

Identification of Urban Rainstorm Waterlogging Based on Multi-source Information Fusion: A Case Study in Futian District, Shenzhen

Zongjia Zhang^{1,2}, Fanyu Meng^{2,3}, Yiping Zeng², Junguo Liu⁴, Diping Yuan⁵, Simon Fong⁶, and Lili Yang^{2,*}

¹School of Environment, Harbin Institute of Technology, 150001 Harbin, China

²Department of Statistics and Data Science, Southern University of Science and Technology, 518055 Shenzhen, China

³Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, 518055 Shenzhen, China

⁴School of Environmental Science and Engineering, Southern University of Science and Technology, 518055 Shenzhen, China

⁵Shenzhen Urban Public Safety and Technology Institute, 518048 Shenzhen, China

⁶Department of Computer and Information Science, University of Macau, 999078 Taipa, Macau, China

Abstract. Flood disasters have become one of the most threatening natural disasters in the world, in which waterlogging is the most common form in the context of highly urbanized megacities. The formation of flood disaster is related to many factors and involves information from multiple sources, making it difficult to be predicted. This paper integrates multi-source information data, classifies the study area into different categories according to hydrological analysis results, and combines hydrodynamic theory and ArcGIS to get the quantitative prediction of the range and depth of waterlogging under different rainfall inputs. The evaluation results provide the government with accurate and timely information of waterlogging risks and locations in order to improve promptness of emergency management such as evacuation and managing traffics.

1 Introduction

1.1 Background

With the fast pace of urbanization, cities have been expanding rapidly with boosting population density and rate of land development. Flood disasters have become one of the most threatening natural disasters in the world given their high frequency, wide coverage and significant destructiveness. As a typical type of flood disaster, urban waterlogging, commonly caused by heavy rain, has become an important factor that seriously threatens urban safety. In developed cities, rainfall accumulates in a large number of low-lying areas formed in the process of urbanization. Insufficient drainage capacity exacerbates the risk of waterlogging [1].

When encountering some extreme rainstorm or typhoon events, the intensity of the rainstorm far exceeds the designed drainage capacity, and the surface water cannot be discharged in time, resulting in surface runoff and stagnant water [2]. The identification of urban waterlogging and inundation risk can help predict the distribution range, depth and flow rate of accumulated water under different rainstorm intensities. Besides, visualization technologies are desired to be applied to generate rainstorm risk maps providing decision-making suggestions for predicting urban waterlogging under extreme rainstorms [3]. Therefore, for the vulnerable areas, early warning information should be issued to the citizens in advance to help urban decision-makers

formulate targeted traffic control measures and emergency control plans.

For megacities with large population, numerous buildings, and complex municipal engineering systems, especially the coastal ones that are directly exposed to typhoon disasters, risk of urban waterlogging is much higher [4], yet not much research has focused on such scenarios. The current research combines SWMM, MIKE FLOOD and other models to simulate an area with detailed drainage information and analyze the risk of urban waterlogging. The Futian District of Shenzhen located in southern China is selected as a typical highly urbanized area in a global megacity, upon which waterlogging risk analyses are conducted based on multiple rainfall input scenarios with practical policy implications suggested.

1.2 Literature review

Statistical methods with advanced algorithms have been applied to identify vulnerable areas of floods. Chau et al. used K-nearest neighbors techniques to locate inundated areas in 2005 [5]. Hong et al. used statistical methods to draw flood-sensitive maps in 2016 [6]. Tehrany et al. carried out multivariate statistical method to identify the submerged area in 2014 and 2017, respectively, such as genetic programming, artificial neural network, fuzzy logic model, etc. [7] D'Addabbo et al. used statistical model based on static Bayesian networks to detect flood in 2016.

* Corresponding author: yangll@sustech.edu.cn

Several scholars also use hydrodynamic models to simulate flood routing. Samela et al. applied the idea of river hydrology to flood routing simulation in 2015. Chen developed a two-dimensional coupled model based on SWMM and SWM to analyze rainstorm and waterlogging in 2019 [8]. However, hydrodynamic models require highly precise data, among which detailed urban underlying surface data is generally necessary. Although relatively fine simulation results can be obtained, the comprehensive and large-scale calculations are commonly required. Therefore, it is difficult to be applied to large-scale urban flood risk identification.

In recent years, many scholars have used machine learning methods to identify inundation areas based on historical inundation data. Lamovec et al. used machine learning methods such as decision tree (DT) and random forest (RF) to detect flood-prone areas in 2013 [9]. Tehrany et al. tested the computational efficiency of DT in this application. Some scholars also applied integrated algorithms such as adaptive-network-based fuzzy inference system (ANFIS) and genetic algorithm based artificial neural network to identify waterlogging risk areas [10]. Ke et al. applied machine learning approaches to simulate urban flooding in the Shenzhen city [2].

With the improvement of the accuracy of remote sensing images and digital elevation models (DEM), the submergence information extraction model based on remote sensing images is used to identify the submergence risk. This method compares the remote sensing images before and after the occurrence of flood disaster in a certain area to directly extract the submerged range and depth of water. Combined with meteorological data, it can provide reference for risk identification and early warning. However, due to the lack of simulation of the flood process and the scarcity of remote sensing satellite resources, it is difficult for them to be widely used in a short time [11].

As the process of formation of waterlogging is complicated, the above studies have considered many factors that may lead to urban waterlogging. Generally, factors such as DEM, aspect and slope, rainfall intensity, land cover type and drainage capacity will be considered. However, some of these studies are still in the qualitative risk analysis stage and can only assess the risk of waterlogging in certain areas with only part of the crucial factors considered [11]. Fusion of multi-source information that integrates meteorological, geographic, and municipal engineering information is of great need. Combining statistics and hydrodynamic models is capable of quantifying certain details of the process and more precisely identifying the coverage and depth of the waterlogging area, which is more conducive to the government to judge the extent of disaster damage, classify risk levels and take early warnings.

2 Study area and data

The study area is the Futian District, which is the central district of the Shenzhen city located in the Guangdong Province, China. The total area of the Futian District is

78.8 square kilometers. The terrain is mainly composed of plains, hills, mountains, and beaches. There are mountains in the north and the sea in the south. Located in the south of the Tropic of Cancer, Futian District in Shenzhen is a subtropical maritime climate zone with sufficient rainfall, with an average annual rainfall of 1,866 mm. The rainfall is mainly concentrated in June and August. In August, there are frequent typhoons and heavy rain extreme weather. Futian District has a high level of urban development and high economic density. In 2020, the permanent population will exceed 1.66 million. Although the area only accounts for 4% of the city's total area, its GDP accounts for 16.9% (454.6 billion Yuan) of the city's GDP (2692.7 billion Yuan).

Table 1 shows the factors related to the formation of urban waterlogging disasters for this study, including the data sets corresponding to each factor and the scale and precision of each data set.

Table 1. Data description.

Item	Data description	Scale/Precision
DEM	Realize digital simulation of ground terrain through limited terrain elevation data	5m*5m
Land cover type	Current status of all land use in the city, including construction land, broad-leaved forest land, coniferous forest land, water bodies, wetlands, etc.	30m*30m
Building height	Building vector files containing building location, height, number of floors, floor space, etc.	0.01m
Drainage system	Rainwater outlet vector file, including location, orifice size, orifice shape	0.001m
Runoff coefficient	The proportion of precipitation converted into surface runoff	0.1
Rainfall intensity	Rainfall per unit time	0.01mm/min

3 Methods

3.1 Influential factors

The formation of urban rainstorm waterlogging is related to many factors, mainly containing three categories: positive correlation factors, distribution influencing factors and negative correlation factors [12].

3.1.1 Positive correlation factors

Incremental factors, including rainfall intensity and rainfall duration, are positively correlated with the scale of waterlogging. The return period of a specific rainstorm intensity refers to the average interval time between the occurrence of a rainstorm intensity greater

than or equal to this value, and it is inversely proportional to the frequency. As the rainfall intensity and rainfall duration increase, the total rainfall increases.

3.1.2 Distribution influencing factors

Such factors are mainly related to topography, landforms and surface structures. Rain causes surface runoff to continue to increase, and overflowing surface runoff flows horizontally from high terrain to low terrain areas under the influence of gravity. The continuously increasing surface runoff cannot be discharged in time by the drainage system, thus forming stagnant water in low-lying areas. In terms of rainfall process, the range and depth of stagnant water generally increase continuously.

3.1.3 Negative correlation factors

Certain factors have a negative correlation with the formation of waterlogging, including infiltration capacity, drainage capacity, and evaporation.

This paper uses multi-source data fusion technology to divide the research area into 596 irregular sub-catchments based on geographic characteristics and confluence processes. Each sub-catchment contains all the attributes of factors related to the risk of waterlogging. Taking the sub-catchment area as the basic calculation unit, combined with the Manning equation in hydrodynamics, Horton infiltration curve and other formulas to calculate the depth of water accumulation and flow velocity, the calculation results with higher accuracy can be obtained. Use the analysis and visualization functions of ArcGIS to draw a waterlogging risk map that includes the range and depth of stagnant water.

3.2 Rainfall model

The Futian District of Shenzhen often suffers short-term heavy rainfall brought by typhoon disasters in August. In order to simulate a variety of extreme rainstorm that the area may suffer, based on the rainstorm intensity Equation (1) published by the Shenzhen Meteorological Bureau in 2015, the return period is set to 5, 30, and 100 years, and the peak rainfall is 0.4. By using Chicago rainfall process model, three simulated rainfall input models are obtained, as shown in Fig. 1.

$$q = \frac{167A(1 + Clga)}{(t + b)^n} \quad (1)$$

where q is the rainfall intensity; a is the return period of heavy rain; t is the rainfall duration; A, C, b, n are the rain force formula parameters.

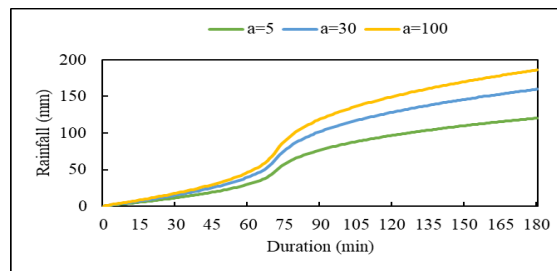


Fig. 1. Chicago rain patterns of 5-year, 30-year, 100-year return period.

3.3 DEM with building height

The original DEM can only display surface elevation data, not including buildings on the ground. But in the process of flood routing, water flow will be blocked by buildings and change its flow direction, which will eventually affect the spatial distribution of stagnant water. The Futian District has a high density of buildings, especially high-rise buildings. In order to simulate the blocking effect of buildings on water flow, the vector data containing building height information is superimposed with the original DEM using ArcGIS raster calculator. Fig. 2 is the new DEM coupled with building height information.

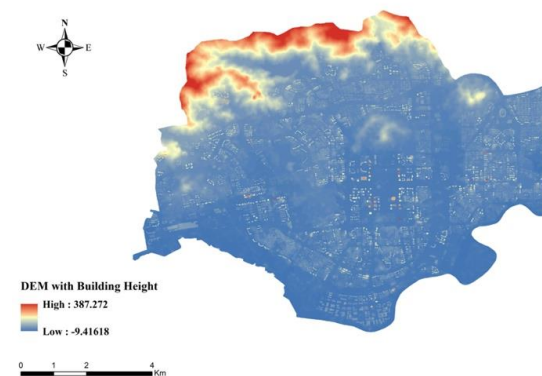


Fig. 2. DEM with building height (min = -9.14 m, max = 387.27 m).

3.4 Analysis of water flow process

The water flow process is mainly based on the D8 algorithm. The principle is to assume that the water flow in a single grid can only flow into the lowest one among the 8 adjacent grids (Fig. 3). It uses the steepest slope method to determine the direction of the water flow (Fig. 4). It is characterized by fast calculation speed and can well reflect the effect of topography on the formation of surface runoff. The D8 algorithm can be used on the DEM data that has been filled to obtain the unique flow direction value of each grid. The hydrological analysis tool of ArcGIS can calculate the flow value of all the grids in the Futian area and further calculate the flow accumulation (Fig. 5, Fig. 6). This value represents the number of grids that the target grid receives upstream confluence. The flow accumulation cannot represent the

final flow result, but it provides theoretical support for runoff movement and sub-catchment division [13].

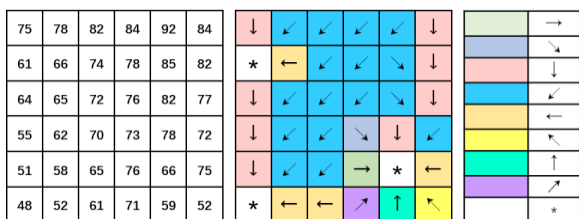


Fig. 3. Schematic diagram of D8 algorithm.

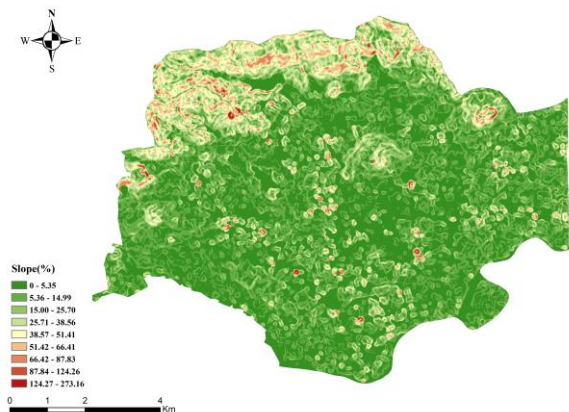


Fig. 4. Topographic slope of Futian District.

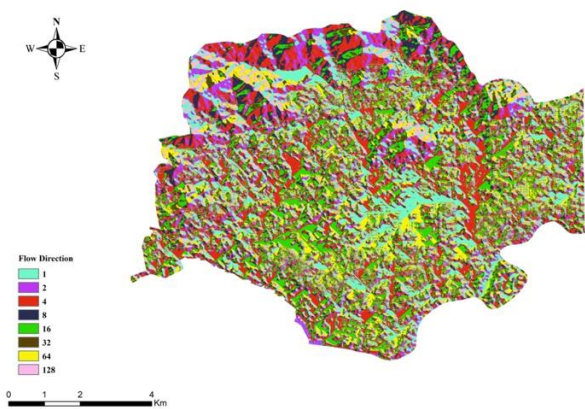


Fig. 5. Flow direction.

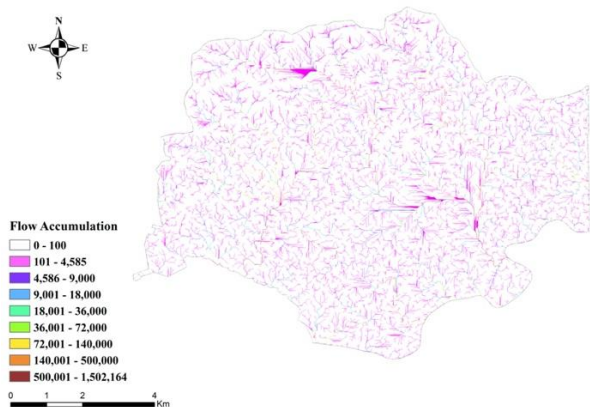


Fig. 6. Flow accumulation.

3.5 Sub-catchment division

In the urban-scale waterlogging prediction research, in order to improve the prediction accuracy and calculation efficiency of the model, the research area can be divided into tiny sub-catchments based on the results of hydrological analysis. When dividing the sub-catchment area, we define the lower limit of the catchment area that can generate surface runoff as 40,000 square meters, which is equivalent to a rectangular area of 200m × 200m. Confluence areas smaller than the lower limit are no longer divided separately. It is unified into other confluence areas. The basis for this division is that on one hand, it is difficult to form large-scale stagnant water in an area generally smaller than 200m × 200m in reality; on the other hand, it is to improve the calculation efficiency of the model. In the study, the catchment area of 40,000 square meters corresponds to the cumulative value of 1600. Fig. 7 further classifies and links the cumulative flow. Fig. 8 shows the 596 sub-catchments.

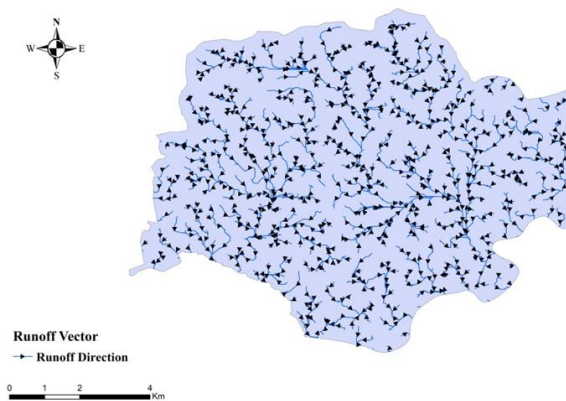


Fig. 7. River link and runoff vector.

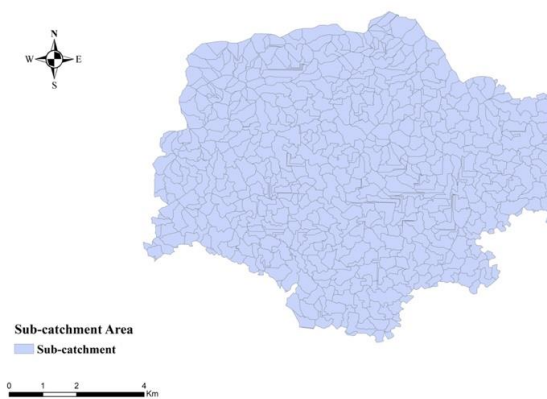


Fig. 8. Total 596 sub-catchments.

3.6 Surface runoff coefficient

The Futian District is the central district of Shenzhen, with 5909 hectares of construction land, accounting for 74.9%. Urban construction land is mostly asphalt, concrete, and cement surfaces, which have high surface runoff coefficients and weak rainfall infiltration capacity. The runoff coefficient of woodland, grassland and wetland is small, and the ability to accumulate rainfall is

strong. According to the land cover type data in the Futian District (Fig. 9), the surface runoff coefficient in this area is 0.95 for construction land, 0.65 for broad-leaved forest and grassland, 0.65 for coniferous forest, 0.7 for water bodies, and 0.75 for wetland and irrigated fields [14]. In view of the actual situation of land use in Futian District, this study extracts the proportion of each land type, and obtains the average runoff coefficient of each sub-catchment through weighted calculation (Fig. 10).

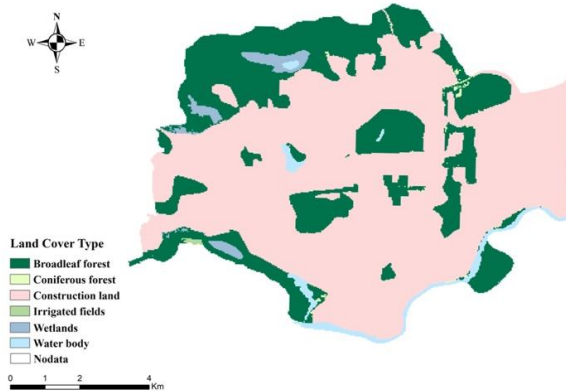


Fig. 9. Land cover type.

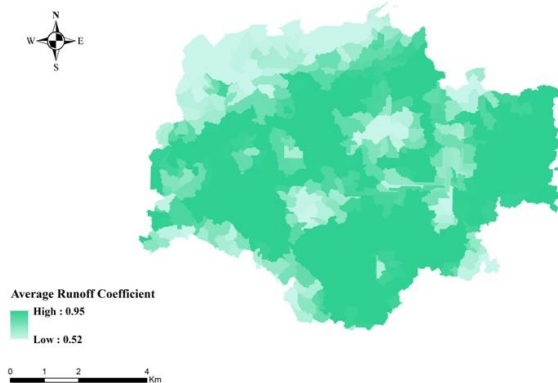


Fig. 10. Average runoff coefficient of each sub-catchment.

3.7 Evaluation of drainage capacity

By analyzing the distribution of 37,845 rainwater outlets in the Futian District, a generalized evaluation model of drainage capacity was established. Almost all rainwater discharge outlets in the Futian District are flat rectangular discharge outlets. According to the number, shape and size of the drainage outlets in each sub-catchment area, the drainage capacity per unit time can be expressed as Equation (2). Combined with rainfall duration, the total drainage volume of each sub-catchment area can be calculated [15].

$$Q_d = \mu s \sqrt{2gh} \quad (2)$$

where Q_d is the drainage flow (m^3/s); μ is orifice flow coefficient; s is drainage area (m^2); h is the water depth (m).

3.8 Depth of waterlogging

Different types of data from multiple information sources are merged with sub-catchments as storage units. Each sub-catchment contains DEM with building height, average slope gradient, average surface runoff coefficient, catchment area, average roughness, rainfall, drainage intensity and other index attributes [12].

Although each sub-catchment area is divided according to an independent water catchment process, the flood routing has integral characteristics. When the Futian District is used as a unified research area, the water flow connection of each sub-catchment area needs to be considered. Because urban roads are generally low-lying, rainwater drains are on both sides of the road and the road network has strong connectivity. In the flooding routing, floods flow and spread through urban roads in the form of surface runoff. By analyzing the 685 main roads in Shenzhen, the city includes 390 arterial roads, 264 first-class highways, 17 expressways, and 14 highways. Because expressways generally have high roadbeds or an elevated form, waterlogging is not easy to occur. The average width of 671 arterial roads, first-class highways and expressways in the studied area is 23.3 meters. Assuming that the water flow between the upstream and the downstream sub-catchment areas is mainly connected by the lowest road and the road is simulated as a U-shaped channel, Equations (3) to (7) can be used to determine the flow, velocity, and depth of water accumulation [16].

$$Q_r = aq\varphi A \quad (3)$$

where Q_r is the surface runoff flow (m^3/s); q is the average rainfall intensity (mm/min); φ is the average runoff coefficient of the sub-catchment; A is the area of the sub-catchment (m^2); a is the unit conversion factor, which is equal to $1/60000$ in this case.

The surface runoff velocity, v , can be calculated as:

$$v = \frac{R^{2/3} i^{1/2}}{n} \quad (4)$$

where v is the surface runoff velocity (m/s); R is the hydraulic radius; i is the average slope gradient of the sub-catchment area (%); n is the surface roughness.

$$Q = SC(Ri)^{1/2} \quad (5)$$

where S is the cross-sectional area of the water (m^2), C is the Chézy coefficient.

The Chézy formula can be expressed as:

$$C = \frac{R^{1/6}}{n} \quad (6)$$

in which the hydraulic radius, R , is commonly expressed as:

$$R = \frac{S}{X} = \frac{dw}{2d + w} \quad (7)$$

where w is the width of the water cross section (m), that is the average width of the road; X is the wet perimeter of the water cross section (m); d is the depth of waterlogging (m). Waterlogging generally appears at the boundary of the sub-catchment area, because these areas are the lowest point of terrain and have a large flow accumulation value.

4 Results and discussions

After the range and depth of the waterlogging under each rainfall input scenario are calculated, the waterlogging results can be superimposed on the satellite image for display through ArcGIS. Fig. 11 shows the spatial distribution of waterlogging with depths. The darker color represents the deeper waterlogging. The depth of the waterlogging (Table 2) ranges from 0 meters to 0.189 meters for the 5-year-return period, from 0 meters to 0.225 meters for the 30-year-return period, from 0 meters to 0.246 meters for 100-year-return period. Comparing the three rainfall models, as the rainfall intensity increases, the area and depth of waterlogging increase, and the risk of waterlogging increases. There are 159, 175, 186 catchments with waterlogging for three different return period of heavy rain. Average depths of waterlogging are 6.94cm, 7.78cm and 8.35cm, respectively. In Fig. 11, the distribution of waterlogging is mainly concentrated in low-lying areas, almost no water in high altitude areas. At the foot of the mountain, because of the insufficient drainage facilities in the mountainous area, the rainfall flows down the slope and gathers in the low-lying areas at the foot of the mountain. The drainage facilities in the urban area are very complete, and the accumulated water can be discharged timely. The result of the waterlogging in the southwest and southeast is that these two areas are new land formed by reclamation and lack drainage data in these areas. The municipal department should strengthen the construction of drainage facilities in waterlogged areas especially the neighboring mountain areas and pump water in the waterlogged areas when suffering from heavy rain.

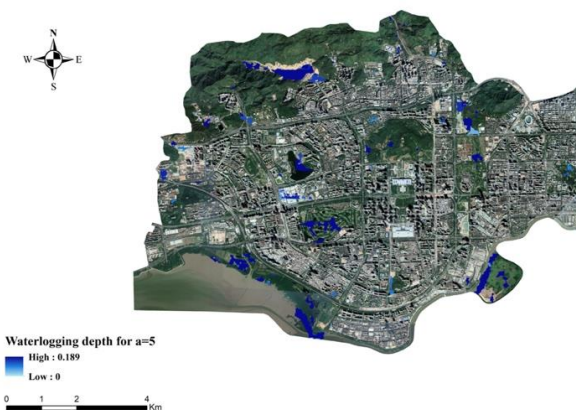


Fig. 11(a). Waterlogging for a = 5.

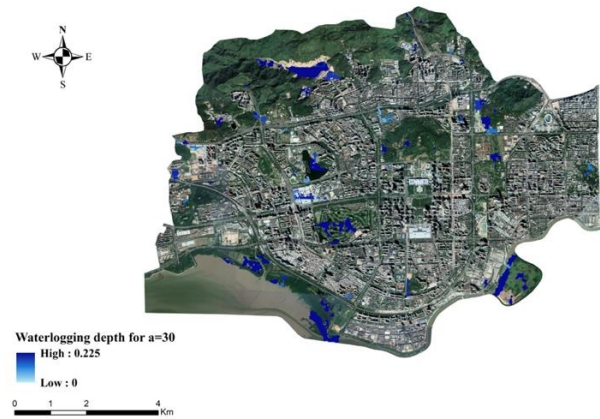


Fig. 11(b). Waterlogging for a = 30.

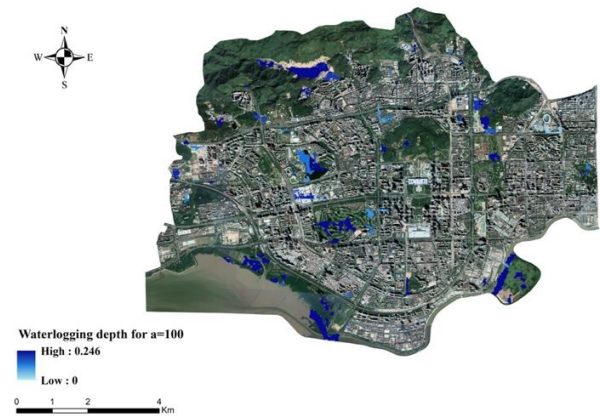


Fig. 11(c). Waterlogging for a = 100.

Table 2. Statistics of waterlogging depth in 596 catchments.

	a = 5	a = 30	a = 100
Number of catchments without waterlogging	437	421	410
Number of catchments with waterlogging	159	175	186
Average of waterlogging depth (cm)	6.94	7.87	8.35
Median of waterlogging depth (cm)	6.46	7.36	7.66
Variance of waterlogging depth	14.90	21.62	23.30

5 Conclusions

This paper uses multi-source information fusion technology to deeply integrate meteorological information, geographic information, and municipal engineering information. ArcGIS is used to realize the hydrological analysis process of horizontal flow after rainfall is transformed into surface runoff. Based on a tradeoff between the efficiency and accuracy of the calculations, 596 irregular sub-catchments were defined,

giving rise to a macroscopic analysis of urban rainstorm waterlogging. Comprehensive consideration of rainfall intensity, rainfall duration, DEM, building height, slope, aspect, current land use, and regional drainage capacity to perform urban flooding prediction and risk identification.

This research provides emergency management authorities with timely and quantitative decision-making information by reasonably identifying risky areas to release early warning signs and reduce casualties and property losses caused by urban waterlogging. For the traffic planning department, it is possible to predict the stagnant water area in advance and release information on the roads affected by disasters for citizens in real time. For the navigation service providers, the navigation algorithm can be improved to avoid the stagnant water section in advance and create a more reasonable and efficient route plan. It can also provide disaster assessment references for emergency dispatching departments, assess the accessibility of rescue services in important areas, optimize the layout of rescue stations and the rescue force dispatch plan.

This research was funded by National Key R&D Program of China (2018YFC0807000), Natural Science Foundation of China (71771113), and National Key R&D Program of China (2019YFC0810705). It was also partly supported by the High-level Special Funding of the Southern University of Science and Technology (grant nos. G02296302 and G02296402).

References

1. Du, J., Fang, J., Xu, W., & Shi, P. Analysis of dry/wet conditions using the standardized precipitation index and its potential usefulness for drought/flood monitoring in Hunan Province China. *Stochastic Environmental Research and Risk Assessment*, **27(2)**, 377–387 (2013)
2. Ke Q., Tian X., Bricker J., Tian Z., Guan G., Cai H., Huang X., Yang H., Liu J. Urban pluvial flooding prediction by machine learning approaches – a case study of Shenzhen city, China. *Advances in Water Resources*, **145**: 103719 (2020)
3. Elkhachy, & Ismail. Flash flood hazard mapping using satellite images and GIS tools: a case study of Najran city, Kingdom of Saudi Arabia (KSA). *Egyptian Journal of Remote Sensing & Space Sciences*, **18(2)**, 261-278 (2015)
4. Cai, T., Li, X., Ding, X., Wang, J., & Zhan, J. Flood risk assessment based on hydrodynamic model and fuzzy comprehensive evaluation with GIS technique. *International Journal of Disaster Risk Reduction*, **2019.101077** (2019)
5. Chau, K. W., Wu, C. L., & Li, Y. S. Comparison of several flood forecasting models in Yangtze river. *Journal of Hydrologic Engineering*, **10(6)**, 485–491 (2005)
6. Hong, H., Pradhan, B., Bui, T., Xu, C., Youssef, M., & Chen, W. Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China). *Geomatics, Natural Hazards and Risk*, **8(2)**, 544–569 (2016)
7. Tehrany, S., Pradhan, B., & Jebur, N. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, **512(2014)**, 332–343 (2014)
8. Wenjie C. Urban Flood Hydrological and Hydrodynamic Model Construction and Flood Management Key Issues Exploration. *Unpublished doctoral dissertation*, South China University of Technology, Guangzhou, China (2019)
9. Lamovec, P., Veljanovski, T., Mikos, M., & Ostir, K. Detecting flooded areas with machine learning techniques: Case study of the Selska Sora river flash flood in September 2007. *Journal of Applied Remote Sensing*, **7(1)**, 1–13 (2013)
10. Tehrany, S., Pradhan, B., & Jebur, N. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, **504(2013)**, 69–79 (2013)
11. Mukherjee F., Singh D. Detecting flood prone areas in Harris County: a GIS based analysis. *GeoJournal*. **85(3)**: 647-663 (2019)
12. Xitao H. Study on urban waterlogging vulnerability evaluation and disaster impact model. *Unpublished doctoral dissertation*, Xi'an University of Technology, Xi'an, China. (2020)
13. Abedin, Hossain, S. J., Stephen, & Haroon. GIS framework for spatiotemporal mapping of urban flooding. *Geosciences*. **9(2)** (2019)
14. Sen, Z., & Abdüsselam Altunkaynak. A comparative fuzzy logic approach to runoff coefficient and runoff estimation. *Hydrological Processes*, **20(9)**, 3991-3991 (2006)
15. Deepak, Singh, Bish, Chandranath, Chatterjee, & Shivani, et al. Modeling urban floods and drainage using SWMM and MIKE URBAN: a case study. *Natural Hazards*, **84(2)**, 749–776 (2016)
16. Tyrna, B., Assmann, A., Fritsch, K., & Johann, G. Large-scale high-resolution pluvial flood hazard mapping using the raster-based hydrodynamic two-dimensional model floodareaHPC. *Journal of Flood Risk Management*, **11(S2)** S1024-S1037 (2018)