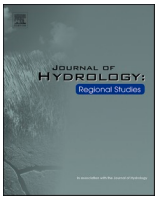
Journal of Hydrology: Regional Studies 41 (2022) 101087

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh



Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model

Faria T. Zahura [a,b], Jonathan L. Goodall [a,b,*]

[a] Department of Engineering Systems and Environment, University of Virginia, 151 Engineers Way, Charlottesville, VA 22904, USA
[b] Link Lab, School of Engineering and Applied Sciences, University of Virginia, Charlottesville, VA, USA

ARTICLE INFO

Keywords:
Coastal flooding
Machine learning
Surrogate modeling
Flood modeling

ABSTRACT

Flooding increases in recent years, in particular for coastal communities facing sea level rise, have brought renewed attention to real-time, street-scale flood forecasting. Such flood models using conventional physics-based modeling approaches are often unrealistic for real-time decision support use cases due to their long model runtime. Machine learning offers an alternative strategy whereby a surrogate model can be trained to mimic relationships present within the physics-based model and, after training, can run in seconds rather than hours. This study used the Random Forest (RF) algorithm to emulate a 1D/2D physics-based model simulating surface water depths in an urban coastal watershed in Norfolk, Virginia. Environmental features from a selected set of pluvial and tidal flood events and topographic information of the roadway were the input variables to train the surrogate model. Results show the potential for the surrogate model to predict flood extent and depth for both pluvial and tidal flood events. Furthermore, the surrogate model can differentiate between flooding locations dominated by pluvial or tidal flooding or impacted by both flooding mechanisms. Flood reports from the mobile app Waze were used for model validation and show 90% agreement with flooding locations from the surrogate model. Finally, feature importance methods were investigated to interpret the performance of the RF models and understand the contribution of different physical features to localized flooding.

1. Introduction

The projected global increase in sea level rise (SLR) and heavy rainfall threatens to increase the frequency and severity of coastal flooding worldwide (Church et al., 2013; USGCRP, 2017). While most past studies estimating the impacts of SLR focused on rarely occurring extreme events (Hallegatte et al., 2013; Hinkel et al., 2014; Hsiang et al., 2017), the impacts of seemingly less threatening high tide flooding have been highlighted in recent studies. High tide flooding, also known as "nuisance flooding," has increased 5- to 10-fold since 1960 and will continue to increase with SLR projections (Dahl et al., 2017; Moftakhari et al., 2015; Sweet et al., 2018; USGCRP, 2017). Although short-duration nuisance flooding is less destructive on a per-event basis, it can disrupt the transportation systems multiple times annually (Jacobs et al., 2018) and negatively impacts the local economy (Hino et al., 2019). In coastal cities, high tide levels compromise the sewer system by inundating the outlets and often flowing backward to inland areas (Sadler et al., 2020; Shen et al., 2019). The joint occurrence of rainfall and high tide can worsen the impact of nuisance flooding events (Lian et al., 2015; Shen et al., 2019). Therefore, real-time flood forecasting at the street-scale is needed for early flood warning systems and flood risk

* Correspondence to: Link Lab, 2nd floor of Olsson Hall, 151 Engineers Way, Charlottesville, VA 22904, USA.
E-mail address: goodall@virginia.edu (J.L. Goodall).

https://doi.org/10.1016/j.ejrh.2022.101087
Received 1 December 2021; Received in revised form 20 March 2022; Accepted 14 April 2022
Available online 22 April 2022

management (Henonin et al., 2013).

The use of physics-based 1D/2D dual drainage models is a traditional approach for simulating the depth and extent of surface flooding in urban areas (Kourtis et al., 2017). These models simulate flow through sewer/river systems and overland flow by coupling 1D and 2D hydrodynamic models (Fan et al., 2017; Lin et al., 2006; Seyoum et al., 2012). Many software packages such as HEC-RAS 2D, MIKE FLOOD, SOBEK (Deltares, 2018), or TUFLOW (BMT WBM, 2016) are commercially available to facilitate 1D/2D flood simulation in urban scenarios. Although this approach can simulate realistic urban flooding with high resolution and accuracy, the associated computational time can limit the application for early flood warning systems (Lhomme et al., 2006; Tanaka et al., 2011). While the advancement in graphical processing unit (GPU) based parallel computing has sped up the execution of these 2D shallow water equation models for flood simulation, high-resolution, real-time flood forecasting is still challenging (Guo et al., 2020).

To enable real-time flood forecasting, recent studies have used "response surface surrogates" or "metamodels" (Box and Wilson, 1951; Razavi et al., 2012; Simpson et al., 2001; Zhang et al., 2018), where machine learning (ML) acts as a surrogate to emulate

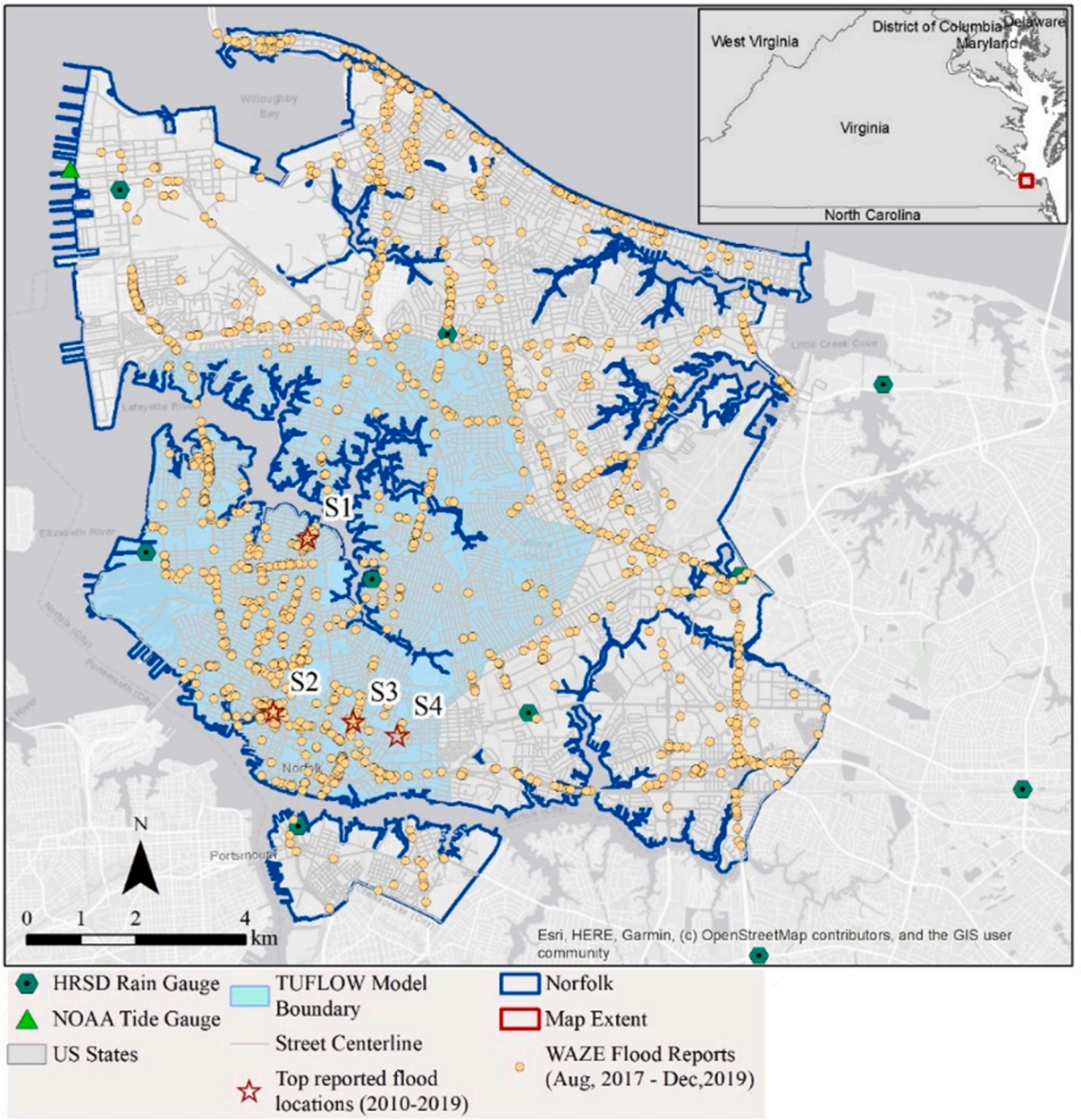


Fig. 1. Map showing the City of Norfolk with rainfall and tide gauges, physics-based model TUFLOW boundary, locations of flood reports from Waze, and road segments (S) with highest flood reports within the model boundary.

complex, high-fidelity physics-based models (Berkhahn et al., 2019; Contreras et al., 2020). These models are similar to the digital twin approach, which is popular in manufacturing systems and building information modeling (Lu et al., 2019; Schleich et al., 2017) and is beginning to be applied in water resources (Bartos and Kerkez, 2021; Inan et al., 2021). In this approach, a database is generated with several explanatory variables and simulation outputs from a physics-based model to train the surrogate model. The surrogate model learns from the input/output relationship rather than simulating complex, physical processes to emulate the physics-based model generated outputs. After training, the surrogate model is validated on independent data and then can make predictions on provided new input data with predictions made in seconds rather than the hours required by the original physics-based model (Berkhahn et al., 2019; Zahura et al., 2020).

Response surface surrogate models for real-time flood forecasting applications have been used by several studies to predict flood depths, volume, and velocity (Bass and Bedient, 2018; Bermúdez et al., 2019, 2018; Chang et al., 2010; Contreras et al., 2020; Jhong et al., 2017; Liu and Pender, 2015; Löwe et al., 2021). Most past studies have explored the ability of surrogate models in approximating pluvial and fluvial flood events or extreme typhoon events. However, the past studies have not emphasized relatively frequent nuisance flooding events due to high tide and rainfall in an urban coastal environment. Only two studies used tidal data as inputs for pluvial flood prediction (Bermúdez et al., 2019; Zahura et al., 2020). None of the past studies investigated the potential of ML surrogates to predict flooding due to high tide or the combined occurrence of rainfall and high tide in coastal cities. Given that tidal flooding has increased in US coastal cities with cumulative impacts surpassing the impacts of rare extreme events (Moftakhari et al., 2018), a real-time flood forecasting model predicting not only flooding from extreme events but also nuisance flooding in coastal cities is urgently needed. Additionally, most of the prior studies ignored topographic data in training the surrogate models. Some past studies used reference points to make flood predictions followed by interpolation based on the topology. Only studies by Löwe et al. (2021) and Zahura et al. (2020) used topographic information from urban settings in model training and found them essential in emulating pluvial flood depths. The capability of the ML surrogates in approximating different flood dynamics based on topography is yet to be investigated.

To address the existing knowledge gaps, this study builds on the work done by Zahura et al. (2020) to create a response surface surrogate model that emulates pluvial and tidal flooding simulated by a coupled 1D/2D hydrodynamic model, TUFLOW, calibrated and validated for the coastal city Norfolk, Virginia, USA. An ML algorithm, Random Forest (RF), was trained using two types of input features, (i) roadway topography and (ii) environmental information, for selected flood events, while the output was TUFLOW-simulated hourly water depths on the roadways in the study domain. The RF surrogate's performance in predicting both pluvial and tidal flooding was evaluated for two different training approaches (i) training with only pluvial and no tidal events and (ii) training with both pluvial and tidal events. These two training approaches were compared to evaluate the value of including tidal flooding explicitly within the training data. Sixteen rainfall-dominant and eight tide-dominant events were used for this purpose. The surrogate models were tested on four rainfall and two high tide dominant events independent of the training events. The physics-based and surrogate models were validated using crowdsourced flood reports from the mobile app Waze to build trust in the model-predicted flooding. The surrogate model's potential to differentiate between tidal-dominate and rainfall-dominate flooding while making predictions was also investigated, as were approaches for gaining insight and interpretability of the RF model output. Overall, this study advances the potential for creating surrogate ML flood models for real-time, street-scale decision support that can capture combined pluvial and tidal flooding in coastal-urban environments.

2. Study area and data

2.1. Study area

The City of Norfolk, Virginia (Fig. 1), located along the US east coast, is one of the most vulnerable coastal cities to SLR (Fears, 2012). This coastal city and the surrounding Hampton roads region are experiencing two times faster local SLR than the global SLR due to land subsidence and ocean circulation dynamics (Atkinson et al., 2013). The frequency of nuisance flooding in Norfolk has increased by 325% since 1960 (Burgos et al., 2018). This increased flood frequency threatens the city, which is home to 244,000 people and vital for the US economy and national security. The City of Norfolk is actively adapting coastal resilience strategies to become resilient to extreme events.

2.2. Environmental data

The surrogate model required environmental data, such as rainfall and tide level observations, as inputs to predict water depths on streets based on characteristics of the given events. These data were collected from Hampton Roads Sanitation District (HRSD) and U.S. National Oceanic and Atmospheric Administration (NOAA) stations from Jan 1, 2016, to Dec 31, 2018. Ten HRSD rainfall stations were used to collect 15-minute rainfall observations. Daily rainfall and hourly tide level observations referenced to the North American Vertical Datum (NAVD88) were collected from NOAA's Norfolk International Airport (NOAA, 2018a) and Sewells Point Station (NOAA, 2018b), respectively.

2.3. Topographic and roadway data

Topographic data were required as model input to distinguish between the topographic characteristics of the roadways causing flooding and validate the models with crowdsourced data. From the U.S. Geological Survey (USGS), a Digital Elevation Model (DEM)

with 1 m resolution (USGS, 2016) and LiDAR (Light Detection and Ranging) point cloud data (USGS, 2013) for Norfolk were collected. A shapefile of Norfolk street centerlines was collected through the city's official website (City of Norfolk GIS Bureau, 2018).

2.4. Crowdsourced street-scale flood data

Crowdsourced flood reports were collected from two sources: the City of Norfolk's System to Track, Organize, Record, and Map (STORM) (Sep 2010- Dec 2019) and Waze (Aug 2017-Dec 2019). The city uses STORM to record the impacts from storms, including the coordinates of flooded locations reported by the city workers at a daily timescale. Waze, owned by Google, is a GPS navigation app that provides real-time traffic updates and allows riders to report live updates on the street condition, such as accidents, flooded streets, etc. The exact time and coordinates of reported flooding were collected from Waze. The four locations with maximum reported flooding were identified using both STORM and Waze reports (Fig. 1). Flood reports around these locations mainly occurred during high tide (S1 in Fig. 1), during both high tide and rainfall (S2 in Fig. 1), and during rainfall events (S3 and S4 in Fig. 1). As Waze provided more extensive and diverse flood data with exact reporting time than STORM, it was used for model validation.

2.5. Simulated street-scale flood depth data

A 1D/2D dual drainage model, TUFLOW, built for a large area in Norfolk by Shen et al. (2019), was used to simulate street-level flood depth and generate output to train the surrogate model. TUFLOW links 1D pipe flow solved using ESTRY to 2D surface flow solved using 2D shallow water equations. The urban flood model covered an area of 56.4 km^2 in Norfolk. The pluvial flooding inside this area was generated internally, as no rainfall-generated flow from the surrounding basins entered the study domain. The tide level at the outlet was considered the same as Sewells Point station, which is the closest tide gauge to the study area. This model required rainfall and tide level to simulate detailed surface flooding maps for every hour of the event at a spatial resolution of 5 m, which was used to build the surrogate model.

3. Methodology

3.1. Model data preprocessing

Four environmental and three topographic features were generated from the raw environmental and topographic data, respectively (Table 1). The 15-min interval rainfall data from HRSD were aggregated to produce three rainfall features: hourly rainfall to describe the rainfall during the hour of interest, cumulative rainfall in the last 2 hr and cumulative rainfall in the last 72 h to describe antecedent soil moisture conditions or exceedance of stormwater system capacity from the past occurrences of rainfall. Due to the spatial variability of rainfall among stations, rainfall features were interpolated across the study domain using inverse distance weighted interpolation. As tide level is one of the dominant factors for flooding in coastal cities, hourly tide level was used as an input feature.

The three topographic features derived from 1 m DEM were elevation, topographic wetness index (TWI), and depth to water (DTW). Elevation was used to describe the variation of street elevations within the study domain. TWI explains the likelihood of an area accumulating surface runoff based on local topography, with higher TWI indicating a higher tendency for runoff accumulation. TWI is defined as

$$\mathrm{TWI} = \ln\left(\frac{\alpha}{tan(\beta)}\right), \tag{1}$$

where α = contributing area per unit contour length and $tan(\beta)$ = local slope (Beven and Kirkby, 1979).

DTW measures the soil moisture condition (Murphy et al., 2007) for *i,* a pixel on the landscape as

$$\mathrm{DTW} = \left[\sum\left(\frac{dz_i}{dx_i}\right)a\right]x_c \quad , \tag{2}$$

where $\sum\left(\frac{dz_i}{dx_i}\right)$ = sum of slopes across the least-slope path from *i* to the closest pixel of waterbodies, a = 1 or $\sqrt{2}$ based on whether the

Table 1
Description of input features (Zahura et al., 2020).

Input Features	Feature abbreviation	Unit	Variability
	Environmental features		
Total hourly rainfall	RH	mm	Spatial and temporal
Cumulative rainfall in the previous 2 hr	HR_2	mm	
Cumulative rainfall in the previous 72 hr	HR_72	mm	
Hourly tide level	TD_HR	m	Temporal
	Topographic features		
Elevation	ELV	mm	Spatial
Topographic wetness index	TWI	–	
Depth to water	DTW	cm	

pixel boundary is crossed parallelly or diagonally by the path and x_c = pixel resolution. The lower value of DTW indicates the pixel on the landscape is closer to waterbodies in terms of both vertical and horizontal distance; therefore, wetter soil.

The input file for the surrogate model was prepared by extracting the input features for the streets within the study domain. Regarding the restriction of vehicular movement on a road during flood events, we assumed that the deepest flood depth along a street would be sufficient for road closure decisions. Therefore, the street centerline shapefile for Norfolk was divided into segments with 50 m length. Due to lack of road width data, we assumed the roads consisted of two lanes, each 3.6 m wide, the average lane width in the US (US Department of Transportation, 2014). The ArcGIS software system (Esri, 2020) was implemented to prepare a shapefile of street segments, where each segment was 50 m long and 7.2 m wide. This shapefile was used to extract the mean and maximum of the input features and TUFLOW-simulated hourly water depths, respectively, at each segment. The information on pump stations operational at the underpasses in Norfolk was unavailable to be incorporated in the TUFLOW model. Therefore, the street segments representing underpasses were excluded from the analysis, resulting in 16,914 segments to develop the surrogate model.

3.2. Training and testing data

The RF surrogate models were developed using training data containing only rainfall events (RF-P) and both rainfall and high tide events (RF-P&T). The daily rainfall data and higher high tide data collected from NOAA were used to select the top 20 daily rainfall events (Table 2a) and the top 10 high tide events (Table 2b). These events were divided into training and testing with an 80%/20% split (Agranoff et al., 2006) with 16 rainfall events in training for RF-P, and 16 rainfall and 8 tide events in training for RF-P&T. The number of hours for the rainfall and tidal training events were 375 and 282, respectively. In our previous study, we found that oversampling the minority class with flood depths $\geq$ 0.3 m and maintaining a one-to-one ratio between minority and majority (depth $<$ 0.3 m) groups improved predictive performance (Zahura et al., 2020). Using a 1:1 ratio between these groups resulted in training samples with 12,227,103 rows and 7 columns, and 21,754,398 rows and 7 columns for RF-P and RF-P&T, respectively.

The predictive capability of both models was tested on four rainfall and two tide events. Test events were chosen based on the availability of Waze data. The chosen rainfall events for testing had a wide variety of storm events, while testing high tide events were selected to represent tide events with considerable and minor occurrence of rainfall.

Table 2a
Rainfall events to train and test the surrogate models.

Date	Daily rainfall (mm)	Maximum hourly rainfall averaged across stations (mm)	Higher high tide (m from NAVD)	Number of flood reports from Waze	Train or test
10/8/2016	188.98	29.74	0.739	–	Train
7/31/2016	177.29	34.29	0.475	–	Train
9/21/2016	99.82	21.84	0.734	–	Train
8/29/2017	99.82	13.97	1.094	–	Train
8/11/2018	94.49	27.94	0.609	105	Test
9/19/2016	77.22	23.37	0.643	–	Train
5/6/2018	65.28	18.03	0.244	15	Test
9/3/2016	61.21	13.21	1.377	–	Train
9/20/2016	60.45	10.67	0.750	–	Train
7/30/2018	59.94	18.54	0.479	60	Train
6/22/2018	57.91	15.49	0.684	9	Train
6/5/2016	53.59	28.19	0.624	–	Train
10/29/2017	53.09	7.06	0.603	12	Test
8/20/2018	52.32	35.31	0.610	24	Train
5/28/2018	47.75	13.46	0.633	19	Test
10/9/2016	45.72	19.30	1.270	–	Train
7/15/2017	45.47	28.96	0.441	–	Train
8/9/2016	44.70	15.49	0.401	–	Train
1/2/2017	43.94	8.38	0.390	–	Train
8/7/2017	43.94	16.76	0.445	–	Train

Table 2b
High tide events to train and test the surrogate models.

Date	Daily rainfall (mm)	Maximum hourly rainfall averaged across stations (mm)	Higher high tide (m from NAVD)	Number of flood reports from Waze	Train or test
1/23/2016	37.80	13.61	1.114	–	Train
9/19/2017	0	0	1.098	28	Train
9/9/2018	39.40	6.45	1.082	22	Test
9/10/2018	0	0	1.076	28	Train
5/5/2016	1.30	0.66	1.072	–	Train
3/21/2018	6.40	0.66	1.021	3	Train
9/29/2016	9.90	4.90	1.012	–	Train
2/9/2016	0.30	0.30	0.996	–	Train
1/24/2017	0	0	0.994	–	Train
11/8/2017	8.60	2.82	0.981	19	Test

3.3. ML algorithm: RF regressor

RF was used in this study because it is faster to tune and train than the other ML methods, and preprocessing inputs to improve predictive accuracy is generally not needed (Ahmad et al., 2017). The RF regressor used in this study is an ensemble ML algorithm that uses many decorrelated regression trees to make predictions (Breiman, 2001; Breiman et al., 1984). Tree-based approaches tend to overfit to the training data. Random forest solves this problem by creating a collection of multiple trees, where each tree learns from randomly chosen samples and input features, and predictions are made by averaging the predictions from each tree. Additionally, the overfitting problem can be solved by not using fully grown trees in the forest or pruning the trees. This can be done by controlling the maximum depth of trees and the minimum number of samples required for splitting. In addition to making predictions, RF also provides an overall interpretation of feature importance, known as Gini importance, by measuring the total decrease in mean squared error by each feature. The RF regression was performed using sklearn.ensemble.RandomForestRegressor from python Scikit-learn module (Pedregosa et al., 2011; Scikit-learn Developers, 2018).

The RF regressor was optimized by tuning hyperparameters: number of trees (n_estimators), number of input features at each split (max_features), and the maximum depth of a tree (max_depth). The function GridSearchCV in the Scikit-learn module can find the optimal values of model hyperparameters by iterating all combinations of predefined hyperparameter values using k-fold cross-validation. Using the training data, this function was applied to tune n_estimators, max_features, and max_depth (Table 3) with 4-fold cross-validation, and the best values were 50, 5, and 'None', respectively. However, with max_depth = 'None', there was minimal improvement in the average accuracy from cross-validation for n_estimators above 30 (Fig. 2(a)). Similarly, with n_estimators = 30, increasing max_depth beyond 40 caused little improvement in accuracy (Fig. 2(b)). As the best values of max_depth and n_estimators did not improve model performance significantly and increased run time by consuming up to 97% computer memory, 40 and 30 were set as the parameter values, respectively. A sample weight of 2 was assigned for samples with water depths $\geq$ 0.2 m to maximize the predictive performance while fitting the RF models.

3.4. Model evaluation

The surrogate models accuracy and bias towards overprediction or underprediction in emulating TUFLOW-simulated time-series water depths on streets during the flood events were evaluated using root mean squared error (RMSE) and mean error (ME) for each street segment as

$$\mathrm{RMSE_s(m)} = \sqrt{\frac{\sum \left(y_{\mathrm{tuflow,s,h}} - y_{\mathrm{rf,s,h}}\right)^2}{n}} \tag{3}$$

Table 3
Hyperparameters optimized to improve model performance.

Hyperparameters	n_estimators	max_features	max_depth
Parameter values	1, 10–50 (incremented by 10)	1–7 (incremented by 1)	1, 5–50 (incremented by 5), and None
Best parameter values	50	5	None

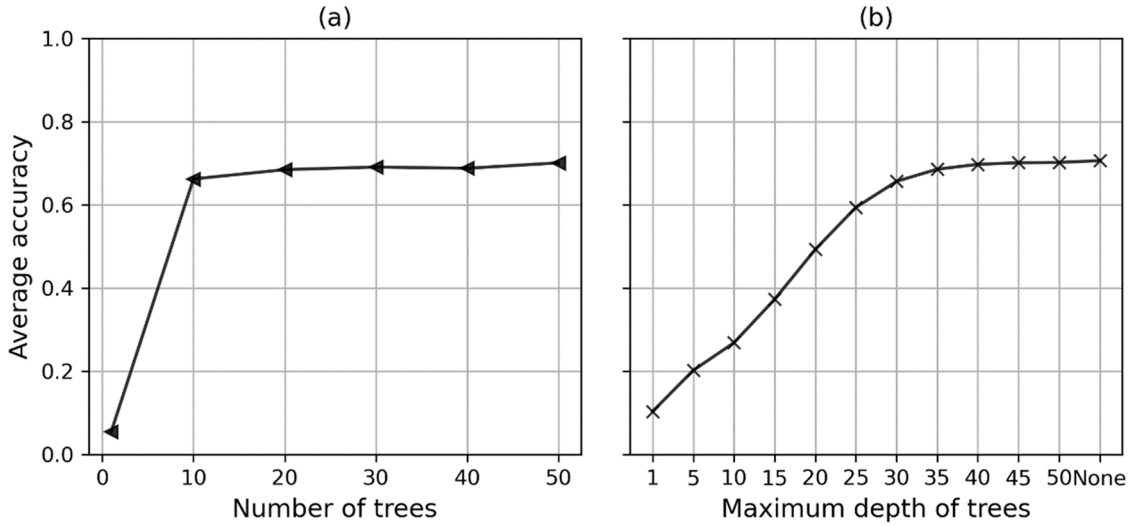


Fig. 2. Improvement in RF model performance for (a) increasing n_estimators with max_depth = None, and (b) increasing max_depth with n_estimators = 30.

$$ME_s(m) = \frac{\sum (y_{tuflow,\ s,\ h} - y_{rf,\ s,\ h})}{n}, \tag{4}$$

where y_{tuflow} and y_{rf} were TUFLOW-simulated and RF-predicted water depth, respectively, on a street segment (s) during an hour (h). "n" represented the total number of hours with rainfall > 0 mm or tide level > 0.8 m from NAVD or TUFLOW-simulated flood depths above 0.1 m during the events. According to NOAA, a tide level above 0.845 m from NAVD is the threshold for minor flooding in Norfolk (Sweet et al., 2018), and flooding can occur below this threshold (Burgos et al., 2018). Therefore, a tide level of 0.8 m above NAVD was used to identify flood depth relevant to tidal flooding.

The spatial performance of the surrogate models in mapping the maximum inundation extent and magnitude was evaluated for threshold depths: 0.1, 0.2, and 0.3 m. The inundation extent was defined with a wet/dry threshold of 0.1 m (Bermúdez et al., 2019; Shen et al., 2019), limiting high-velocity water depth on streets (Shand et al., 2011). Thresholds of 0.2 and 0.3 m were used to identify different flood intensities. As the air inlet height of passenger vehicles ranges within 0.25–0.35 m, a street becomes impassable at 0.3 m of water depth (AusRoads, 2008; Yin et al., 2016). Also, 0.3 m water depth can wash away cars and should be closed to ensure safety (NOAA National Weather Service, 2018). Therefore, replicating the extent of such streets was important. Three performance metrics, precision, recall, and F1 scores, were used to evaluate RF's potential in emulating spatial flood patterns.

Recall, representing the percentage of TUFLOW-simulated flooded streets, emulated correctly by RF, is calculated as

$$\text{Recall} = \frac{\text{RF-predicted true flooded streets}}{\text{Total TUFLOW-simulated flooded streets}}, \tag{5}$$

Precision represents the percentage of correct RF-predicted flooded streets and calculated as

$$\text{Precision} = \frac{\text{RF-predicted true flooded streets}}{\text{Total RF-predicted flooded streets}}, \tag{6}$$

Both precision and recall values range between 0 and 1. Lower recall and precision indicate underprediction and overprediction of flooding, respectively, compared to TUFLOW simulations. Higher values of these metrics are important for prompt action in hazardous locations and avoiding unnecessary actions in less threatening areas. F1-score, the harmonic mean of precision and recall, is calculated as

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{7}$$

The F1-score can reach a maximum value of 1 when both recall and precision are 1, indicating RF-predicted and TUFLOW-simulated inundation extents matched perfectly. The sklearn.metrics module in Python (Scikit-learn Developers, 2018) was used to calculate the performance measures.

3.5. Crowdsourced data processing and model validation

Flood reports from Waze were used to validate the location and time of flood occurrences during the six test events. Waze data provide binary "yes" or "no" values rather than the flood depth. To identify the flooded locations surrounding Waze flood reports, the

"Viewshed" tool on ArcGIS was implemented. This tool determines locations on a raster surface visible from the observer location, which is the location of Waze reports in this study. LiDAR point cloud data for Norfolk were utilized to produce a DEM raster. The LiDAR point cloud data contain tree canopy heights over streets resulting in a higher elevation in the LiDAR-derived DEM than the street elevation and interfered with viewshed analysis. At these locations, elevation from LiDAR-derived DEM was substituted with

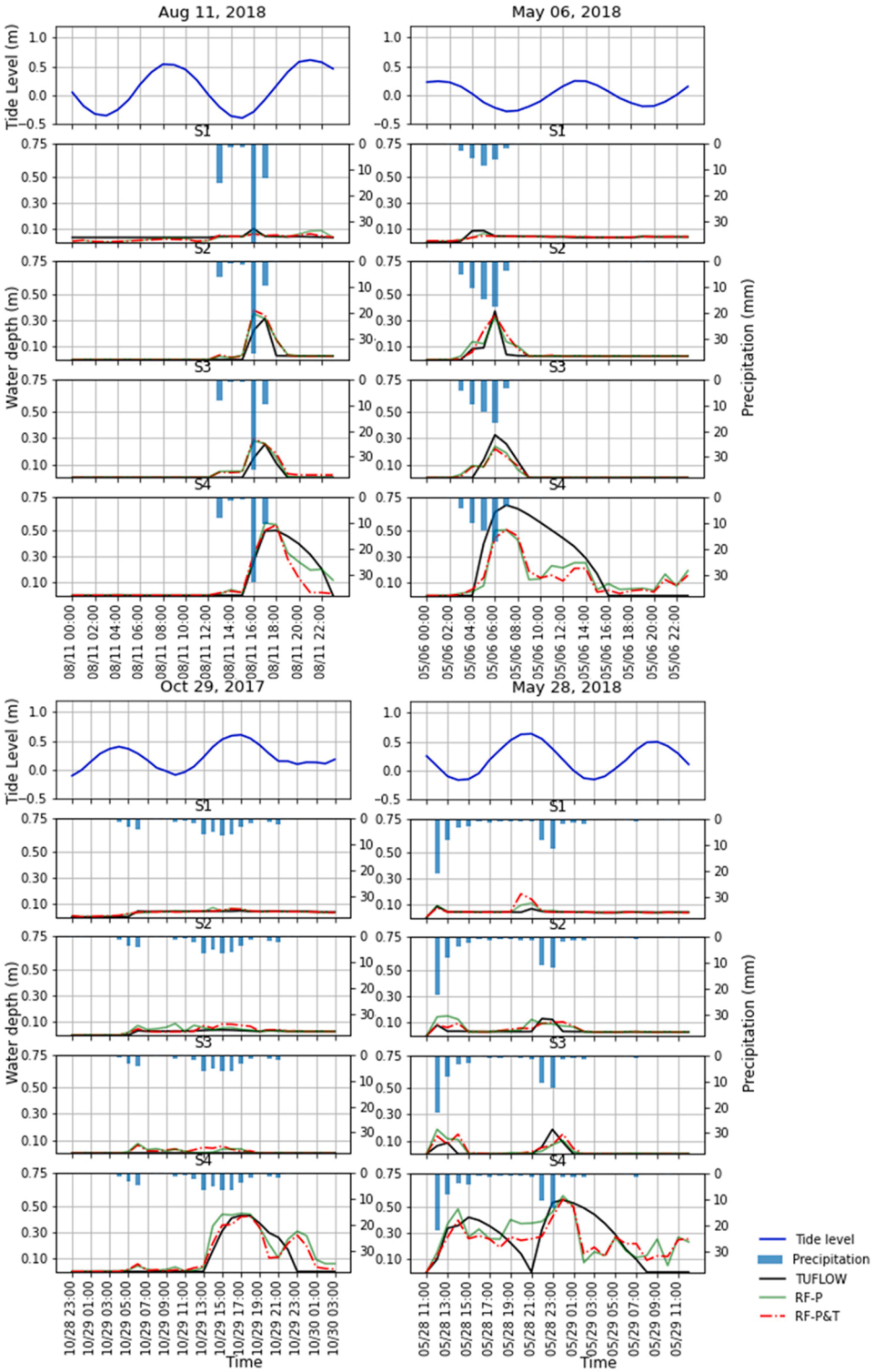


Fig. 3. Time series water depth on the four most reported flood locations from TUFLOW, RF-P, and RF-P&T during the pluvial testing events.

street elevation from bare earth DEM to obtain continuous viewshed. Street segments with > 50% area covered by viewshed were used to extract the maximum water depth from TUFLOW and RF during the flood events.

The performance of the models was analyzed using hit rate (H), which describes the proportion of Waze reported flood locations identified with > =0.10 m of water depth by the models, defined as

$$\mathrm{H}(\%) = \frac{\mathrm{Hits}}{\mathrm{Hits} + \mathrm{Misses}} \tag{8}$$

where "Hits" and "Misses" indicated a flood report from Waze with ≥ 0.10 m and < 0.10 m of water depth, respectively, within its viewshed from the flood models. The time difference between Waze reports and peak flood depths was also evaluated.

4. Results and discussion

4.1. Approximating flood depth

Comparison between TUFLOW-simulated, and RF-P and RF-P&T predicted water depths on the four highest reported flooded streets during the pluvial and tidal test events are shown in Fig. 3 and Fig. 4, respectively. The RMSEs between TUFLOW-simulated and surrogate model predicted water depths at the four locations and across the study domain are listed in Table 4.

As flooding on S1 is mostly tide-driven, the intensity of TUFLOW-simulated flooding on this segment was low during the pluvial events, which both surrogate models emulated. At S2 and S3, the average difference between the TUFLOW-simulated and RF-P predicted peaks on Aug 11, 2018, and May 6, 2018 events were 0.041 and 0.057 m, respectively. On Oct 29, 2017 event, no flooding was simulated on S2 and S3. The May 28, 2018 event had two separate rainfall peaks resulting in two flood peaks simulated by TUFLOW. The first peak was overpredicted on S2 by 0.069 m by RF-P and S3 by 0.10 and 0.065 m by RF-P and RF-P&T, respectively. At S4, as the surrogate models drained out the accumulated water earlier than TUFLOW, this segment had RMSE higher than the 90th

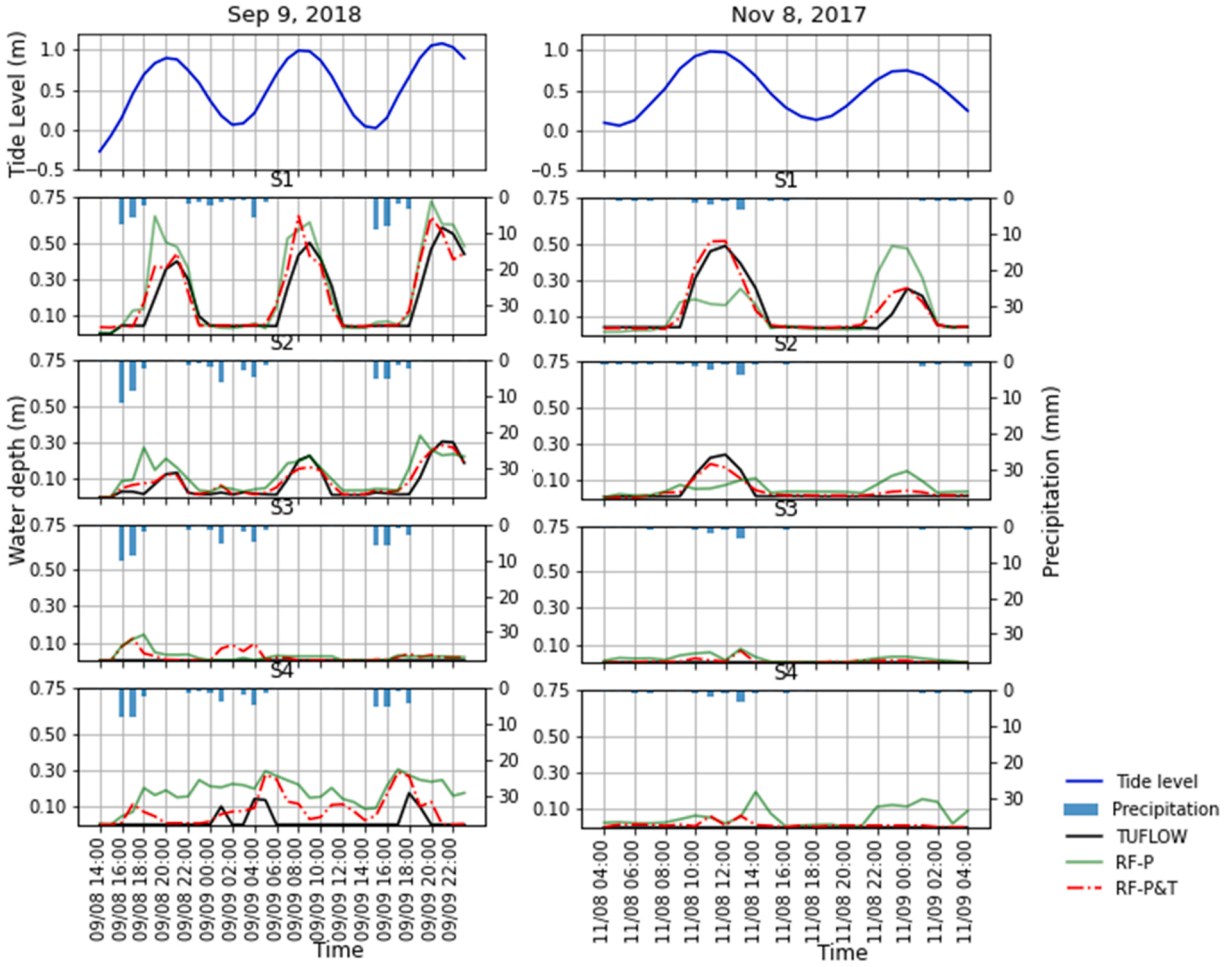


Fig. 4. Time series water depth on the four most reported flood locations from TUFLOW, RF-P, and RF-P&T during the high tide testing events.

Table 4
RMSEs (m) of water depth from the surrogate models at the four most reported flood locations with average and 90th percentile RMSEs across the study domain during the testing events.

Events	Surrogate Model	S1	S2	S3	S4	Average RMSE	90th percentile RMSE
				Pluvial events			
Aug 11, 2018	RF-P	0.012	0.075	0.066	0.076	0.054	0.078
	RF-P&T	0.017	0.082	0.074	0.149	0.054	0.078
May 6, 2018	RF-P	0.015	0.039	0.045	0.213	0.045	0.054
	RF-P&T	0.017	0.057	0.048	0.216	0.050	0.060
Oct 29, 217	RF-P	0.01	0.021	0.024	0.093	0.026	0.035
	RF-P&T	0.008	0.020	0.026	0.083	0.029	0.041
May 28, 2018	RF-P	0.015	0.042	0.045	0.180	0.030	0.038
	RF-P&T	0.034	0.026	0.047	0.150	0.032	0.043
				Tidal events			
Sep 9, 2018	RF-P	0.121	0.080	0.041	0.177	0.063	0.090
	RF-P&T	0.084	0.031	0.040	0.097	0.036	0.050
Nov 8, 2017	RF-P	0.146	0.074	0.030	0.077	0.046	0.063
	RF-P&T	0.046	0.025	0.017	0.020	0.029	0.041

percentile for the pluvial events. However, the surrogate models predicted the peak time precisely, and the predicted peak depths were close to the TUFLOW-simulated peak and above 0.3 m to cause road closure decision. During the pluvial events, the predicted time series water depths by RF-P and RF-P&T on the four segments were close, and the differences between peak values ranged within 0–0.059 m.

On Nov 8, 2017 event, TUFLOW-simulation showed increasing flood depths on S1 and S3 during the astronomical high tide periods, which receded during low tide. RF-P underpredicted the first and overpredicted the second peak on both S1 and S2. The differences between maximum depths were 0.23 and 0.13 m on S1 and S2, respectively. RF-P&T predicted the peak depth and time of flooding on both segments accurately with RMSE 0.046 and 0.025 m for S1 and S2, respectively. As S3 and S4 are prone to rainfall-driven flooding,

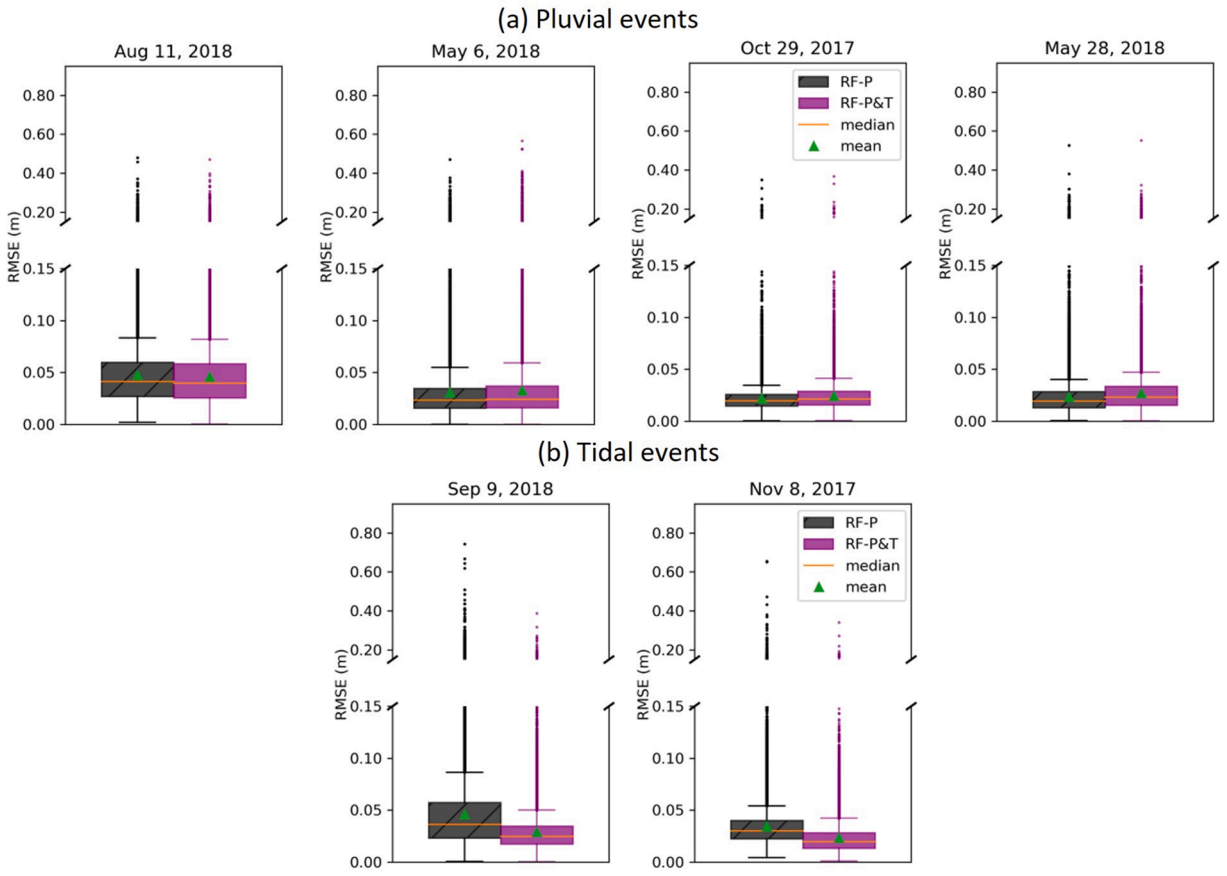


Fig. 5. Box plots of RMSE between water depths from surrogate model prediction and TUFLOW simulation during (a) pluvial events and (b) tidal events at different street segments.

no flooding was simulated on these segments. While RF-P overpredicted the depth on S4, RF-P&T emulated no flooding on this street. On Sep 9, 2018 event, RF-P overpredicted the peak depths on S1 with RMSE 0.121 m. RF-P&T prediction on S1 showed better agreement with TUFLOW-simulation compared to RF-P. However, RF-P&T predicted peaks occurred an hour earlier than the TUFLOW-simulated peak resulting in an RMSE (0.084 m) higher than 90th percentile RMSE (0.05 m). As S4 experiences rainfall-driven flooding, two flood peaks with depth $>$ 0.1 m appeared in the TUFLOW simulation during hours with rainfall rather than the high tide period. RF-P could not identify the separate flood peaks during the rainfall period and predicted water depth ranging within 0.09–0.3 m throughout the day. RF-P&T emulated the two flood peaks on S4 during the rainfall period; however, the peaks were overpredicted by 0.13 and 0.11 m.

Although RF-P predicted the presence of tidal flooding, the estimated water depths showed a high deviation from TUFLOW-simulated values. On the contrary, RF-P&T predicted flood depths with low errors for both pluvial and tidal events. Also, it differentiated between the locations driven by pluvial and tidal flooding and accurately predicted the time of occurrence.

The capability of the surrogate models to predict water depths across the street segments was measured using segment-wise RMSE, represented as histograms in Fig. 5. During Aug 11, 2018 event with the maximum daily rainfall among the testing events, the average and 90th percentile RMSE from RF-P were the highest among the pluvial events. Among 16,914 segments, 65% and 31% had RMSE $<$ 0.05 and within 0.05–0.10 m, respectively and only 4% had RMSE $>$ 0.10 m. 88% of the segments had RMSE $<$ 0.05 m on May 6, 2018 event with $<$ 3% falling above RMSE 0.10 m. The rest of the two pluvial events had $>$ 96% and $<$ 1% of the segments with RMSEs $<$ 0.050 and $>$ 0.10 m, respectively. For RF-P&T, having high tide events in training increased RMSEs during the pluvial events except for Aug 11, 2018 event. However, the increase in average and 90th percentile RMSEs was less than 0.005 and 0.006 m, respectively, indicating the inclusion of tide events did not significantly affect the performance of RF-P&T.

The predictive ability of RF-P was worse for the tidal events compared to the pluvial events. The 90th percentile RMSE on Sep 9, 2018 (0.09 m) was higher than Aug 11, 2018 (0.078 m) event, and on Nov 8, 2017 event was higher than RMSEs during the other three pluvial events. RF-P&T showed a significant decrease in RMSEs during the tidal events. On average, 92% and $<$ 1% segments had RMSE $<$ 0.05 and $>$ 0.10 m, respectively, from RF-P&T predictions.

Fig. 6 shows the ME between the simulated and predicted flooding on each road segment during pluvial and tidal flood events. Negative values of ME indicate the mean water depths were overpredicted, whereas positive values indicate underprediction. ME values for RF-P during pluvial events were closer to zero, with 1.3% and 2% of segments having ME $<$ -0.025 and ME $>$ 0.025 m, respectively. Using RF-P&T, the bias in prediction increased; however, only 3% and 4.2% of segments had ME $<$ -0.025 and ME $>$ 0.025 m, respectively. RF-P showed more bias toward overprediction of mean depth with ME $<$ -0.025 m for 35% of streets for tidal flood events, while 14.5% of the segments had ME $>$ 0.025 m. Using RF-P&T, this bias in tidal flood prediction was reduced to ME $<$ -0.025 m at 5% road segments and ME $>$ 0.025 m at 9% segments.

Although tide level was one of the input features for RF-P, there were high discrepancies between predicted and simulated water depth on tidal events. While RF-P&T maintained similar RMSEs like RF-P during pluvial events, it improved predictions during high tide events. Therefore, RF-P&T could serve as a unified flood prediction model, as it emulated TUFLOW-simulated flooding with low RMSEs on most of the segments for both types of flood events.

4.2. *Flood extent mapping*

The recall, precision, and F1 scores evaluating the maximum extent of flooding from RF-P and RF-P&T during the flood events for thresholds 0.1, 0.2, and 0.3 m are listed in Table 5. During the pluvial events for threshold $>$ =0.1 m defining flood boundary, recall scores ranged within 0.83–0.96, indicating RF-P emulated above 83% of the flood locations from TUFLOW-simulation. Precision scores ranged within 0.72–0.92, indicating more than 72% of the predicted flood locations were correct for the pluvial events. While precision scores remained consistently high throughout all the threshold values, recall values dropped with an increased threshold. RF-P

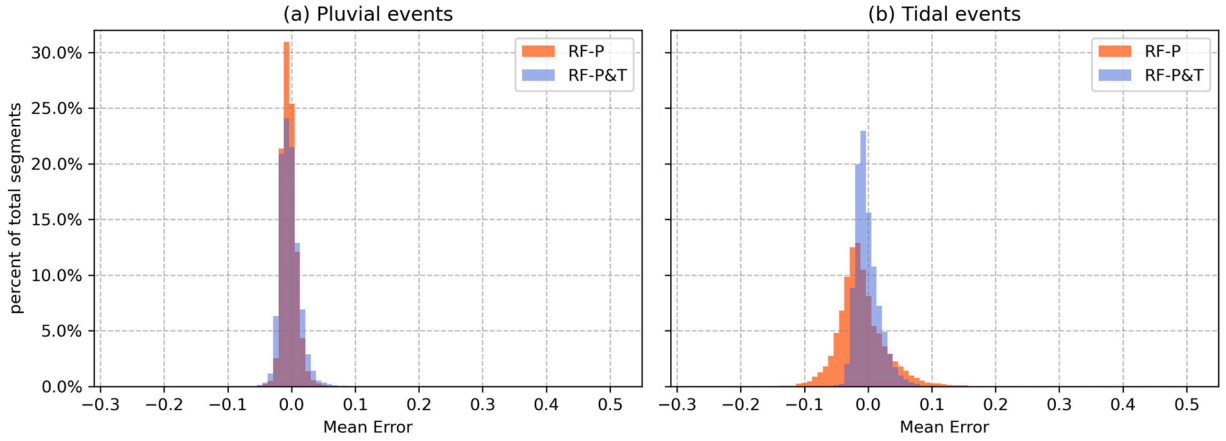


Fig. 6. Histograms of ME (m) between TUFLOW simulated and RF predicted water depths for all (a) pluvial and (b) tidal events.

correctly predicted 55%– 87% and 50%– 85% of the flooded streets for 0.2 and 0.3 m thresholds, respectively. Like RMSEs, introducing tidal events in training for RF-P&T caused minimal alteration in recall and precision scores, and the decrease in F1 scores was at most 0.04.

For the tidal events, with a threshold > =0.1 m, RF-P predicted the extent of flooding less accurately than the pluvial events, as reflected by the performance scores. For thresholds > =0.2 and > =0.3 m, F1-scores dropped below 0.58 for both tidal events, suggesting RF-P's poor predictive performance in emulating tidal flooding. With high tide events in training for RF-P&T, F1 improved up to 0.80 and 0.84 for events on Nov 8, 2017, and Sep 9, 2018, respectively. On the Nov 8, 2017, and Sep 9, 2018 events, RF-P&T correctly emulated 67% and 93% of the TUFLOW-simulated flood extent, and 71% and 78% of the roads unsafe to passenger vehicles, respectively. This suggested that using tide level as an input feature to develop surrogate flood models in an urban-coastal environment was insufficient to correctly emulate the extent of tidal flooding or combined pluvial and tidal flooding. Training events should have representation from different flood types.

4.3. Validation using crowdsourced flood report

Fig. 7 and Fig. 8 demonstrate the rainfall, tide level, and time of Waze reports, and the maximum water depth during the events within the viewshed of the Waze reports, respectively. Table 6 lists the H values quantifying the agreement between Waze reports and flooding from TUFLOW and RF-P&T. A maximum of 58 flood locations was reported on Aug 11, 2018, and 93% and 96.6% of the reports had maximum water depth ranging within 0.1–1 m inside viewshed, according to TUFLOW and RF-P&T, respectively. The lowest H was found for TUFLOW on Sep 9, 2018 event. By investigating the streets with base map imagery in ArcMap and TUFLOW-simulated water depth, it was found that two reports had water depth > 0.1 m within viewshed. However, creating segments with 7.2 m width did not cover the actual road width, and missed the corners of the roads where flooding happened, decreasing H. Similarly, 1, 1, and 3 flood reports on Nov 8, 2017, May 6, and Aug 11, 2018, respectively, had water depths > 0.10 m within viewshed, which were missed due to the geometric bias, reducing H. Overall, 87.5% and 90.4% of the Waze flood reports were associated with water depth > 0.10 m from TUFLOW and RF-P&T, respectively. While this agreement validated both models, it also suggested that RF-P&T could predict flood depth on the four highest reported flood locations and other user-reported locations across the study domain. The minor mismatch between Waze and flood models could be resulting from the uncertainty regarding crowdsourced data reliability (Boutsis et al., 2016). As experts do not generate these reports, the perception of a flood would widely vary among the users.

The distance between 95% of the flood reports and the location of maximum flood depth ranged within 0–150 m. Reports located > 150 m from the maximum flood depth had > 0.10 m of water depth within 0–47 m on streets. Fig. 9 demonstrates the difference between the time of the Waze reports and the peak depths. 68% of flood reports appeared within 1 h of the peak depth for all the test events, as indicated by the boxplots. The reports appearing 3–4 h from the peak depth on Oct 29, 2017, Nov 8, 2017, and Aug 11, 2018, had > 0.10 m of water depth during or within 1 h of the reports. The May 28, 2018 event had two peak rainfalls generating two flood peaks. Although the maximum flood depth was simulated during the second peak, flood reports appeared during the first peak (Fig. 7), with depths ranging between 0.08 and 0.23 m. Therefore, the time difference was calculated between the first peak and the flood reports for this event. Overall, 17%, 28%, and 55% of the reports appeared before, during, and after the time of peak flood depth,

Table 5
Recall, precision, and F1 scores for maximum flood extent.

Event	Metrics	Thresholds (m)					
		> =0.1		> =0.2		> =0.3	
		RF-P	RF-P&T	RF-P	RF-P&T	RF-P	RF-P&T
Pluvial events							
Oct 29, 2017	Recall	0.94	0.95	0.87	0.85	0.85	0.84
	Precision	0.90	0.83	0.95	0.89	0.93	0.92
	F1 score	0.92	0.89	0.90	0.87	0.89	0.88
May 6, 2018	Recall	0.83	0.80	0.55	0.54	0.50	0.48
	Precision	0.86	0.89	0.95	0.93	0.98	0.95
	F1 score	0.84	0.84	0.7	0.68	0.67	0.63
May 28, 2018	Recall	0.96	0.95	0.82	0.81	0.8	0.80
	Precision	0.73	0.72	0.85	0.83	0.91	0.91
	F1 score	0.83	0.82	0.83	0.82	0.85	0.85
Aug 11, 2018	Recall	0.95	0.93	0.68	0.7	0.61	0.68
	Precision	0.72	0.74	0.66	0.66	0.73	0.62
	F1 score	0.82	0.83	0.67	0.68	0.67	0.65
Tidal events							
Nov 8, 2017	Recall	0.60	0.67	0.51	0.60	0.61	0.71
	Precision	0.62	0.94	0.59	0.97	0.48	0.93
	F1 score	0.61	0.78	0.55	0.74	0.54	0.80
Sep 9, 2018	Recall	0.79	0.93	0.54	0.77	0.56	0.78
	Precision	0.60	0.76	0.63	0.87	0.59	0.84
	F1 score	0.68	0.84	0.58	0.82	0.57	0.81

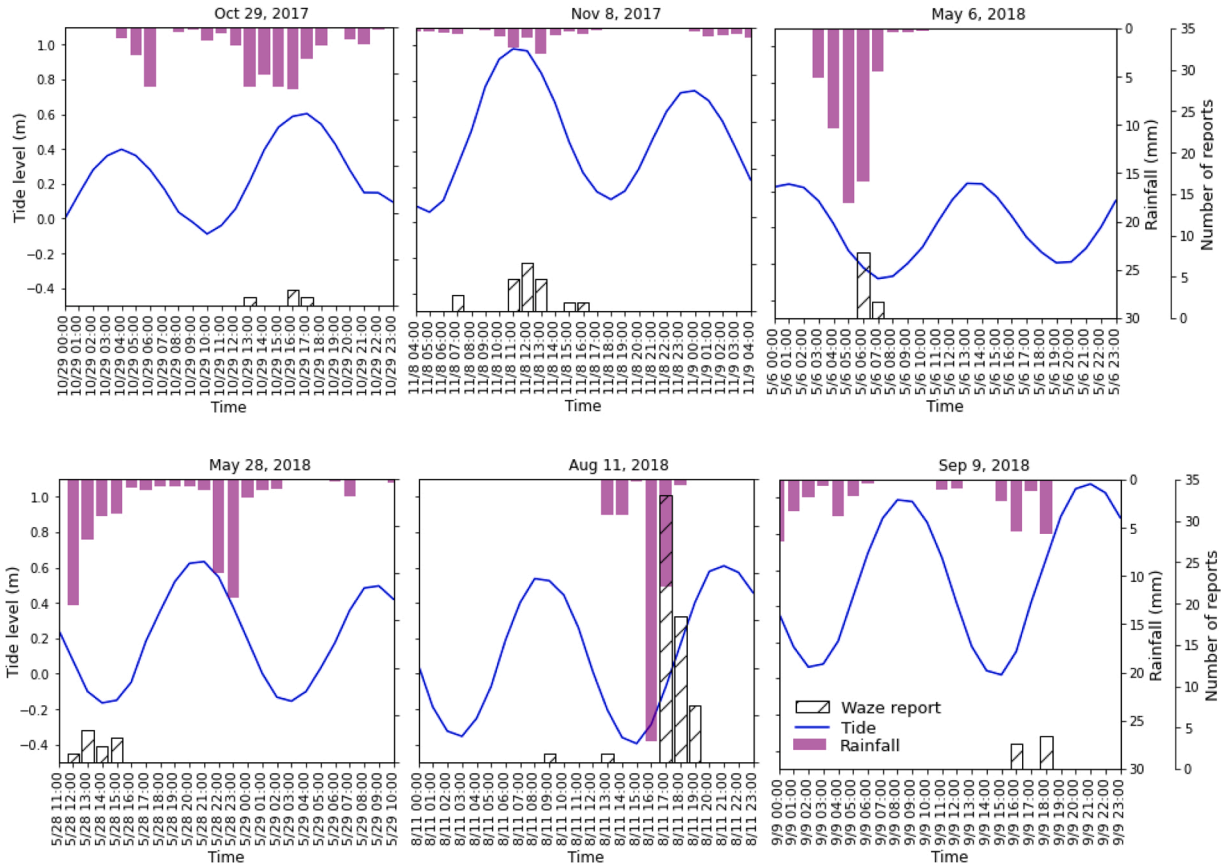


Fig. 7. Tide level, average rainfall across the rainfall stations, and time of reported flooding on Waze during the test events.

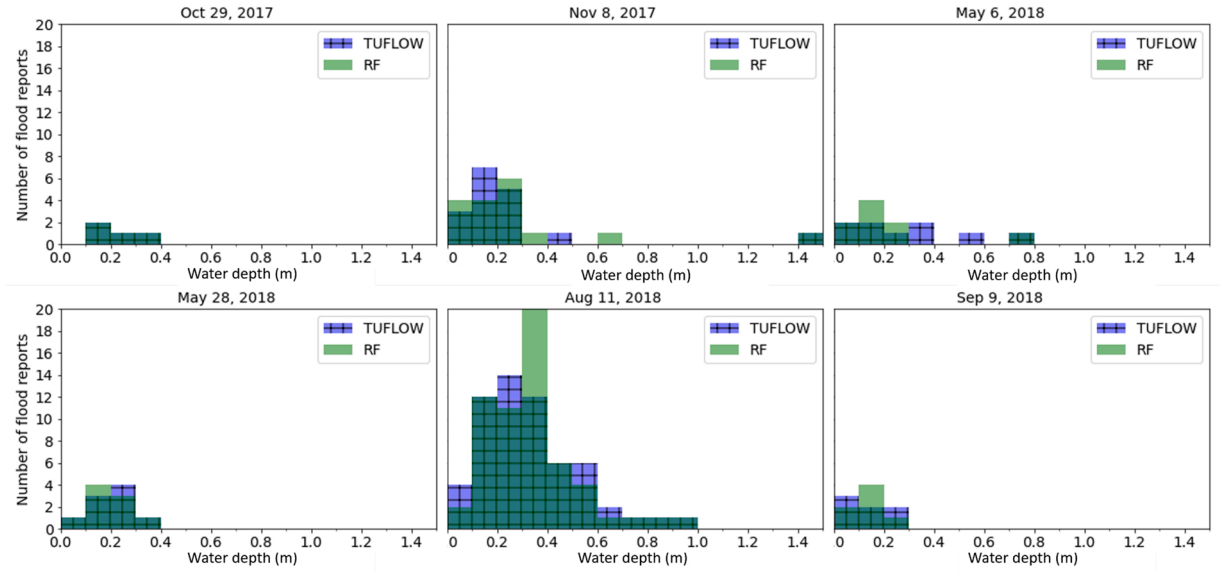


Fig. 8. Maximum water depths within the visible distance from Waze flood reports.

Table 6
Estimating the agreement between Waze flood report and flood occurrence from TUFLOW and RF-P&T.

Events	H (%)	
	TUFLOW	RF-P&T
Oct 29, 2017	100%	100%
Nov 8, 2017	82.4%	76.5%
May 6, 2018	77.8%	77.8%
May 28, 2018	88.9%	100%
Aug 11, 2018	93%	96.6%
Sep 9, 2018	57.1%	71.4%
Overall	87.5%	90.4%

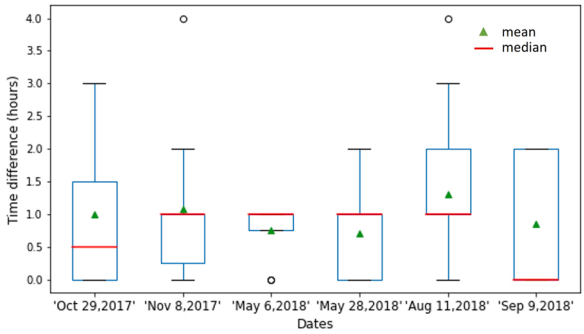


Fig. 9. Time difference between the occurrence of Waze report and peak flood depth.

respectively.

4.4. Feature importance

RF-generated Gini importance measures the global importance of each feature for the whole dataset (Saarela and Jauhiainen, 2021). Among the seven input features, RF-P found TWI to be the most important in predicting street flooding, followed by ELV and HR_72 (Fig. 10). TD_HR was the least important feature in RF-P, which could be due to the lack of representation of tidal flood events in training. However, RF-P&T computed importance of TD_HR dropped by 0.008 and was the least important feature, despite tidal flood incidents improving flood prediction. The importance of HR_72 increased by 0.10 and was the most important feature calculated by RF-P&T, while the importance of TWI dropped. While rainfall and tide are the commonly known features for flooding in coastal cities, RH and TD_HR were unimportant, according to RF. This deviation of feature importance from physical expectations could be due to the bias in Gini importance towards variables varying in scale, which are artificially preferred during splitting (Strobl et al., 2007). Also, Gini importance fails to find the true predictors when the relationship between predictors and response is weak (Archer and Kimes, 2008). Correlations between training features and water depths varied between − 0.30 and 0.43 (RF-P&T), indicating a weak predictor-response relationship. TWI had the highest correlation coefficient: 0.452 and 0.436 for RF-P and RF-P&T, respectively. The second most correlated feature was HR_72, which experienced an increase in correlation coefficient from 0.349 (RF-P) to 0.432 (RF-P&T), suggesting that correlation between the features and response might have influenced the global feature importance from RF.

An alternative approach for determining variable importance for RF is Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016). We applied LIME for S1 and S3 during hours 17:00 and 9:00 for the Aug 11, 2018, and Sep 9, 2018 events, respectively (Fig. 11). For the Aug 11, 2018 event, ELV and RH were the top two features contributing positively towards flood prediction for both segments. TWI was important for both segments, but contributed negatively to S1 and positively for S3. This may be because high TWI for S3 caused runoff accumulation during rainfall, whereas S1 had low TWI and was impacted more by tidal flooding. For the event on Sep 9, 2018, ELV and TD_HR contributed positively to the TUFLOW-simulated 0.499 m depth of water, while all the other features had a negative effect. In all four examples, ELV was the most important feature in explaining flood depth prediction, while HR_72 was consistently one of the least important features. While RH, HR_2, and TD_HR were the least important features according to Gini importance, they played a crucial role in flood prediction according to the local analysis, which is what one might expect given the physical understanding of flood generation. Therefore, we suggest there is more value in analyzing variable importance locally using an approach like LIME than global approaches like Gini for flood applications.

4.5. Computational cost

The comparison between the computational cost for TUFLOW and RF is listed in Table 7. The computational time for TUFLOW

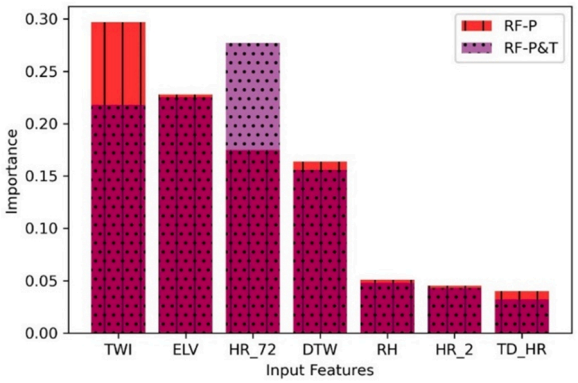


Fig. 10. Feature importance from RF-P and RF-P&T.

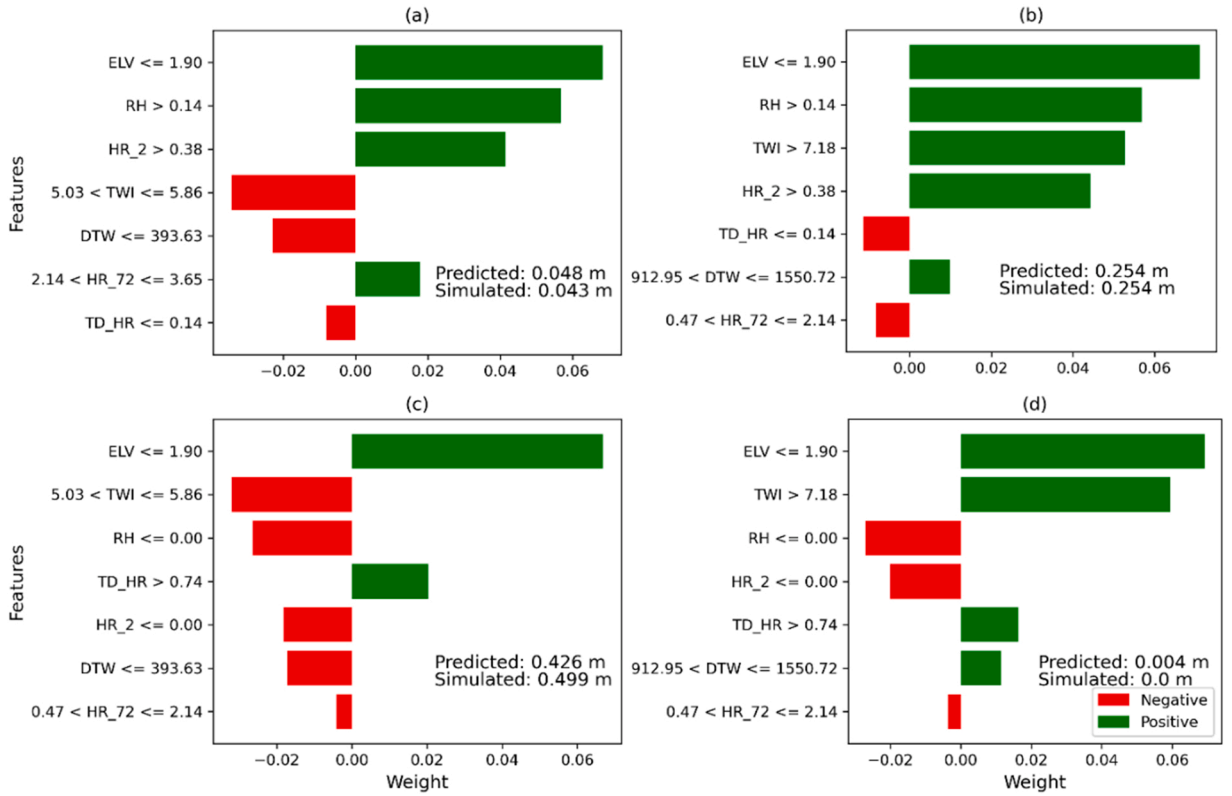


Fig. 11. Explaining individual predictions using LIME for Aug 11, 2018 event on (a) S1 and (b) S3, and Sep 9, 2018 event on (c) S1 and (d) S3. The green bars towards the right and red bars towards the left indicate the feature's positive and negative contribution to flood depth prediction, respectively. The features are ranked according to their level of contribution for each prediction.

Table 7
The computational cost for TUFLOW and surrogate models.

Models	CPU	RAM	GPU	Computational time
TUFLOW	4.4 GHz, 4cores	64 GB	2 NVIDIA(R) GeForce(R) Titan X each with 12 GB GDDR5X	4.5–6 hr
RF-P	3.6 GHz, 4 cores	16 GB	–	Train: 30 min Test: 4.5 ± 2 s
RF-P&T	3.6 GHz, 4 cores	16 GB	–	Train: 56 min Test: 4.2 ± 1.5 s

model execution on CPUs would be > 120 hr per event (Morsy et al., 2018). TUFLOW Heavily Parallelized Compute (HPC) engine allowed simulation on two GPUs and reduced execution time to 4.5–6 hr per event. Training time for RF-P trained with 16 events containing 375 hr was 30 min, which increased to 56 min for RF-P&T due to increased training samples (627 hr). However, the average time predicting each event was 4.5 and 4.2 s for RF-P and RF-P&T, respectively. Thus, RF-P&T could make predictions 3800 times faster than TUFLOW with reasonable flood depth and extent prediction errors.

5. Conclusion

This study resulted in RF surrogate models for real-time, street-scale forecasting of both pluvial and tidal flooding using the coastal city of Norfolk, VA, as a study region. Responses from a physics-based model, TUFLOW, were approximated by learning from inputs containing topographic variables for roadways and environmental variables for flood events and the corresponding target variable, TUFLOW-simulated water depths. Two surrogate models were trained with (i) only pluvial (RF-P) and (ii) both pluvial and tidal (RF-P&T) flood events. Comparison between these approaches showed that training on pluvial events with the tide as an input inadequately predicted the depth and extent of inundation during high tide or combined rainfall and high tide occurrence in coastal cities. The representation of both pluvial and tidal flood events in training significantly improved the predictive performance on test events by predicting 67%− 95% of the TUFLOW-simulated flood extent with average RMSEs of water depth ranging within 0.029–0.054 m. The predictive capability of the surrogate model for locations not in training was not explored in this study. To apply this model in a different urban area may require retraining using corresponding topographic data and physics-based model outputs. Future work should explore the transferability of this ML modeling approach to other urban areas.

Analyzing the locations of Waze flood reports showed that 87.5% and 90.4% of the reports had > 0.10 m water depth within the reporter's visibility from TUFLOW-simulated and RF-P&T predicted flooding, respectively. The distance between the reports and simulated flooding ranged within 0–150 m. 68% of the reports were within 1 hr from the time of peak depth. The agreement between time and location of flood reports with TUFLOW-simulated and RF-P&T predicted flooding served as validation for both models. The high agreement between the models and the flood reports suggested that the surrogate model could identify problematic flood spots on the roadway network in both space and time to support real-time decisions. For the locations with disagreement between flood reports and modeled results, further work is needed to determine if this was due to errors in the crowdsourced data or errors with the model's predictive skill. If the latter, it would be interesting and worth determining if the error is in the physics-based model itself or the ML surrogate model.

The RF algorithm can facilitate real-time street flood forecasting due to its short runtime, while maintaining an acceptable loss of accuracy compared to the high-fidelity physics-based model, TUFLOW. RF-P&T required 56 mins for training and 4.2 s on average for predicting each test event, increasing the computational speed by a factor of 3800 compared to TUFLOW. Additionally, RF provides global feature importance for the whole dataset, which was analyzed to understand the contribution of input features in flood prediction. The least important features from RF-P&T were surprisingly found to be hourly rainfall and tide level, which are the main physical drivers of flooding in a coastal city. Analyzing local feature importance using LIME revealed that these two features were significant in flood depth estimation with either positive or negative contribution levels based on the type of flood events. Thus, while the global feature importance provides insight into the RF model's general behavior, it can be biased towards the training data and affected by the correlation between inputs and target variable. Therefore, to interpret any black-box RF model for flood prediction or to perform feature selection, the local contribution of features should be investigated in future studies.

CRediT authorship contribution statement

Faria Tuz Zahura: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft preparation, Writing – review & editing, Visualization, Investigation, Formal analysis. **Jonathan L. Goodall:** Supervision, Writing – original draft preparation, Writing – review & editing, Funding acquisition.

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.

Data Availability

The script for Random Forest model and corresponding inputs and outputs are available on HydroShare (Zahura, 2021).

Acknowledgements

The work was funded by the National Science Foundation, United States under Award Number CBET-1735587. We acknowledge BMT for the TUFLOW HPC license and Waze for providing live traffic data.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2022.101087.

References

Agranoff, D., Fernandez-Reyes, D., Papadopoulos, M.C., Rojas, S.A., Herbster, M., Loosemore, A., Tarelli, E., Sheldon, J., Schwenk, A., Pollok, R., Rayner, C.F., Krishna, S., 2006. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. Lancet 368, 1012–1021. https://doi.org/10.1016/S0140-6736(06)69342-2.

Ahmad, M.W., Mourshed, M., Rezgui, Y., 2017. Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 147, 77–89. https://doi.org/10.1016/J.ENBUILD.2017.04.038.

Archer, K.J., Kimes, R.V., 2008. Empirical characterization of random forest variable importance measures. Comput. Stat. Data Anal. 52, 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015.

Atkinson, L.P., Ezer, T., Smith, E., 2013. Sea level rise and flooding risk in Virginia. Sea Grant Law Policy J.

AusRoads, 2008. Guide to Road Design, Part 5: Drainage Design.

Bartos, M., Kerkez, B., 2021. Pipedream: An interactive digital twin model for natural and urban drainage systems. Environ. Model. Softw. 144, 105120 https://doi.org/10.1016/J.ENVSOFT.2021.105120.

Bass, B., Bedient, P., 2018. Surrogate modeling of joint flood risk across coastal watersheds. J. Hydrol. 558, 159–173. https://doi.org/10.1016/j.jhydrol.2018.01.014.

Berkhahn, S., Fuchs, L., Neuweiler, I., 2019. An ensemble neural network model for real-time prediction of urban floods. J. Hydrol. 575, 743–754. https://doi.org/10.1016/j.jhydrol.2019.05.066.

Bermúdez, M., Cea, L., Puertas, J., 2019. A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. J. Flood Risk Manag. 1–14. https://doi.org/10.1111/jfr3.12522.

Bermúdez, M., Ntegeka, V., Wolfs, V., 2018. Development and Comparison of Two Fast Surrogate Models for Urban Pluvial Flood Simulations 2801–2815. ⟨https://doi.org/10.1007/s11269–018-1959–8⟩.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69. https://doi.org/10.1080/02626667909491834.

BMT WBM, 2016. TUFLOW User Manual.

Boutsis, I., Kalogeraki, V., Guno, D., 2016. Reliable crowdsourced event detection in smartcities, in: 2016 1st International Workshop on Science of Smart City Operations and Platforms Engineering (SCOPE) in Partnership with Global City Teams Challenge (GCTC), SCOPE - GCTC 2016. Institute of Electrical and Electronics Engineers Inc. ⟨https://doi.org/10.1109/SCOPE.2016.7515060⟩.

Box, G.E.P., Wilson, K.B., 1951. On the experimental attainment of optimum conditions. J. R. Stat. Soc. Ser. B 13, 1–38. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x.

Breiman, L., 2001. Random forests. Mach. Learn 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification an Regression Trees. Wadsworth, Belmont, CA.

Burgos, A.G., Hamlington, B.D., Thompson, P.R., Ray, R.D., 2018. Future nuisance flooding in Norfolk, VA, from astronomical tides and annual to decadal internal climate variability, 432-12,439 Geophys. Res. Lett. 45, 12. https://doi.org/10.1029/2018GL079572.

Chang, L.C., Shen, H.Y., Wang, Y.F., Huang, J.Y., Lin, Y.T., 2010. Clustering-based hybrid inundation model for forecasting flood inundation depths. J. Hydrol. 385, 257–268. https://doi.org/10.1016/j.jhydrol.2010.02.028.

Church, J.A., Clark, P.U., Cazenave, A., Gregory, J.M., Jevrejeva, S., Levermann, A., Merrifield, M.A., Milne, G.A., Nerem, R.S., Nunn, P.D., Payne, A.J., Pfeffer, W.T., Stammer, D., Unnikrishnan, A.S., 2013. Sea level change. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A.Y., Xia, V.B., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 493–499. https://doi.org/10.1016/B978-0-12-409548-9.10820-6.

City of Norfolk GIS Bureau, 2018. Street Centerline | Norfolk Open GIS Data [WWW Document]. URL ⟨https://norfolkgisdata-orf.opendata.arcgis.com/datasets/street-centerline⟩ (accessed 10.1.19).

Contreras, M.T., Gironás, J., Escauriaza, C., 2020. Forecasting flood hazards in real-time: a surrogate model for hydrometeorological events in an Andean watershed. Nat. Hazards Earth Syst. Sci. 1–24. https://doi.org/10.5194/nhess-2019-384.

Dahl, K.A., Fitzpatrick, M.F., Spanger-Siegfried, E., 2017. Sea level rise drives increased tidal flooding frequency at tide gauges along the U.S. East and Gulf Coasts: projections for 2030 and 2045. PLoS One 12, e0170949. https://doi.org/10.1371/journal.pone.0170949.

Deltares, 2018. Hydrodynamics, Rainfall Runoff and Real Time Control. User Manual SOBEK.

Esri, 2020. ArcGIS Desktop | Desktop GIS Software Suite - Esri [WWW Document]. URL ⟨https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview⟩ (accessed 8.20.20).

Fan, Y., Ao, T., Yu, H., Huang, G., Li, X., 2017. A coupled 1D–2D hydrodynamic model for urban flood inundation. Adv. Meteorol. 2017. https://doi.org/10.1155/2017/2819308.

Fears, D., 2012. Built on sinking ground, Norfolk tries to hold back tide amid sea-level rise [WWW Document]. Washington Post. URL ⟨https://www.washingtonpost.com/national/health-science/built-on-sinking-ground-norfolk-tries-to-hold-back-tide-amid-sea-level-rise/2012/06/17/gJQADUsxjV_story.html⟩ (accessed 9.30.19).

Guo, K., Guan, M., Yu, D., 2020. Urban surface water flood modelling-a comprehensive review of current models and future challenges. Hydrol. Earth Syst. Sci. Discuss. https://doi.org/10.5194/hess-2020-655.

Hallegatte, S., Green, C., Nicholls, R.J., Corfee-Morlot, J., 2013. Future flood losses in major coastal cities. Nat. Clim. Change 3, 802–806. https://doi.org/10.1038/nclimate1979.

Henonin, J., Russo, B., Mark, O., Gourbesville, P., 2013. Real-time urban flood forecasting and modelling - a state of the art. J. Hydroinform. 15, 717–736. https://doi.org/10.2166/hydro.2013.132.

Hinkel, J., Lincke, D., Vafeidis, A.T., Perrette, M., Nicholls, R.J., Tol, R.S.J., Marzeion, B., Fettweis, X., Ionescu, C., Levermann, A., 2014. Coastal flood damage and adaptation costs under 21st century sea-level rise. Proc. Natl. Acad. Sci. U.S.A. 111, 3292–3297. https://doi.org/10.1073/pnas.1222469111.

Hino, M., Belanger, S.T., Field, C.B., Davies, A.R., Mach, K.J., 2019. High-tide flooding disrupts local economic activity. Sci. Adv. 5, 1–10. https://doi.org/10.1126/sciadv.aau2736.

Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., Rasmussen, D.J., Muir-Wood, R., Wilson, P., Oppenheimer, M., Larsen, K., Houser, T., 2017. Estimating economic damage from climate change in the United States. Science. https://doi.org/10.1126/science.aal4369.

Inan, C.A., Artigue, G., Kurtulus, B., Pistre, S., Johannet, A., 2021. IOP Conference Series: Earth and Environmental Science A Hydrological Digital Twin by Artificial Neural Networks for Flood Simulation in Gardon de Sainte-Croix Basin, France A Hydrological Digital Twin by Artificial Neural Networks for Flood Simulation in Gardon de Sainte-Croix Basin, France 906, 12112. ⟨https://doi.org/10.1088/1755–1315/906/1/012112⟩.

Jacobs, J.M., Cattaneo, L.R., Sweet, W., Mansfield, T., 2018. Recent and future outlooks for nuisance flooding impacts on roadways on the U.S. east coast. Transp. Res. Rec. 2672, 1–10. https://doi.org/10.1177/0361198118756366.

Jhong, B.C., Wang, J.H., Lin, G.F., 2017. An integrated two-stage support vector machine approach to forecast inundation maps during typhoons. J. Hydrol. 547, 236–252. https://doi.org/10.1016/j.jhydrol.2017.01.057.
Kourtis, I.M., Bellos, V., Tsihrintzis, V.A., 2017. Comparison of ^{1}D–^{1}D and ^{1}D–^{2}D urban flood models, in: 15th International Conference on Environmental Science and Technology. pp. 1–6.
Lhomme, J., Bouvier, C., Mignot, E., Paquier, A., 2006. One-dimensional GIS-based model compared to two-dimensional model in urban floods simulation. ⟨https://doi.org/10.2166/wst.2006.594⟩.
Lian, J.J., Xu, K., Ma, C., 2015. Joint impact of rainfall and tidal level on flood risk in a coastal city with a complex river network: a case study of Fuzhou City, China. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-17-679-2013.
Lin, B., Wicks, J.M., Falconer, R.A., Adams, K., 2006. Integrating 1D and 2D hydrodynamic models for flood simulation. Proc. Inst. Civ. Eng. - Water Manag. 159, 19–25. https://doi.org/10.1680/wama.2006.159.1.19.
Liu, Y., Pender, G., 2015. A flood inundation modelling using v-support vector machine regression model. Eng. Appl. Artif. Intell. 46, 223–231. https://doi.org/10.1016/j.engappai.2015.09.014.
Löwe, R., Böhm, J., Jensen, D.G., Leandro, J., Rasmussen, S.H., 2021. U-FLOOD – Topographic deep learning for predicting urban pluvial flood water depth. J. Hydrol. 603, 126898 https://doi.org/10.1016/J.JHYDROL.2021.126898.
Lu, Q., Xie, X., Heaton, J., Parlikad, A.K., Schooling, J., 2019. From BIM towards digital twin: strategy and future development for smart asset management. Stud. Comput. Intell. 853, 392–404. https://doi.org/10.1007/978-3-030-27477-1_30.
Moftakhari, H.R., AghaKouchak, A., Sanders, B.F., Allaire, M., Matthew, R.A., 2018. What is nuisance flooding? Defining and monitoring an emerging challenge. Water Resour. Res. 54, 4218–4227. https://doi.org/10.1029/2018WR022828.
Moftakhari, H.R., AghaKouchak, A., Sanders, B.F., Feldman, D.L., Sweet, W., Matthew, R.A., Luke, A., 2015. Increased nuisance flooding due to sea-level rise: past and future. Geophys. Res. Lett. 42, 9846–9852. https://doi.org/10.1002/2015GL066072.
Morsy, M.M., Goodall, J.L., O'Neil, G.L., Sadler, J.M., Voce, D., Hassan, G., Huxley, C., 2018. A cloud-based flood warning system for forecasting impacts to transportation infrastructure systems. Environ. Model. Softw. 107, 231–244. https://doi.org/10.1016/j.envsoft.2018.05.007.
Murphy, P., Ogilvie, J., Connor, K., Arp, P., 2007. Mapping wetlands: a comparison of two different approaches for New Brunswick, Canada. Wetlands. https://doi.org/10.1672/0277-5212(2007)27[846:MWACOT]2.0.CO;2.
NOAA, 2018a. Daily Summaries Station Details: NORFOLK NAS, VA US, GHCND:USW00013750 | Climate Data Online (CDO) | National Climatic Data Center (NCDC) [WWW Document]. URL ⟨https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00013737/detail⟩ (accessed 10.6.19).
NOAA, 2018b. Sewells Point - Station Home Page - NOAA Tides & Currents [WWW Document]. URL ⟨https://tidesandcurrents.noaa.gov/waterlevels.html?id=8638610⟩ (accessed 10.1.19).
NOAA National Weather Service, 2018. Turn Around Don't Drown® [WWW Document]. URL ⟨https://www.weather.gov/safety/flood-turn-around-dont-drown⟩ (accessed 4.1.21).
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Peter, P., Weiss, R., Doubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res.
Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011527.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should i trust you?" Explaining the predictions of any classifier, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 1135–1144. ⟨https://doi.org/10.1145/2939672.2939778⟩.
Saarela, M., Jauhiainen, S., 2021. Comparison of feature importance measures as explanations for classification models. SN Appl. Sci. 3. https://doi.org/10.1007/s42452-021-04148-9.
Sadler, J., Goodall, J., Behl, M., Bowes, B., Morsy, M., 2020. Exploring real-time control of stormwater systems for mitigating flood risk due to sea level rise. J. Hydrology 583, 124571. https://doi.org/10.1016/J.JHYDROL.2020.124571.
Schleich, B., Anwer, N., Mathieu, L., Wartzack, S., 2017. Shaping the digital twin for design and production engineering. CIRP Ann. 66, 141–144. https://doi.org/10.1016/J.CIRP.2017.04.040.
Scikit-learn Developers, 2018. 3.2.4.3.2. sklearn.ensemble.RandomForestRegressor — scikit-learn 0.21.3 documentation [WWW Document]. URL ⟨https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html⟩ (accessed 10.4.19).
Scikit-learn Developers, 2018. Model evaluation: quantifying the quality of predictions [WWW Document]. URL ⟨https://scikit-learn.org/stable/modules/model_evaluation.html⟩ (accessed 10.6.19).
Seyoum, S.D., Vojinovic, Z., Price, R.K., Weesakul, S., 2012. Coupled 1D and noninertia 2D flood inundation model for simulation of urban flooding. J. Hydraul. Eng. 138, 23–34. https://doi.org/10.1061/(asce)hy.1943-7900.0000485.
Shand, T.D., Cox, R.J., Blacka, M.J., Smith, G.P., 2011. Australian Rainfall & Runoff.
Shen, Y., Morsy, M.M., Huxley, C., Tahvildari, N., Goodall, J.L., 2019. Flood risk assessment and increased resilience for coastal urban watersheds under the combined impact of storm tide and heavy rainfall. J. Hydrol. 579, 124159 https://doi.org/10.1016/J.JHYDROL.2019.124159.
Simpson, T.W., Peplinski, J.D., Koch, P.N., Allen, J.K., 2001. Metamodels for computer-based engineering design: Survey and recommendations. Eng. Comput. 17, 129–150. https://doi.org/10.1007/PL00007198.
Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. ⟨https://doi.org/10.1186/1471–2105-8–25⟩.
Sweet, W.V., Dusek, G., Obeysekera, J., Marra, J.J., 2018. Patterns and Projections of High Tide Flooding Along the U.S. Coastline Using a Common Impact Threshold. Silver Spring, Maryland.
Tanaka, S., Bunya, S., Westerink, J.J., Dawson, C., Luettich, R.A., 2011. Scalability of an unstructured grid continuous Galerkin based hurricane storm surge model. J. Sci. Comput. 46, 329–358. https://doi.org/10.1007/s10915-010-9402-1.
US Department of Transportation, 2014. Mitigation Strategies For Design Exceptions - Safety | Federal Highway Administration [WWW Document]. URL ⟨https://safety.fhwa.dot.gov/geometric/pubs/mitigationstrategies/chapter3/3_lanewidth.cfm⟩ (accessed 10.5.19).
USGCRP, 2017. Climate Science Special Report: Fourth National Climate Assessment, Volume I. U.S. Global Change Research Program, Washington, DC, USA. ⟨https://doi.org/10.7930/J0J964J6⟩.
USGS, 2016. TNM Download [WWW Document]. Natl. Map. URL ⟨https://viewer.nationalmap.gov/basic/⟩ (accessed 10.1.19).
USGS, 2013. LidarExplorer [WWW Document]. URL ⟨https://prd-tnm.s3.amazonaws.com/LidarExplorer/index.html#⟩/ (accessed 2.15.21).
Yin, J., Yu, D., Yin, Z., Liu, M., He, Q., 2016. Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China. J. Hydrol. 537, 138–145. https://doi.org/10.1016/j.jhydrol.2016.03.037.
Zahura, F., 2021. Inputs for Random Forest model (pluvial and tidal events) | CUAHSI HydroShare [WWW Document]. URL ⟨https://www.hydroshare.org/resource/d25a394aa0644e51a0b0bfab5ab2dbf0/⟩ (accessed 10.16.21).
Zahura, F.T., Goodall, J.L., Sadler, J.M., Shen, Y., Morsy, M.M., Behl, M., 2020. Training machine learning surrogate models from a high-fidelity physics-based model: application for real-time street-scale flood prediction in an urban coastal community. Water Resour. Res 56. https://doi.org/10.1029/2019WR027038.
Zhang, J., Taflanidis, A.A., Nadal-Caraballo, N.C., Melby, J.A., Diop, F., 2018. Advances in surrogate modeling for storm surge prediction: storm selection and addressing characteristics related to climate change. Nat. Hazard. 94, 1225–1253. https://doi.org/10.1007/s11069-018-3470-1.