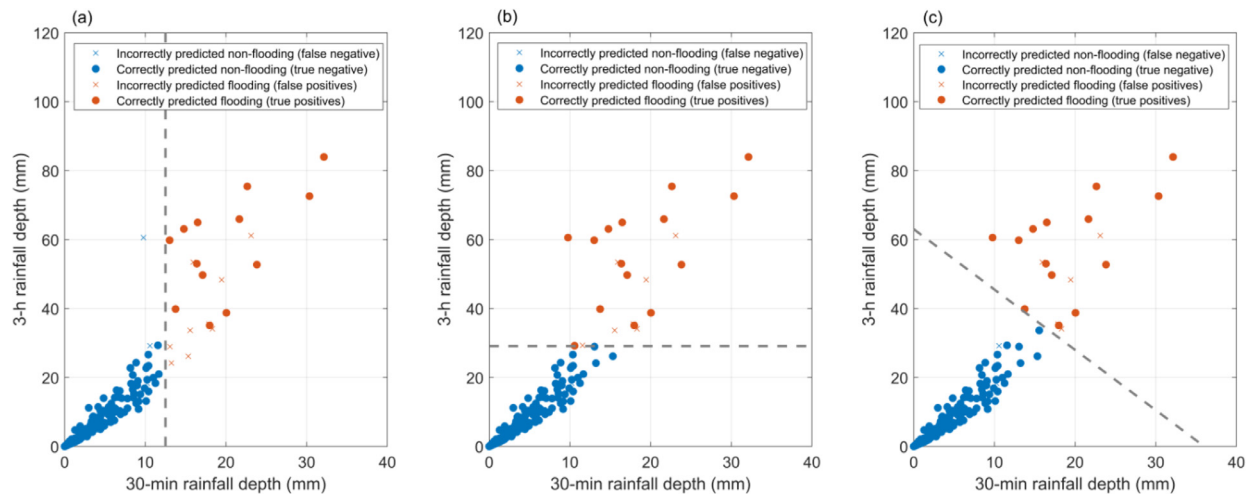


Fig. 10. Performance of single and dual feature rainfall thresholds calculated by the subspace DA model based on observed records. (a) performance of single threshold – 30-mins of 12.5 mm; (b) performance of single threshold – 3-hours of 29.1mm; (c) performance of dual threshold (30-mins together with 3-hours).

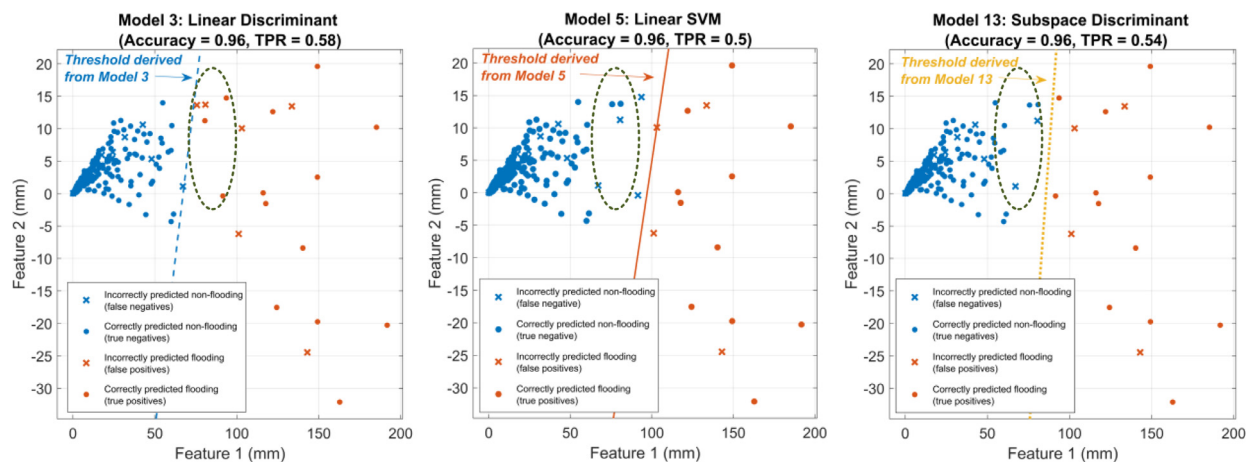


Fig. 11. Performance of outstanding ML models (models 3, 5 and 13) in terms of ACC and TPR using two principle components. Each model categorizes events to the left of its threshold line as non-flooding, and to the right of the line as flooding.

4.2. Pros and cons of the machine learning (ML) model

ML models can successfully produce rainfall thresholds for urban pluvial flooding. The model only needs to be run once and the water system manager/operator can simply use a look-up graph to determine whether a pluvial flood is likely to occur. The features can be flexibly selected, using either the entire 10 features, or fewer representative features by running a PCA.

However, ML is a data-derived method, which largely relies on the quantity and the quality of data available. For example, five points (which are circled in Fig. 11), regarded as ‘tricky events’, can influence the output when using different models. These events have similar rainfall conditions but they are categorized by the ML models differently; in reality, three are flooding events and two are non-flooding events. The models only make decisions based on data, resulting indifferent threshold lines for Models 3, 5 and 13. Model 3 includes these five points in the set of flooding events, thereby making two predictions incorrectly. Model 5 excludes these five points from the set flooding events, thereby making three predictions incorrectly. Model 13 draws a threshold line in between, thereby making only one prediction incorrectly. Potentially, more data points lying in between the threshold line of Model 3 and the

threshold line of Model 5 can improve the model to make predictions more precisely.

In this work, historical inundation records were collected through a flood report system (a smart phone application). However, not all the municipality’s citizens are aware of this reporting system. This limits the number of the records, thus affecting the TPR (i.e. increased missed alerts). In addition, each citizen’s socio-economic background, education level and experience with pluvial flooding influence the records as well. For instance, inundation caused by blockage of sewers/pipes at home can be wrongly reported as inundation caused by rainfall; this undoubtedly increased the number of false positives. Since the current flood report system does not provide information on the reasons for inundation, false inundation records cannot be filtered out.

It should also be noted that our ML models were applied over the entire city of Shenzhen in this study, due to the limited number of data points. If more data become available, the model can be further refined to a district, a community, or a street. Similarly, it can also be applied to other urban/rural catchments given an available rainfall-flooding database. As more available data can be collected in the future, even with images and text descriptions, we also aim to test deep learning algorithms to increase the accuracy of the flood prediction model.

5. Conclusion

Despite uncertainty about the inundation records and ML models, this data-driven method provides a basis for generating rainfall thresholds for flood early warning and emergency response in Shenzhen. The objective of this paper is to predict the occurrence of urban pluvial flooding by ML approaches. It concludes that ML models can determine the rainfall flooding threshold as a line projected in a plane spanned by two principal components, thereby providing a binary result (flood or no flood). Compared to the conventional empirical critical rainfall curve, the proposed models, especially the subspace discriminant analysis algorithm, can better classify flooding and non-flooding events by different combinations of multi-resolution rainfall intensities, greatly raising the ACC to 96.5% and lowering the false alert rate to 25%. Such models are applicable to other urban catchments as well.

Extreme weather events in the future due to global climate change will bring high-intensity rainfall of short duration (Westra et al., 2014). Advanced techniques, such as radar observations, can efficiently improve very short-range rainfall forecasts, which are essential for accurate flood prediction (Yang et al., 2016). Precipitation is the dominant input influencing the flood prediction result. Other factors like soil characteristics, drainage capacity and topography (e.g. land subsidence) would affect the result as well, emphasizing the need for updated, data-driven flooding thresholds. Since rainfall-threshold-based flood prediction can be executed rapidly and simply, this method allows decision makers (e.g. emergency managers) time for a high-level assessment of flood risk, providing valuable lead time for citizens in the flood-prone areas to be warned. Probability thresholds, which can help understand the uncertainties involved, need to be investigated further. Although the inundation records contain information about occurrence locations and (estimated) inundation depths, these data were not utilized/analysed in this study. Further study on the correlation of spatial distribution of inundation and inundation depth with the spatially varying rainfall records will be valuable as well.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Qian Ke: Conceptualization, Methodology, Data curation, Writing - original draft. **Xin Tian:** Conceptualization, Methodology, Data curation, Writing - original draft. **Jeremy Bricker:** Supervision, Writing - review & editing. **Zhan Tian:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Guanghua Guan:** Visualization, Investigation. **Huayang Cai:** Visualization, Investigation. **Xinxing Huang:** Visualization, Investigation. **Honglong Yang:** Data curation. **Junguo Liu:** Writing - review & editing.

Acknowledgements

The authors would like to thank the sponsors of this work: the National Key R&D Program of China (Grant no. 2018YFE0206200), the National Natural Science Foundation of China (Grant no. 41671113 and 51761135024), the Netherlands Organisation for Scientific Research (NWO) (Grant no. ALWSD.2016.007), JPI Urban Europe, ERA-NET Co-fund Smart Urban Futures (Project no. 646453), NWO/VerDuS Smart Urban Regions of the Future (SURF) and the Engineering and Physical Sciences Research Council of UK (Grant no. R034214/1), the High-level Special Funding of the Southern University of Science and Technology (Grant no. G02296302, G02296402). We sincerely appreciate three reviewer's comments and suggestions which improved the quality of the

article. We also would like to show our gratitude to Meteorological Bureau of Shenzhen Municipality and Water Sector of Shenzhen Municipality for providing the rainfall data and inundation records in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.advwatres.2020.103719](https://doi.org/10.1016/j.advwatres.2020.103719).

Appendices

X1. Performance metrics of 14 trained machine learning models based on 10 features.

No	Models	TP	FN	FP	TN	TPR	TNR	PPR	NPR	Accuracy
1	Coarse tree	13	11	7	382	0.54	0.98	0.65	0.97	0.956
2	Medium tree	14	10	11	378	0.58	0.97	0.56	0.97	0.949
3	Linear discriminant	14	10	7	382	0.58	0.98	0.67	0.97	0.959
4	Quadratic discriminant	18	6	29	360	0.75	0.93	0.38	0.98	0.919
5	Linear SVM	11	13	2	387	0.46	0.99	0.85	0.97	0.964
6	Quadratic SVM	11	13	2	387	0.46	0.99	0.85	0.97	0.964
7	Cubic SVM	10	14	6	383	0.42	0.98	0.63	0.97	0.952
8	Fine KNN	11	13	9	380	0.46	0.98	0.55	0.97	0.946
9	Medium KNN	10	14	3	386	0.42	0.99	0.77	0.97	0.959
10	Cos KNN	7	17	4	385	0.29	0.99	0.64	0.96	0.950
11	Cubic KNN	10	14	3	386	0.42	0.99	0.77	0.97	0.959
12	Bagged trees	12	12	6	383	0.50	0.98	0.67	0.97	0.956
13	Subspace discriminant	14	10	5	384	0.58	0.99	0.74	0.97	0.964
14	Subspace KNN	14	10	10	379	0.58	0.97	0.58	0.97	0.951

X2. Performance metrics of 14 trained machine learning models based on 2 features derived from a PCA.

No	Models	TP	FN	FP	TN	TPR	TNR	PPR	NPR	Accuracy
1	Coarse tree	10	14	8	381	0.42	0.98	0.56	0.96	0.947
2	Medium tree	10	14	16	373	0.42	0.96	0.38	0.96	0.927
3	Linear discriminant	14	10	6	383	0.58	0.98	0.70	0.97	0.961
4	Quadratic discriminant	14	10	10	379	0.58	0.97	0.58	0.97	0.952
5	Linear SVM	12	12	3	386	0.50	0.99	0.80	0.97	0.964
6	Quadratic SVM	9	15	4	385	0.38	0.99	0.69	0.96	0.954
7	Cubic SVM	7	17	3	386	0.29	0.99	0.70	0.96	0.952
8	Fine KNN	8	16	12	377	0.33	0.97	0.40	0.96	0.932
9	Medium KNN	7	17	1	388	0.29	1.00	0.88	0.96	0.956
10	Cos KNN	0	24	1	388	0.00	1.00	0.00	0.94	0.939
11	Cubic KNN	5	19	1	388	0.21	1.00	0.83	0.95	0.952
12	Bagged trees	10	14	10	379	0.42	0.97	0.50	0.96	0.942
13	Subspace discriminant	13	11	4	385	0.54	0.99	0.76	0.97	0.964
14	Subspace KNN	7	17	11	378	0.29	0.97	0.39	0.96	0.932

X3. Performance metrics of 14 trained machine learning models based on (a) 30-min rainfall (b) 3-h rainfall (c) both 30-min and 3-h rainfall.

No	Models	TP	FN	FP	TN	TPR	TNR	PPR	NPR	Accuracy
a	Based on 30-min rainfall	13	11	9	380	0.54	0.98	0.59	0.97	0.952
b	Based on 3-h rainfall	15	9	7	382	0.63	0.98	0.68	0.98	0.961
c	Based on 30-min and 3-h rainfall	14	10	5	384	0.58	0.99	0.74	0.97	0.964

References

- Atta-ur-Rahman, Parvin, G.A., Shaw, R., Surjan, A., 2016. Cities, Vulnerability, and Climate Change, in: Urban Disasters and Resilience in Asia. Elsevier, pp. 35–47. <https://doi.org/10.1016/B978-0-12-802169-9.00003-3>.
- Babaei, S., Ghazavi, R., Erfanian, M., 2018. Urban flood simulation and prioritization of critical urban sub-catchments using SWMM model and PROMETHEE II approach. Phys. Chem. Earth 105, 3–11. <https://doi.org/10.1016/j.pce.2018.02.002>.
- BBC, 2019. Mumbai: Heaviest rain in decade triggers chaos [WWW Document]. URL <https://www.bbc.com/news/world-asia-india-48835828>.

- Bengio, Y., Grandvalet, Y., 2004. No unbiased estimator of the variance of K-Fold cross-validation. *J. Mach. Learn. Res.* 5, 1089–1105. [https://doi.org/10.1016/S0006-291X\(03\)00224-9](https://doi.org/10.1016/S0006-291X(03)00224-9).
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and Regression Trees*, 1st edn..
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. *Classification And Regression Trees*. Routledge. <https://doi.org/10.1201/9781315139470>.
- Cai, Y., 2014. Causes analysis of urban flooding in Shenzhen. Shenzhen Water Conservancy Planning Design Institute, in Chinese Technical report.
- Carpenter, T.M., Sperflage, J.A., Georgakakos, K.P., Sweeney, T., Fread, D.L., 1999. National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems. *J. Hydrol.* 224, 21–44. [https://doi.org/10.1016/S0022-1694\(99\)00115-8](https://doi.org/10.1016/S0022-1694(99)00115-8).
- Chan, F.K.S., Wright, N., Cheng, X., Griffiths, J., 2014. After Sandy: Rethinking Flood Risk Management in Asian Coastal Megacities. *Nat. Hazards Rev.* 15, 101–103. [https://doi.org/10.1061/\(asce\)nh.1527-6996.0000117](https://doi.org/10.1061/(asce)nh.1527-6996.0000117).
- Costabile, P., Macchione, F., 2015. Enhancing river model set-up for 2-D dynamic flood modelling. *Environ. Model. Softw.* 67, 89–107. <https://doi.org/10.1016/j.envsoft.2015.01.009>.
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27. <https://doi.org/10.1109/TIT.1967.1053964>.
- Cheng, D., Zhang, S., Deng, Z., Zhu, Y., Zong, M., 2014. kNN Algorithm with Data-Driven k Value, pp. 499–512. https://doi.org/10.1007/978-3-319-14717-8_39.
- Douglas, I., Garvin, S., Lawson, N., Richards, J., Tippet, J., White, I., 2010. Urban pluvial flooding: A qualitative case study of cause, effect and nonstructural mitigation. *J. Flood Risk Manag.* 3, 112–125. <https://doi.org/10.1111/j.1753-318X.2010.01061.x>.
- Falconer, R.H., Cobby, D., Smyth, P., Astle, G., Dent, J., Golding, B., 2009. Pluvial flooding: New approaches in flood warning, mapping and risk management. *J. Flood Risk Manag.* 2, 198–208. <https://doi.org/10.1111/j.1753-318X.2009.01034.x>.
- Fawcett, R., Stone, R., 2010. A comparison of two seasonal rainfall forecasting systems for Australia. *Aust. Meteorol. Oceanogr. J.* 61, 15–23. <https://doi.org/10.22499/2.6001.002>.
- García-Pedrajas, N., Ortiz-Boyer, D., 2009. Boosting k-nearest neighbor classifier by means of input space projection. *Expert Syst. Appl.* 36, 10570–10582. <https://doi.org/10.1016/j.eswa.2009.02.065>.
- García-urquía, E., Axelsson, K., 2015. Rainfall thresholds for the occurrence of urban landslides in tegucigalpa, honduras: an application of the critical rainfall intensity. *Geogr. Ann. Ser. A, Phys. Geogr.* 97, 61–83. <https://doi.org/10.1111/geoa.12092>.
- Gianacchini, R., Galanti, Y., D'Amato Avanzi, G., 2012. Critical rainfall thresholds for triggering shallow landslides in the Serchio River Valley (Tuscany, Italy). *Nat. Hazards Earth Syst. Sci.* 12, 829–842. <https://doi.org/10.5194/nhess-12-829-2012>.
- Golian, S., Fallahi, M.R., Behbahani, S.M., Sharifi, S., Sharma, A., 2015. Real-time updating of rainfall threshold curves for flood forecasting. *J. Hydrol. Eng.* 20, 1–9. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001049](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001049).
- Golian, S., Saghaian, B., Maknoon, R., 2010. Derivation of Probabilistic Thresholds of Spatially Distributed Rainfall for Flood Forecasting. *Water Resour. Manag.* 24, 3547–3559. <https://doi.org/10.1007/s11269-010-9619-7>.
- Hong, M., Kim, J., Jeong, S., 2018. Rainfall intensity-duration thresholds for landslide prediction in South Korea by considering the effects of antecedent rainfall. *Landslides* 15, 523–534. <https://doi.org/10.1007/s10346-017-0892-x>.
- Hua, C., 2019. Sudden rainstorm kills 11 people in Shenzhen. <https://www.chinadailyhk.com/articles/154/181/191/1555144049585.html>. (Accessed 09 May 2019).
- Jhong, Y.-D., Chen, C.-S., Lin, H.-P., Chen, S.-T., 2018. Physical Hybrid Neural Network Model to Forecast Typhoon Floods. *Water* 10, 632. <https://doi.org/10.3390/w10050632>.
- Jiang, Y., Zevenbergen, C., Fu, D., 2017. Can “Sponge Cities” Mitigate China's Increased Occurrences of Urban Flooding? *Aquademia Water. Environ. Technol.* 1, 1–5. <https://doi.org/10.20897/awet.201703>.
- Jiang, Y., Zevenbergen, C., Ma, Y., 2018. Urban pluvial flooding and stormwater management: A contemporary review of China's challenges and “sponge cities” strategy. *Environ. Sci. Policy* 80, 132–143. <https://doi.org/10.1016/j.envsci.2017.11.016>.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.3390/rs10060911>.
- Ledoit, O., Wolf, M., 2004. Honey, I Shrunk the Sample Covariance Matrix: Problems in Mean-Variance Optimization. *J. Portf. Manag.* 30, 110–119.
- Li, X., 2020. A Hybrid Model for Fast and Probabilistic Urban Pluvial Flood Prediction. *Water Resour. Res.* 56. <https://doi.org/10.1029/2019WR025128>.
- Li, X., Willems, P., 2019. Probabilistic flood prediction for urban sub-catchments using sewer models combined with logistic regression models. *Urban Water J.* 16, 687–697. <https://doi.org/10.1080/1573062X.2020.1726409>.
- Liu, F., Xu, F., Yang, S., 2017. A Flood Forecasting Model Based on Deep Learning Algorithm via Integrating Stacked Autoencoders with BP Neural Network. In: 2017 IEEE Third International Conference on Multimedia Big Data (BigMM), pp. 58–61. <https://doi.org/10.1109/BigMM.2017.29> IEEE.
- Madsen, R.E., Kai, H., Winther, O., 2004. Singular Value Decomposition and Principal Component Analysis.
- Martelloni, G., Segoni, S., Fanti, R., Catani, F., 2012. Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides* 485–495. <https://doi.org/10.1007/s10346-011-0308-2>.
- Martina, M.L.V., Todini, E., Libralon, A., 2006. A Bayesian decision approach to rainfall thresholds based flood warning. *Hydrol. Earth Syst. Sci.* 10, 413–426. <https://doi.org/10.5194/hess-10-413-2006>.
- Montesarchio, V., Ridolfi, E., Russo, F., Napolitano, F., 2011. Rainfall threshold definition using an entropy decision approach and radar data. *Nat. Hazards Earth Syst. Sci.* 11, 2061–2074. <https://doi.org/10.5194/nhess-11-2061-2011>.
- Ng, A., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. In: Christianini, N., J.C.S.-T. (Eds.), *Intelligent Systems and Their Applications* IEEE. Cambridge University Press, pp. 1–25. <https://doi.org/10.1016/j.jaca.2011.07.027>.
- Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: Literature review. *Water (Switzerland)* 10, 1–40. <https://doi.org/10.3390/w10111536>.
- Nikolopoulos, E.I., Crema, S., Marchi, L., Marra, F., Guzzetti, F., Borga, M., 2014. Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. *Geomorphology* 221, 286–297. <https://doi.org/10.1016/j.geomorph.2014.06.015>.
- Norbiato, D., Borga, M., DegliEsposti, S., Gaume, E., Anquetin, S., 2008. Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. *J. Hydrol.* 362, 274–290. <https://doi.org/10.1016/j.jhydrol.2008.08.023>.
- Noymanee, J., Nikitin, N.O., Kalyuzhnaya, A.V., 2017. Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin. *Procedia Comput. Sci.* 119, 288–297. <https://doi.org/10.1016/j.procs.2017.11.187>.
- Pan, H.L., Jiang, Y.J., Wang, J., Ou, G.Q., 2018. Rainfall threshold calculation for debris flow early warning in areas with scarcity of data. *Nat. Hazards Earth Syst. Sci.* 18, 1395–1409. <https://doi.org/10.5194/nhess-18-1395-2018>.
- Ramaswamy, S., Rastogi, R., Shim, K., 2000. Efficient algorithms for mining outliers from large data sets. In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data - SIGMOD '00*. ACM Press, New York, New York, USA, pp. 427–438. <https://doi.org/10.1145/342009.335437>.
- Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.J., Ganguly, S., Hsu, K.L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., Tsai, W.P., 2018. HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrol. Earth Syst. Sci.* 22, 5639–5656. <https://doi.org/10.5194/hess-22-5639-2018>.
- Shi, P.J., Yuan, Y., Zheng, J., Wang, J.A., Ge, Y., Qiu, G.Y., 2007. The effect of land use/cover change on surface runoff in Shenzhen region, China. *Catena* 69, 31–35. <https://doi.org/10.1016/j.catena.2006.04.015>.
- Skurichina, M., Duin, R.P.W., 2002. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal. Appl.* 5, 121–135. <https://doi.org/10.1007/s100440200011>.
- SMB, 2019. Rough determination of meteorological disaster in Shenzhen [www Document]. URL http://weather.sz.gov.cn/qixiangfuwu/fangzaijianzai/fangzaijianzai/content/post_3576505.html in Chinese.
- SSB (Statistical Bureau of Shenzhen Municipality), 2019. Shenzhen Statistical Yearbook. China Statistics Press online PDF <http://tj.sz.gov.cn/attachment/0/418/418268/6765070.pdf>.
- Hastie, R. T., Tibshirani, J.F., 2008. The Elements of Statistical Learning -Data Mining, Inference, and Prediction. pp.106–119.
- Tayfur, G., Singh, V., Moramarco, T., Barbetta, S., 2018. Flood Hydrograph Prediction Using Machine Learning Methods. *Water* 10, 968. <https://doi.org/10.3390/w10080968>.
- Thorndahl, S., Nielsen, J.E., Jensen, D.G., 2016. Urban pluvial flood prediction: A case study evaluating radar rainfall nowcasts and numerical weather prediction models as model inputs. *Water Sci. Technol.* 74, 2599–2610. <https://doi.org/10.2166/wst.2016.474>.
- Tian, X., Veldhuis, M., Schleiss, M., 2019. Critical rainfall thresholds for urban pluvial flooding inferred from citizen observatories. *Sci. Total Environ.* 21, 4890.
- Ho, Tin Kam, 1998. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 832–844. <https://doi.org/10.1109/34.709601>.
- Tracy, A., Trumbull, K., Loh, C., 2007. Impact of climate change in Hong Kong and the pearl river delta. *China Perspect.* 2007, 18–29.
- van Asch, T.W.J., Tang, C., Alkema, D., Zhu, J., Zhou, W., 2014. An integrated model to assess critical rainfall thresholds for run-out distances of debris flows. *Nat. Hazards* 70, 299–311. <https://doi.org/10.1007/s11069-013-0810-z>.
- Vogel, R.M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P.K., Hooper, R.P., Matalas, N.C., 2015. Hydrology: The interdisciplinary science of water. *Water Resour. Res.* 51, 4409–4430. <https://doi.org/10.1002/2015WR017049>.
- Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F., Kendon, E.J., Lenderink, G., Roberts, N.M., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. *Rev. Geophys.* 52 (3), 522–555. <https://doi.org/10.1002/2014RG000464>.
- Wilks, D., 2005. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Elsevier, Burlington.
- Xing, Y., Liang, Q., Wang, G., Ming, X., Xia, X., 2019. City-scale hydrodynamic modelling of urban flash floods: the issues of scale and resolution. *Nat. Hazards* 96, 473–496. <https://doi.org/10.1007/s11069-018-3553-z>.
- Yan, H., He, X., Lei, Y., Wang, Y., Su, H., Jiang, S., 2019. Land use-induced change in trophic state of Shenzhen Bay (South China) over the past half-century. *Mar. Pollut. Bull.* 145, 208–213. <https://doi.org/10.1016/j.marpolbul.2019.05.046>.
- Yang, T.H., Hwang, G.Do, Tsai, C.C., Ho, J.Y., 2016. Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts. *Hydrol. Earth Syst. Sci.* 20, 4731–4745. <https://doi.org/10.5194/hess-20-4731-2016>.
- Zhai, X., Guo, L., Liu, R., Zhang, Y., 2018. Rainfall threshold determination for flash flood warning in mountainous catchments with consideration of antecedent soil moisture and rainfall pattern. *Nat. Hazards* 94, 605–625. <https://doi.org/10.1007/s11069-018-3404-y>.

- Zhang, Z., 2016. Introduction to machine learning: k-nearest neighbors. *Ann. Transl. Med.* 4, 218. <https://doi.org/10.21037/atm.2016.03.37>.
- Zhou, J., Peng, T., Zhang, C., Sun, N., 2018. Data Pre-Analysis and Ensemble of Various Artificial Neural Networks for Monthly Streamflow Forecasting. *Water* 10, 628. <https://doi.org/10.3390/w10050628>.
- Zhang, J., Hou, G., Ma, B., Hua, W., 2018. Operating characteristic information extraction of flood discharge structure based on complete ensemble empirical mode decomposition with adaptive noise and permutation entropy. *J. Vib. Control* 24, 5291–5301. <https://doi.org/10.1177/1077546317750979>.
- Zhou, L., Wu, X., Ji, Z., Gao, G., 2017. Characteristic analysis of rainstorm-induced catastrophe and the countermeasures of flood hazard mitigation about Shenzhen city. *Geomat. Nat. Hazards Risk* 8, 1886–1897. <https://doi.org/10.1080/19475705.2017.1392368>.
- Ziegler, A.D., 2012. Reduce urban flood vulnerability. *Nature* 481, 145. <https://doi.org/10.1038/481145b>.