Impact of Land Use/Land Cover Changes on Urban Flooding: A Case Study of the Greater Bay Area, China

Yuqi Lei , Hui Cao, Xuanyan Zhou , Jon Mills, and Wen Xiao , *Member, IEEE*

Abstract—Urban flooding caused by heavy rainfall is a common natural hazard in cities globally. Impervious surfaces are often increased during urban development, but there is limited research on the impact of large-scale and long-term land use/land cover (LULC) changes on urban flooding, while considering the influence of using different remote sensing data sources. In this study, a framework to evaluate the correlation between LULC changes and flooding extents is proposed, mainly comprising: 1) classifying remote sensing time series, using different sources but adopting the same classifier, to obtain the LULC of the Greater Bay Area, China, over a one decade period; 2) designing flooding scenarios with different rainfall intensities, and using the soil conservation service curve number (SCS-CN) model and local equal volume method to extract the inundation extent of urban flooding; and 3) analyzing the influence of different data sources on flood simulation results, and calculating the correlations between LULC and inundation area over the decade to analyze the impact of LULC changes on urban flooding. The resulting correlation coefficients of water and built-up land are 0.93 and 0.42, and those of bare land, grassland, orchard, and forest are -0.40, -0.61, -0.57, and -0.75, respectively. The inundation derived by Sentinel and Landsat data showed around 99% consistency, while Landsat tends to derive more inundation areas, with the differences mainly scattered in flat areas.

Index Terms—Guangdong–Hong kong–Macao greater bay area, land use/land cover (LULC) change, landsat, soil conservation service curve number (SCS-CN) model, Sentinel, urban flooding.

I. INTRODUCTION

URBAN flooding is a regional flooding phenomenon caused by excessive rainfall and low-lying terrain that cannot be drained in time when short-term heavy precipitation exceeds the drainage capacity [1], [2]. Under the global process of climate change and urban sprawl, the frequency and severity of urban flooding induced by extreme heavy rainfall are continuously increasing [3], [4], [5], [6]. One of the reasons is that land use/land cover (LULC) changes in cities affect the surface water flux, altering the runoff generation and convergence processes of watersheds, which may increase flood risk. The impact of LULC changes on urban flooding has gained extensive attention in the related research communities [7], [8], [9], [10].

LULC changes can be obtained through remote sensing observations. As remote sensing technology has evolved over the years, the availability of satellite data has become increasingly abundant [11]. Common optical satellite data sources, such as the Landsat series, have the advantage of having long time series imagery, while Sentinel-2 optical data provides higher spatial resolution. In addition, microwave satellite data sources, e.g., Sentinel-1, have the advantage of all-weather observation. At the same time, acquiring multitemporal multiplatform remote sensing images from the same area has become more convenient, which is conducive to multisource data fusion. Previous studies have demonstrated that the fusion of multiple remote sensing data sources for LULC classification, including optical images and synthetic aperture radar (SAR) data, can improve the classification accuracy based on pixel-based methods, and the classification results are superior to those using only multispectral or microwave data [12], [13], [14], [15], [16], [17]. Different remote sensing data sources have their own strengths and weaknesses, and it is necessary to evaluate the impact of using different data on LULC classification and subsequent flood modeling experiments.

Surface runoff is currently mainly simulated through various hydrological models. Commonly used hydrological models include soil conservation service curve number (SCS-CN), Mike, storm water management model (SWMM), InfoWorks ICM, and Wallingford models [18], [19], [20], among which the SCS-CN model requires few parameters. It is easy to operate and is commonly used to compute surface runoff in coastal areas [21], [22], [23], [24]. Many studies have used it to simulate the generation and convergence processes of flood disasters in small watersheds [25], [26], [27]. Liu and Li [28] used the SCS-CN model to calculate the runoff process and the model accuracy was above 75%. The results proved that the SCS-CN model could simulate the runoff process in a typical subwatershed on the Loess Plateau. The model was also combined with the local equal volume method [29] and proved to be suitable for flooding extent extraction in larger watersheds [30].

The Greater Bay Area, China, is a coastal region with expanding built-up land, increasing population, and industrial concentration [31]. The rapid urbanization there has led to an increased risk of urban flooding [20], [32], [33], [34], [35]. Prior studies of the Greater Bay Area have evaluated the effects of urbanization on precipitation [36], [37], and calculated the depth of surface runoff under extreme rainfall [38], [39], [40], but did not further simulate the inundation extent after a rainstorm, or explore the impact of LULC change on the inundation range. In addition, LULC classification is an important input for runoff calculation, whereas little research has studied the impact of classification results from different remote sensing data sources on urban flooding simulation.

In this study, multitemporal satellite imagery is used to obtain LULC changes in the Greater Bay Area over the past decade. With the help of the SCS-CN model, surface runoffs under various rainfall intensity scenarios are calculated, and the inundation extent of urban flooding is extracted. Then, the inundation extent of each administrative region and various LULC types are statistically analyzed over a decade to calculate the correlation coefficient between the extent of urban flooding and LULC changes. Additionally, Sentinel-1 and Sentinel-2 images are jointly used for LULC classification to compare with Landsat data, and urban flooding simulation is conducted on both data sources to explore the influence of different data sources and feature selection on classification and flood simulations.

The main contributions of this study are threefold: 1) obtain the LULC changes in the Greater Bay Area over a decade, and simulate regional scale urban flooding under four rainfall scenarios on this basis; 2) assess the influence of using different remote sensing data sources, Landsat and Sentinel, on urban flood simulation, revealing the consistency between the two in both flooded and unflooded areas; 3) quantitatively evaluate the correlation coefficient between each of the eight LULC changes and urban flooding extents for the first time, providing more detailed and accurate references on the impact of urban development on flooding. The results reported in this article provide scientific support for the Greater Bay Area to improve its capacity to mitigate urban flooding.

The rest of this article is organized as follows. Section II of this article details the study area and the data utilized. In Section III, the experimental procedures will be followed to introduce preprocessing, LULC classification, runoff modeling, flooding extent estimation, and statistical methods. In Section IV, it will sequentially present the interannual changes in LULC, the analysis of flood simulation, the impact of LULC changes on urban flooding, and the effects of different data sources on the experimental results. In Section V, the deficiencies of the experiment will be summarized, and future prospects will be outlined. Finally, Section VI concludes this article.

II. STUDY SITE AND DATA

A. Study Site

As shown in Fig. 1, the Guangdong-Hong Kong-Macao Greater Bay Area consists of the cities of Guangzhou, Zhuhai, Shenzhen, Foshan, Dongguan, Huizhou, Jiangmen, Zhongshan,
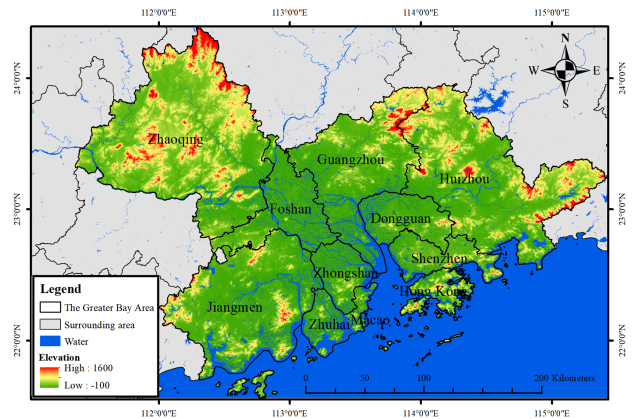

Fig. 1. Elevation map of the Greater Bay Area, China.

and Zhaoqing in Guangdong Province, and Hong Kong and Macao Special Administrative Regions, China. In 2022, the Greater Bay Area had a population of 86.62 million and a gross domestic product (GDP) of 1943.54 billion US dollars. With its dense population and developed economy, urban flooding hazards in the region can cause significant losses. Unfortunately, the region is vulnerable to urban flooding owing to geographical and climatic conditions. The Greater Bay Area has a subtropical maritime monsoon climate with concentrated and heavy rainfall. The terrain of a high north and low south makes it susceptible to flash floods. The central urban areas have low-lying and flat terrain with little gradient, resulting in poor natural drainage capacity [41], [42], [43].

B. Data

The data used for the study include remote sensing imagery, soil texture, soil type, weather station locations, etc. Table I shows all utilized data sources.

Landsat images are sourced from the Landsat 8 Level 2, Collection 2, Tier 1 dataset on Google Earth Engine (GEE). The years of image acquisition are 2013, 2015, 2018, 2020, and 2022. Sentinel-1 images are sourced from Sentinel-1 SAR GRD: C-band SAR Ground Range Detected, log scaling dataset on GEE. Sentinel-2 images are sourced from harmonized Sentinel-2 MSI: Multispectral Instrument, Level-2 A dataset on GEE. The year of image acquisition is 2022.

The spatial distribution of soil texture in China, provided by the Data Center of Resource and Environmental Sciences, Chinese Academy of Sciences (RESDC), is divided according to the content of sand, silt, and clay particles. The content of particles with different textures is reflected by percentage. Based on this, a soil type map with a one-kilometer resolution was obtained according to the International Standard for Soil Texture Classification [44].

The locations of 33 weather stations in the study area were obtained from the China National Meteorological Information Center and NOAA National Centers for Environmental Information. Among them, three stations are located in Hong Kong, one is in Macau, and 29 belong to the other cities in Guangdong province.

TABLE I
RESEARCH DATA AND SOURCES

Data	Source
Landsat 8	Landsat 8 Level 2, Collection 2, Tier 1 dataset on GEE
Sentinel-1	Sentinel-1 SAR GRD: C-band SAR Ground Range Detected, log scaling dataset on GEE
Sentinel-2	Harmonized Sentinel-2 MSI: Multi Spectral Instrument, Level-2A dataset on GEE
Digital Elevation Model	NASA SRTM Digital Elevation 30m dataset on GEE
Spatial distribution of soil texture	RESDC (https://www.resdc.cn/)
Soil types map	Obtained according to the International Standard for Soil Texture Classification
Weather stations	China National Meteorological Information Center (https://data.cma.cn/) & NOAA National Centers for Environmental Information (https://www.ncei.noaa.gov/)
Rainstorm intensity formula	New Urban Storm Intensity Formula in China [45] & Regulations of various administrative regions

III. METHODOLOGY

As shown in Fig. 2, the research methodology in this article has four steps: LULC classification, SCS-CN runoff modeling, flood extent estimation, and statistical analysis. In LULC classification, two different data sources are used for random forest classification. The first source is Landsat 8 imagery from the years 2013, 2015, 2018, 2020, and 2022. The second source involves Sentinel-1 and Sentinel-2 imagery from 2022, for comparison with Landsat results from the same year. The classification results are then used in subsequent experiments. Next, in the runoff modeling, four rainfall intensity scenarios are simulated using the SCS-CN model to obtain surface runoff. Then, the watersheds are divided based on the digital elevation model (DEM) and the locations of reservoirs, and flooded areas are extracted using the local equal volume method. Finally, LULC and inundation areas over the decade are statistically analyzed after normalization, i.e., calculating the correlation coefficient between the flood area and each LULC. Moreover, a comparative analysis is conducted on the classification results and flooded areas obtained from the two satellite data sources.

A. Preprocessing

Before LULC classification, we preprocessed the remote sensing images.

Landsat 8 images were obtained on GEE for 2013, 2015, 2018, 2020, and 2022. The imagery has undergone preprocessing such as atmospheric correction and orthorectification. Further cloud removal and fusion are needed to obtain the annual Landsat 8 images for five epochs. A mask is made from the cloud shadows and cloud data of the QA_PIXEL band of each image, and the pixels blocked by clouds and cloud shadows are removed through the mask. After cloud removal, the median pixel values of the annual images in the research area are extracted to form five new images.

Seasonal Sentinel-2 imagery was acquired to create an image with temporal features. Cloud removal processing was applied to each seasonal image, and the pixel values were fused by taking the median value of all pixel values at that location. For Sentinel-1 imagery, in addition to preprocessing steps such as eliminating ground range detected (GRD) border noise and thermal noise, as well as radiometric calibration and terrain correction, the study

also performed coherent speckle filtering and shadow removal on mountainous areas. The terrain data required for processing comes from SRTM30. The latest data for each month within the study period were obtained and combined to create monthly images. Due to the cloudy weather conditions in the Greater Bay Area, there were few images that could be used for classification most of the time.

B. Land Use/Land Cover Classification

1) Classification Using Landsat 8 Data: Considering the need to capture long-term LULC changes, the random forest method was used to classify five years of Landsat 8 data over the last decade. The LULC types were categorized into eight classes: farmland, orchard, forest, grassland, bare land, built-up land, water, and wetland. Random forest is an algorithm based on classification trees [46]. Among various classifiers, random forest has demonstrated higher accuracy, efficiency, and relatively lower computational complexity. It has become a reliable choice for studying LULC changes [47].

The random forest model in the study was trained using sample data from the 2020 imagery. Due to the difficulty in distinguishing some LULCs on the 30-m resolution images of Landsat 8, we first recorded the location of LULCs based on high-resolution images of Google Earth, and then confirmed and selected them as samples on the Landsat 8 images. The number of decision trees was tested at an interval of 5 within the range of 0–150, and the one with the highest classification accuracy was taken as the final number of decision trees. During training, all classes were treated as balanced, which helped improve the extraction of classes with fewer samples. The model was then used to classify imagery from the years 2013, 2015, 2018, and 2022. The annual validation data were selected by visual interpretation on Google Earth, ensuring that they did not overlap with the training samples. The overall accuracy represents the proportion of correctly classified samples relative to the total number of samples. Given the substantial variability in the areas of different LULC types within the study area, a completely random selection of validation samples would result in a small sample size for underrepresented LULC types. This could lead to an overall accuracy biased towards the dominant LULC types. Therefore, balancing the number of validation samples across classes is preferable. To address the imbalance, we obtained
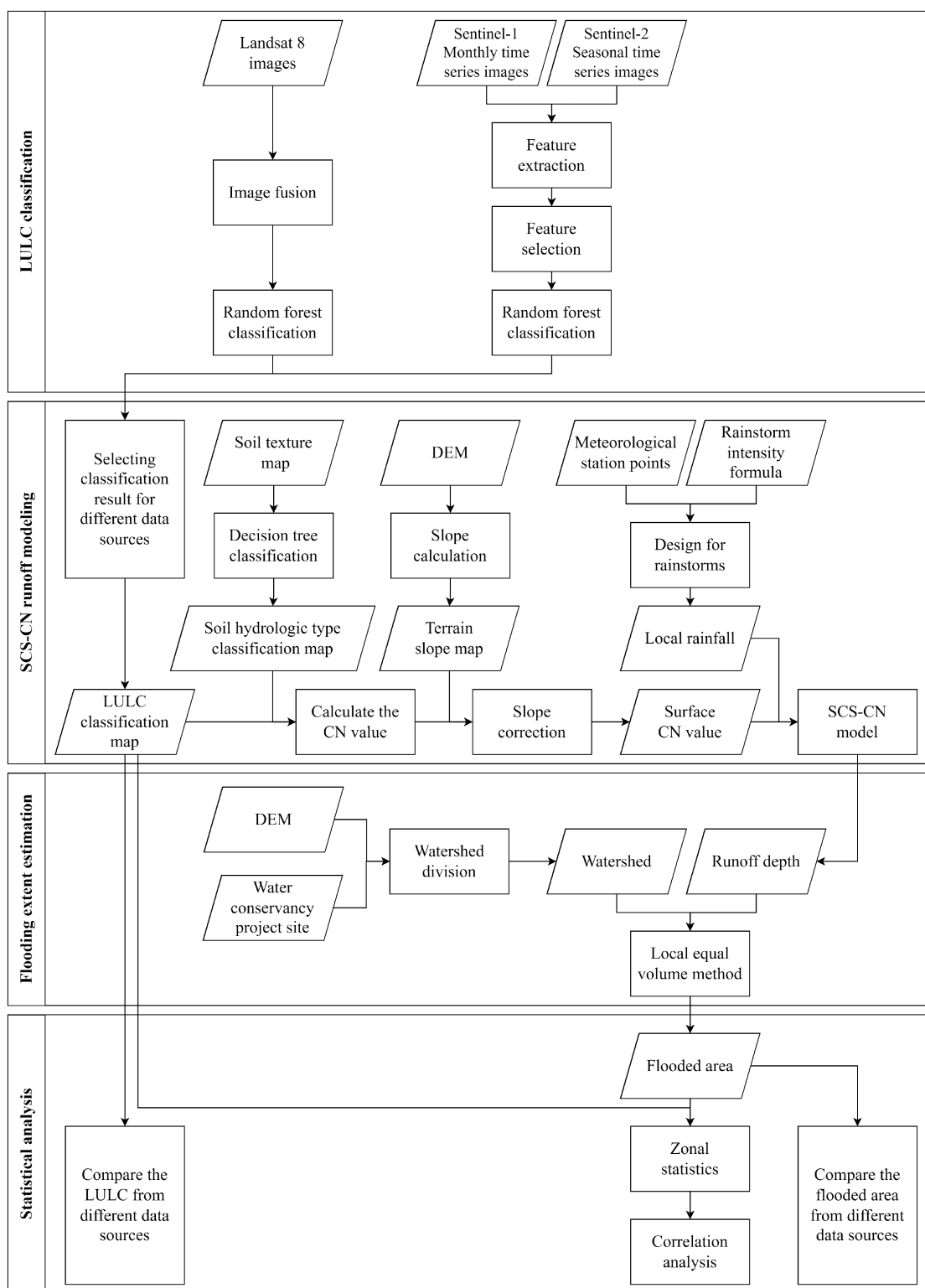
Fig. 2.　Framework of the research methodology.

samples for underrepresented LULC types as much as possible and selected representative samples for dominant LULC types. For example, in the case of wetlands, we selected samples within each catchment area, wherever available. Due to the absence of wetlands in some catchment areas, the number of catchment areas containing forests is approximately three times that of those containing wetlands. As forests generally cover a larger area, we therefore limited the number of forest samples in each catchment area. The resulting sample difference is due to the absence of wetlands in some catchment areas, which cannot be

further adjusted without compromising the representativeness of the forest samples across the study region. Therefore, even after balancing the interclass quantity of validation samples, the gap still exists. The forest sample was the largest, approximately 16 square kilometers, while the wetland sample was the smallest, approximately 5.7 square kilometers. The total number of pixels used for validation in each image should be greater than 100 000 and evenly distributed.

2) Classification Using Sentinel-1/2 Data: Sample selection here is consistent with that of Landsat imagery, with eight LULC types (farmland, orchard, forest, grassland, bare land, built-up land, water, and wetland) of samples selected by visual interpretation on Google Earth.

During preliminary feature selection, bands 2, 3, 4, 5, 6, 8, 11, and 12 were selected for the Sentinel-2 image, and then ten spectral indices such as normalized difference vegetation index (NDVI), normalized difference water index (NDWI), soil adjusted vegetation index (SAVI), enhanced vegetation index (EVI), land surface water index (LSWI), bare soil index (BSI), and 18 texture features such as angular second moment (ASM), contrast, correlation, variance, inverse differential moment (IDM), sum average were calculated for each image [48]. The above indices were combined with the original bands to form a new image. We selected the backscatter intensity of two polarization modes, VV and VH, for the Sentinel-1 image, and calculated six indices such as ratio and radar forest degradation index (RFDI), which were combined into a new image. In addition, we obtained four temporal features of each index in the temporal image: mean, maximum, minimum, and standard deviation.

Random forest can calculate out-of-bag (OOB) error to evaluate the importance of each feature [46]. Initially, features were inputted into the random forest in batches to calculate and rank their importance. The ten least important features were replaced or removed in each iteration until the classification accuracy no longer improved.

In summary, a total of 79 features were selected for Sentinel-2 optical imagery, including spectral, index, and texture features. For Sentinel-1 radar imagery, 16 features were selected, including terrain scattering characteristics and indices. Additionally, the DEM was included, resulting in a total of 96 features, as shown in Table II. All image bands were resampled to 10-m spatial resolution.

In this study, the classification was carried out with the decision tree number starting from 0 to 150 with a step size of 5. The decision tree number that yielded the best performance was selected as the final parameter for the random forest, and it was set to 80 based on the test.

C. SCS-CN Runoff Modeling

The SCS-CN model, produced by the United States Soil and Water Conservation Bureau, reflects the impact of varied land uses and soil types on surface runoff in watersheds [49]. The model requires relatively few input data, and the simulation results are highly reliable [50], [51], hence it is widely used in the calculation of surface runoff in coastal areas [22]. The SCS-CN

TABLE II
FEATURES EXTRACTED FROM SENTINEL 1/2 DATA FOR RANDOM
FORESTS CLASSIFICATION

Type	Feature	Number
Spectral characteristics	B2	4
	B3	4
	B4	4
	B5	4
	B6	4
	B7	4
	B8	4
	B11	4
	B12	4
Texture features	CONTRAST	1
	VAR	1
	IDM	1
	SAYG	4
Index characteristics	EVI	4
	NDVI	4
	NDWI	4
	NDTI	4
	SAVI	4
Terrain features	DEM	1
Temporal characteristics of spectral indices	standard deviation	5
	minimum	5
	maximum	5
	mean	5
Temporal characteristics of scattering	standard deviation	2
	minimum	2
	maximum	2
	mean	2
Temporal characteristics of the SAR index	standard deviation	1
	minimum	1
	maximum	1
	mean	1

method considers four main characteristics of runoff producing basins, namely, land use, soil type, hydrological conditions, and antecedent water conditions (AMC).

The formula for calculating surface runoff through the SCS-CN model is

$$\begin{cases} Q = \frac{(P-I_a)^2}{P-I_a+S} & (P > I_a) \\ Q = 0 & (P \le I_a) \end{cases} \quad (1)$$

where Q is the depth of surface runoff; P is the precipitation at a time; S is the maximum possible retention capacity; I_a is the initial loss value of rainfall. All units are in millimeters.

The formula for calculating the initial loss of rainfall is

$$I_a = \lambda S \quad (2)$$

where λ is dimensionless, usually taken as 0.2.

The formula for calculating the maximum possible retention capacity is

$$S = \frac{25400}{\text{CN}} - 254 \quad (3)$$

where CN is a dimensionless parameter that ranges from 0 to 100.

There are three grades of AMC, AMC III (wet), AMC II (normal), and AMC I (dry), which are based on the Antecedent Precipitation Index (API), i.e., the total amount of rainfall (mm) in the previous five days [52]. Based on the rainfall conditions

TABLE III
CN VALUE OF ANTECEDENT MOISTURE CONDITION LEVEL 2 (AMC II)

LULC	Soil hydrology type A	Soil hydrology type B	Soil hydrology type C	Soil hydrology type D
Farmland	67	78	85	89
Orchard	43	65	76	82
Forest	25	55	70	77
Built-up land	85	90	94	96
Water	98	98	98	98
Wetland	32	58	72	79
Bare land	77	86	91	94
Grassland	39	61	74	80

in the study area, the normal AMC II conditions, which are commonly used in local studies, are chosen in this study to calculate the CN value. The CN value is an empirical relationship obtained through statistical analysis based on measured data. The subsequent extensive application results have also proven its rationality. This study refers to the study by Peng et al. [53] for the CN values set for different LULC types, which updates the corrections of relevant research in the Pearl River Delta and surrounding regions based on the CN value lookup table provided by Mishra and Singh [54]. Finally, the CN values corresponding to different underlying surface types under the conditions of early medium soil moisture in the Greater Bay Area were determined as shown in Table III.

Due to the significant impact of terrain slope changes on surface runoff, when using the SCS-CN model, slope correction is also necessary for the CN value. The Williams slope correction formula [55] and Huang slope correction formula [56] are commonly used, between which the Huang slope correction formula is more suitable for mountainous areas. Since the Greater Bay Area is not mountainous, the Williams slope correction formula is adopted in this article, as follows:

$$\text{CN3} = \text{CN2} * \exp[0.00673 * (100 - \text{CN2})] \quad (4)$$

where CN2 is the initial value of CN collected above; CN3 is the curve value at another degree of wetting

$$\text{CN2}s = \frac{(\text{CN3} - \text{CN2})}{3} * [1 - 2 * e^{-13.86*\text{slope}}] + \text{CN2} \quad (5)$$

where CN2s is the CN value after slope correction; slope is calculated using SRTM30 m data in the GEE platform; CN3 is the curve value obtained from the previous formula for another degree of wetting.

Rainfall in the SCS-CN model is obtained through rainstorm design. For the rainstorms in this article, the designed scenarios of rainfall recurrence periods are 10 years, 20 years, 50 years, and a century, and all last for 60 min. Note we only need the final flooded area but not the rainfall process and rainfall temporal distribution. To improve the spatial rationality of rainfall, Thiessen polygons are used to divide meteorological stations in each city [57]. The rainfall within each Thiessen polygon is calculated using the rainstorm intensity formula corresponding to the administrative region to which the meteorological station belongs.

D. Flooding Extent Estimation

The process of each runoff from the urban surface converging to the corresponding runoff unit outlet is called confluence. As the study area is relatively large, watershed extraction and watershed division are required [58]. Taking the river network as input and combining with the locations of large-scale water projects such as reservoirs and dams within the watershed, the study site is divided into 160 catchment areas.

Based on the runoff in each catchment area, the accumulated water in that area is obtained. Next, the range and depth of flooding were simulated using the local equal volume method. The method treats rainfall as a passive submerged state, regardless of the actual process of convergence, and fills each depression with the total amount of runoff according to the surface elevation from high to low [59]. The formula is

$$W = \iint\limits_{A} [E_w(x, y) - E_g(x, y)]\mathrm{d}\delta \quad (6)$$

where W represents the total volume of water; A is the area of the flooded area; $E_w(x, y)$ is the water elevation of a certain point; $E_g(x, y)$ is the elevation of a certain ground point; $\mathrm{d}\delta$ the flooded units.

Because the flow rate of flooding and ponding in urban rainstorm is slow, the submerged water surface can be approximated as a plane, and the formula is

$$W = \iint\limits_{A} [E_w - E_g(x, y)]\mathrm{d}\delta \quad (7)$$

where E_w is the unified water surface elevation of the region.

Because the DEM data representing the surface elevation is a discrete regular grid, the formula is then discretized as follows:

$$W = \sum_{i=1}^{N} [E_w - E_g(i)]\Delta\sigma \quad (8)$$

where $\Delta\sigma$ is the area of the grid unit; N is the total grids number in the flooded area; $E_g(i)$ is the elevation of the ith grid.

After solving Ew for each catchment area, the submerged and nonsubmerged areas can be exported as elevation images. Then, the correlation between LULC and inundation areas over the decade can be statistically analyzed.

	2013 Landsat 8	2015 Landsat 8	2018 Landsat 8	2020 Landsat 8	2022 Landsat 8	2022 Sentinel-1/2
Overall accuracy	92.58%	87.83%	87.15%	89.97%	88.90%	93.20%
Kappa coefficient	0.91	0.84	0.83	0.87	0.86	0.92

E. Statistical Analysis

The entire study area was divided into 11 sections by administrative regions, and the LULC area and flooded area were counted within each section. The area of each administrative region varied greatly, hence the data needed to be normalized. Therefore, the proportion of each type of LULC area and the proportion of flooded area under each rainfall scenario were calculated for each administrative region over the years. The area proportions will be used to calculate the correlation coefficients between the eight LULC categories and the flooded area under the four rainfall scenarios.

Since our data do not follow a bivariate normal distribution, it is not appropriate to use the Pearson correlation coefficient calculation [60]. In addition, we do not assume a linear relationship between LULC and flooded area, but speculate that there is a monotonic correlation between the two. Thus, Spearman rank correlation is a more appropriate measurement in our case [60].

The Spearman rank correlation coefficient [61] can be regarded as a nonparametric version of the Pearson correlation coefficient. It is calculated between the ranks of two variables instead of their actual values [62]. The Spearman rank correlation coefficient ranges from -1 to 1, with closer to 0 indicating a lower correlation.

As an example, the formula for calculating the correlation coefficient (r_s) between forest and the flooded area under the once in a century rainfall scenario is (9) shown at the bottom of this page, where x is the proportion of forest area; y is the proportion of flooded area, R(x) and R(y) are the rank of x and y, respectively; S_s is the sum of squared differences of ranked variables x and y (x_i, respectively, y_i); n is 55. Other correlation coefficients of LULC can be similarly calculated.

IV. RESULTS AND ANALYSIS

A. LULC Changes

Using the 2020 Landsat 8 image, the random forest classifier was trained and tested, with an overall accuracy and a kappa coefficient of 89.97% and 0.87. The overall accuracy of each Landsat image exceeded 87%, with kappa coefficients above 0.83, as shown in Table IV. Based on Sentinel-1/2 images, the overall accuracy of the LULC classification was 93.20%, and the kappa coefficient was 0.92. The LULC results obtained by the two data sources are shown in Fig. 3.

	2013	2015	2018	2020	2022
Water	4122.69	3753.76	3702.54	3814.67	3865.43
Bare land	1158.13	1100.47	1273.55	1144.15	1122.96
Built-up land	9137.84	9317.98	10404.01	10155.97	9849.35
Grassland	6606.36	7166.65	6711.14	6045.72	5185.27
Orchard	8916.31	10863.40	7751.56	10502.05	11078.59
Forest	19569.84	16840.73	19813.20	18745.98	19451.91
Farmland	6568.10	6905.98	6353.91	5728.22	5561.72
Wetland	896.24	1026.55	965.60	838.77	860.27

The area occupied by each classified category can be found in Table V. It shows that in the past decade, the farmland area in the Greater Bay Area has shown a downward trend, while the areas of water, wetland, and bare land have remained relatively stable. Although bare land may have undergone positional changes due to human activities such as cultivation and construction, the total area has not changed much. The land use in the built-up area increased and peaked around 2018, and then decreased. Since 2015, the grassland area has been decreasing, while the forest and orchard areas have experienced significant fluctuations.

B. Urban Flooding Analysis

The designed rainfall scenarios are with recurrence periods of 10 years, 20 years, 50 years, and a century. The runoff is categorized into Level 1 (0–30 mm), Level 2 (30–50 mm), Level 3 (50–70 mm), Level 4 (70–85 mm), and Level 5 (85–100 mm). The runoff levels in the Greater Bay Area under the four rainfall scenarios obtained by using Sentinel-1/2 data are shown in Fig. 4.

From the runoff map of the 1-in-10-year rainfall scenario, the areas with higher runoff depths in the Greater Bay Area are: the junction of Jiangmen, Zhongshan, and Zhuhai, the junction of Foshan and Zhaoqing, the south coast of the Greater Bay Area, and the area around the estuary of the Pearl River. In the runoff map of the 20-year scenario, the overall runoff is deepened, with the runoff around the estuary of the Pearl River deepening into a circular pattern, while the runoff at the junction of Foshan and Zhaoqing deepens into the cities. With the increase of rainfall, the runoff at the junction of Jiangmen, Zhongshan, and Zhuhai has the greatest depth, mostly reaching Level 5. Meanwhile, the distribution of deep runoff areas in the estuary of the Pearl River spans a wide range, with runoff depths also reaching Level 4 to

$$r_s = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\left(R\left(x_i\right) - \overline{R(x)}\right) - \left(R\left(y_i\right) - \overline{R(y)}\right)\right)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left(R\left(x_i\right) - \overline{R(x)}\right)^2\right) * \left(\frac{1}{n}\sum_{i=1}^{n}\left(R\left(y_i\right) - \overline{R(y)}\right)^2\right)}} \tag{9}$$
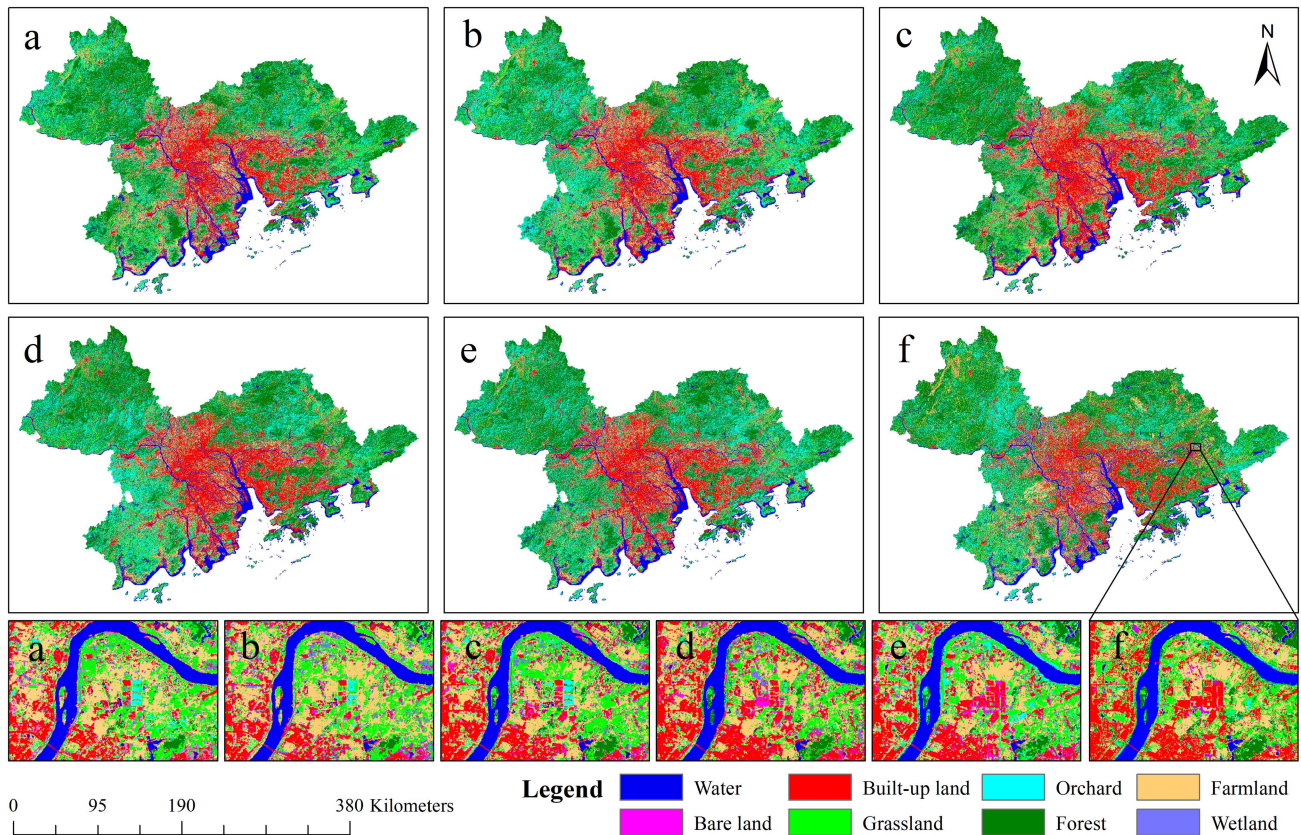
Fig. 3.   Classification results based on Landsat and Sentinel-1/2 data. (a)–(e) Results of Landsat 8 in 2013, 2015, 2018, 2020, and 2022, respectively. (f) Result of Sentinel-1/2 in 2022.

Level 5 in most of Dongguan, Shenzhen, and Hong Kong in the eastern part of the estuary.

Using the local equal volume method within each watershed, the flooded area under different rainfall scenarios was determined and presented in Fig. 5. It illustrates the inundation situation in various administrative regions. Zhaoqing in the northwest and Huizhou in the northeast of the Greater Bay Area are affected by hilly terrains, resulting in discrete blocks of flooded areas. Jiangmen and the Pearl River estuary in the southwest, as well as Dongguan, Shenzhen, Hong Kong, and Macau along the eastern coast, exhibit coastal and riverine distributions of affected areas, with less severe spreading and diffusion. The most severely affected areas are Guangzhou, Foshan, Zhongshan, and Zhuhai, located in the central part and the western part of the Pearl River estuary. They exhibit large-scale inundation, with numerous small blocks of inundated areas forming extensive coverage. These areas are most affected by urban flooding.

Table VI lists the flooded areas under the four rainfall scenarios. As can be seen, over the past decade, flooding trends have varied between scenarios, due to the fluctuation of the terrain and changes in the land. The flooded area of the 1-in-10-year rainfall scenario increases abruptly from 2013 to 2015 and remains high. Then, until 2022, it decreases to the level of 2013. The flooded area of the 1-in-20-year rainfall scenario is relatively stable. The flooded area of the once-in-50-year rainfall scenario increases from 2013 to 2015, then stabilizes for the next three years before

TABLE VI
FLOODED AREA WITH RAINFALL SCENARIOS OF DIFFERENT RECURRENCE PERIODS (km^2)

Rainfall scenarios	2013	2015	2018	2020	2022
Once in 10 years	5695.99	5707.90	5707.14	5704.80	5696.26
Once in 20 years	5814.02	5815.33	5814.33	5814.50	5814.06
Once in 50 years	5920.48	5932.06	5931.37	5930.80	5925.99
Once in a century	6088.79	6103.43	6091.55	6095.04	6102.80

declining in 2022. The flooded area of the once-in-100-year rainfall scenario only decreases from 2015 to 2018, with a larger area in 2015 and 2022.

C. Analysis of the Influence of LULC Change on Urban Flooding

For the five LULC classification results obtained from Landsat 8 data over a decade, the flooded areas of 11 administrative regions were calculated, and 55 sets of data were obtained for each rainfall scenario. Table VII shows the flooded area of each administrative region under the once in a century rainfall scenario. After combination and normalization of the data from the 5 epochs across the decade, the correlation coefficients between the proportion of flood area and the proportion of each LULC are calculated for each of the rainfall scenarios, as shown in Table VIII. Fig. 6 shows the scatter charts of the proportions of
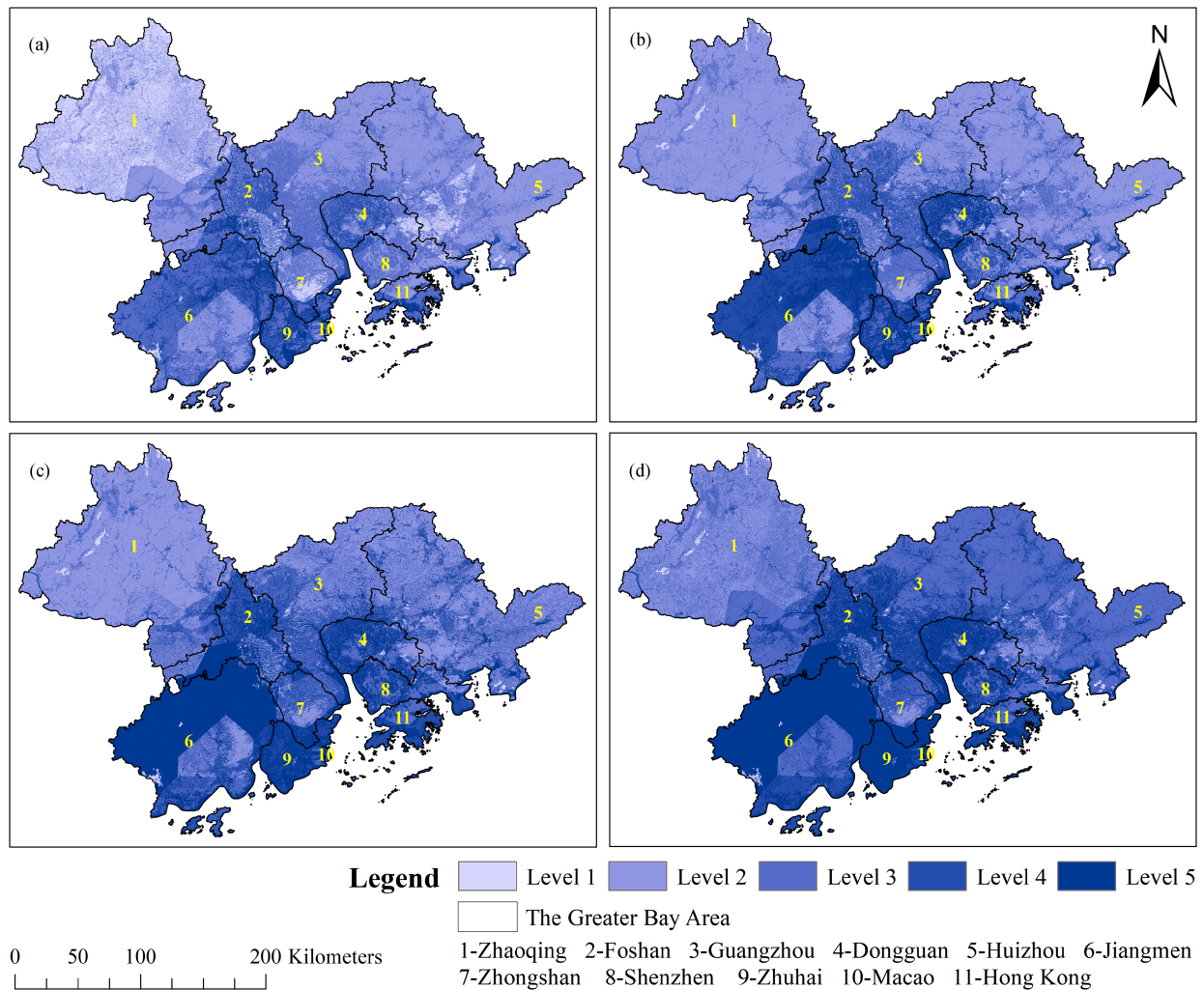
Fig. 4. Runoff levels in 2022 derived from Sentinel-1/2 data. (a)–(d) Results of once-in-10-year rainfall, once-in-20-year rainfall, once-in-50-year rainfall and once-in-a-century rainfall, respectively. Level 1 (0–30 mm); Level 2 (30–50 mm); Level 3 (50–70 mm); Level 4 (70–85 mm); Level 5 (85–100 mm).

TABLE VII
FLOODED AREA (km^2) AND PROPORTION (%) OF EACH ADMINISTRATIVE REGION UNDER A ONCE-IN-A-CENTURY RAINSTORM

Admin. regions	No.	2013		2015		2018		2020		2022	
		Area	Proportion	Area	Proportion	Area	Proportion	Area	Proportion	Area	Proportion
Zhaoqing	1	439.31	2.95	440.44	2.96	439.33	2.95	439.61	2.95	439.91	2.95
Foshan	2	540.19	14.21	540.19	14.21	540.19	14.21	540.19	14.21	540.19	14.21
Guangzhou	3	891.33	12.05	891.33	12.05	891.33	12.05	891.33	12.05	891.33	12.05
Dongguan	4	303.41	12.33	303.41	12.33	303.41	12.33	303.41	12.33	303.41	12.33
Huizhou	5	559.88	4.88	565.82	4.93	559.88	4.88	565.82	4.93	570.55	4.97
Jiangmen	6	1403.29	14.48	1408.11	14.53	1403.29	14.48	1403.29	14.48	1403.29	14.48
Zhongshan	7	499.27	27.50	502.00	27.65	502.00	27.65	499.27	27.50	502.00	27.65
Shenzhen	8	281.23	13.18	281.23	13.18	281.23	13.18	281.23	13.18	281.23	13.18
Zhuhai	9	892.25	45.72	892.25	45.72	892.25	45.72	892.25	45.72	892.25	45.72
Macao, SAR	10	32.40	61.59	32.40	61.59	32.40	61.59	32.40	61.59	32.40	61.59
Hong Kong, SAR	11	246.24	19.16	246.24	19.16	246.24	19.16	246.24	19.16	246.24	19.16
Total submerged area		6088.79		6103.43		6091.55		6095.04		6102.80	

flood area and LULC types under the once-in-a-century rainfall scenario.

The results show that water, built-up land, farmland, and wetland have positive correlation coefficients, among which water has the highest one, whereas bare land, grassland, orchard, and forest have negative correlation coefficients, and forest has

the highest. A positive correlation coefficient indicates that an increase in the corresponding LULC type will lead to an increase in the urban flooding area under the same precipitation conditions, while a negative correlation coefficient indicates the opposite. The larger the correlation coefficient, the stronger the effect of the LULC type on urban flooding. In all the significance tests of
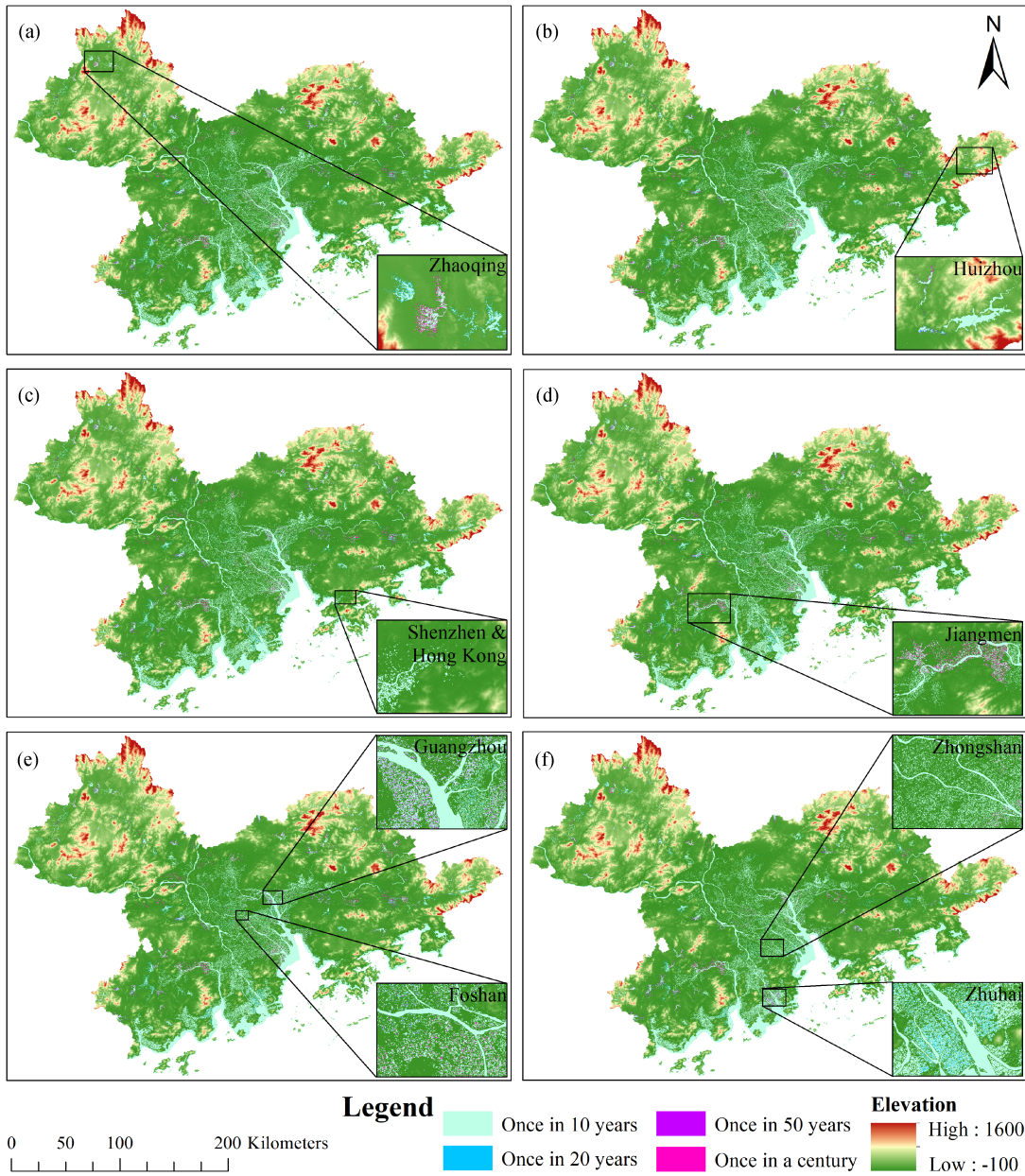
Fig. 5. Flooded areas under different rainfall scenarios. (a)–(e) Results of the five years based on Landsat 8 data. (f) Result based on the 2022 Sentinel-1/2 data.

TABLE VIII
CORRELATION COEFFICIENT BETWEEN DIFFERENT LULC AND FLOODED AREA

Rainfall scenarios	Water	Bare land	Built-up	Grassland	Orchard	Forest	Farmland	Wetland
Once in 10 years	0.93	-0.40	0.39	-0.59	-0.54	-0.74	0.20	0.24
Once in 20 years	0.93	-0.39	0.39	-0.59	-0.54	-0.74	0.20	0.24
Once in 50 years	0.92	-0.40	0.44	-0.63	-0.59	-0.77	0.22	0.25
Once in a century	0.92	-0.40	0.44	-0.63	-0.59	-0.77	0.22	0.26
Average	0.93	-0.40	0.42	-0.61	-0.57	-0.75	0.21	0.25

correlation coefficients, only the p-value of farmland was greater than 0.05. It is considered that there is no significant correlation between the area of farmland and the area of inundation. Beside that, the effect of water, bare land, built-up, grassland, orchard, forest on flooded area was significant at the 0.01 level, and the effect of wetland was significant at the 0.05 level.

It can be seen that bare land, grassland, orchard, and forest all have a certain capacity to reduce surface runoff and flooding. Among them, forest has the best flood control and storage capacity, followed by grassland and orchard, while bare land has the lowest capacity. The increase in water significantly increases the area of urban floods, thus, the water level should be carefully
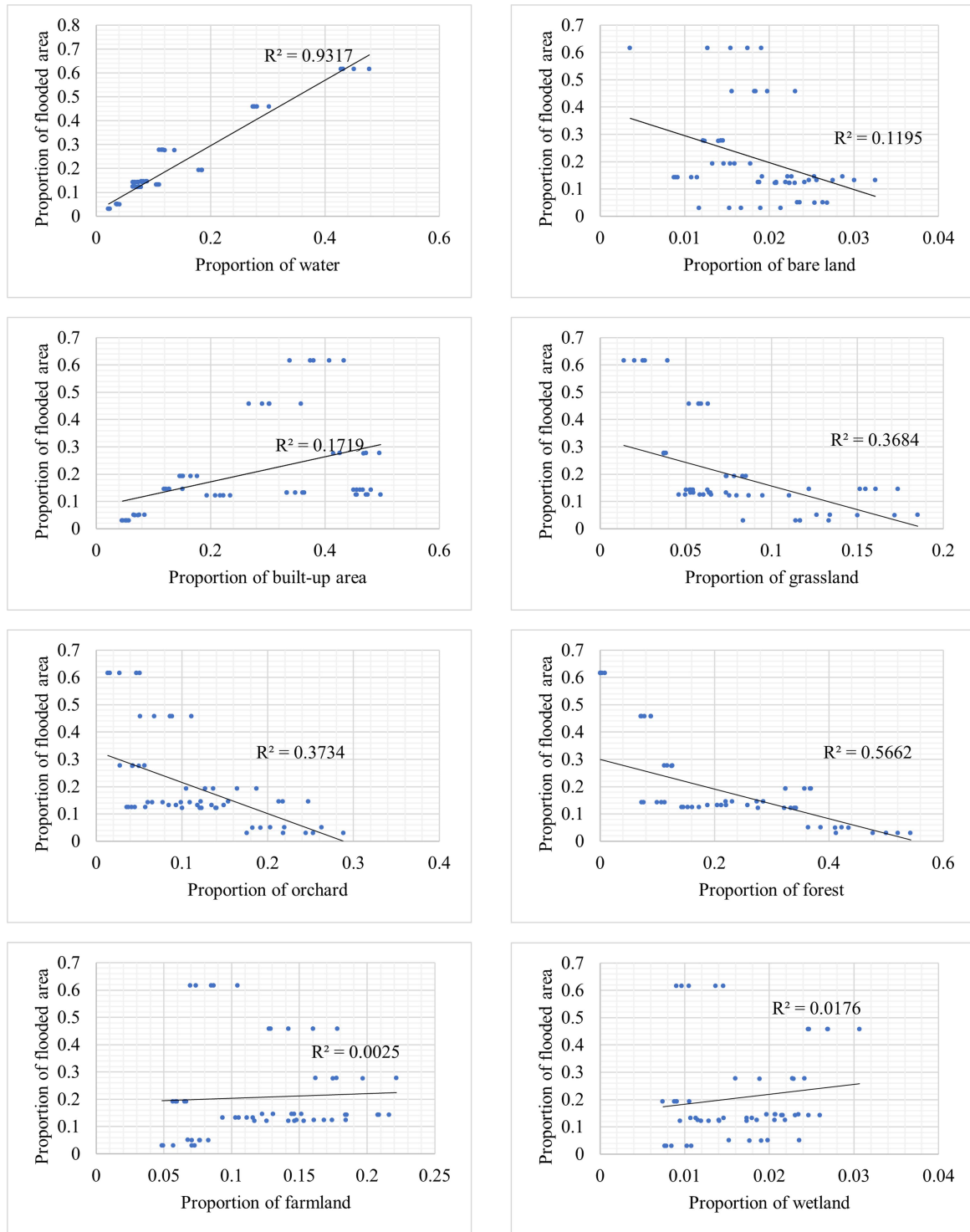
Fig. 6. Scatter plot of proportion of LULC and flooded area in each administrative region.

monitored. The expansion of built-up land also causes a rise in the flooding area. There is no significant correlation between farmland and flooding. The correlation of wetland was small, probably because it has a small area and few changes. However, the role of wetland in regulating flooding cannot be denied. If water is converted into wetland, there can be a notable reduction in surface runoff, which is obviously easier than converting water into forest.

D. Comparison of Experimental Results From Different Data Sources

A comparison is made between the classification results of Landsat and Sentinel data, as shown in Table IX. The comparison of their flooded area results is shown in Table X. The columns of these two tables represent the results based on Landsat data, and the rows on Sentinel data.

TABLE IX
COMPARISON OF CLASSIFICATION RESULTS FROM DIFFERENT DATA SOURCES (%)

Sentinel \ Landsat	User's accuracy	Water	Bare land	Built-up land	Grassland	Orchard	Forest	Farmland	Wetland
User's accuracy		99.89	64.78	93.85	38.80	76.31	90.40	71.50	49.30
Water	99.91	94.408	0.166	1.017	0.041	0.026	0.133	0.406	8.085
Bare land	96.81	0.027	75.113	0.782	0.373	0.001	0.011	0.988	0.39
Built-up land	91.62	2.115	8.791	89.648	0.723	0.01	0.273	6.938	7.474
Grassland	64.67	0.069	4.071	0.522	78.263	2.089	1.981	6.486	0.836
Orchard	77.94	0.033	0.012	0.015	4.433	86.265	4.523	0.456	0.087
Forest	90.11	0.454	0.287	0.579	7.865	11.241	91.1	3.515	3.38
Farmland	95.79	0.502	11.177	6.626	8.241	0.362	1.801	80.185	9.23
Wetland	98.97	2.176	0.383	0.811	0.061	0.005	0.177	1.026	70.517

TABLE X
COMPARISON OF FLOODED AREAS USING LANDSAT AND SENTINEL DATA (%)

Sentinel \ Landsat	Unflooded area	Flooded area	Rainfall scenarios
Unflooded area	99.894	1.084	Once in 10 years
Flooded area	0.106	98.916	
Unflooded area	99.921	1.028	Once in 20 years
Flooded area	0.079	98.972	
Unflooded area	99.922	1.302	Once in 50 years
Flooded area	0.078	98.698	
Unflooded area	99.922	2.135	Once in a century
Flooded area	0.078	97.865	

Comparing the classification results of Landsat 8 with Sentinel-1/2 data, it can be observed that there is high consistency in water, built-up land, orchard, and forest between the two. However, the bare land, grassland, and wetland in the Sentinel results are occasionally classified as farmland in the Landsat 8 results. This is because the Sentinel-1/2 data have higher temporal resolution and can utilize temporal features for classification, while the cloudy and rainy climate in southern China during the summer limits the availability of sufficient temporal features in Landsat 8 data. Additionally, during the harvesting period, farmland and bare land appear similar, making them prone to misclassification when Landsat 8 data are used. The consistency in the classification of wetland between the two sets of results is relatively low, as wetland can appear or disappear seasonally with water level changes, and better classification results can be achieved when temporal features are available. The built-up land in the Sentinel-1/2 results is also prone to being classified as farmland in the Landsat 8 results. This is because most of the bands in Sentinel-1/2 have higher spatial resolution, allowing for the differentiation of scattered built-ups near farmland, while at 30-m resolution, such built-ups are easily mixed within farmland pixels in Landsat 8 data.

Table X shows that the two data sources have produced highly consistent results, with more than 99% agreement for unflooded areas and around 98% agreement for flooded areas under all four rainfall scenarios, with differences mainly observed in scattered flooded areas in relatively flat regions and at the edges of a few rivers and lakes in mountainous areas. There is a tendency to obtain more flooded areas with Landsat data than with Sentinel data. Sentinel data have higher spatial and spectral resolutions, which can better distinguish complex and mixed LULC types in cities, such as grassland and farmland.

V. DISCUSSION

Previous studies have shown that built-up land increases the impermeability of the surface, thereby increasing surface runoff and leading to more severe urban flooding, while forest can reduce surface runoff and hence the risk of flooding [63], [64]. This is consistent with the results of this study. However, we made a more fine-grained categorization of vegetation, revealed the difference in the flood mitigation capacity of forest, orchard, and grassland, and also quantified the correlation between the eight LULC types and flooding. More detailed categorization and quantification can be beneficial for accurate modeling and assessment of flood risk. For example, the impact of vegetation has been accounted for in some flood hazard risk assessments, but these studies tend to lump all vegetation together [65], [66], [67]. In this study, we have demonstrated that different vegetation resists flooding differently and quantitatively assessed the relationship between different vegetation types and flood hazard, which is useful for a more accurate assessment of flood risk. The results of the relationship between LULC and urban flooding serve as a reference for urban flood control decision-making and planning in the Greater Bay Area, while the analysis of different remote sensing data sources provides evidence that the variation of input data sources has an insignificant impact on the final results. There is only an inconsistency of 1%–2% when using mixed data with apparent different spatial and spectral resolutions.

In terms of LULC classification, this study used the random forest method, which has been widely used in experiments and has been proven to have high accuracy and stability [47]. Prasad et al. [68] compared the classification accuracy of different algorithms when using Landsat 8 or fused Sentinel-1/2 data, and

the results showed that random forest classification is better than ANN and most algorithms. However, with the rapid development of deep learning, complex neural networks have been applied to LULC classification and achieved higher accuracy [69], [70]. Future studies can try to use additional classification methods, provided enough training data is available.

In terms of urban flooding area extraction, there is no urban drainage component in the model, limited by the difficulty in obtaining data. Urban drainage data in the study area are not publicly available. Quantitative assessment of urban drainage capacity on a large scale is still lacking and could be a future study after overcoming the data source problem.

In addition, based on the proposed methodology in this study, the experiment can be applied to other regions, and the differences in the impact of LULC on urban flooding in different climates can be analyzed. However, attention should be paid to the subtle differences in surface infiltration relationships caused by geological, soil, terrain slope, weather, and other factors in different regions, taking into account the correction of CN values in different regions through field experiments. Expanding the scope of research could be a direction for future study.

VI. Conclusion

This study examined the LULC changes in the Greater Bay Area, China, over the past decade and simulated the extent of urban flooding under different intensities of rainfall. These two factors are combined to explore the influence of LULC change on urban flooding and analyze the influence of using different remote sensing data on the experiment results. It has been found that water and built-up land show a positive correlation with urban flooding, while bare land, grassland, orchard, and forest exhibit a negative correlation. Meanwhile, wetland has a positive but insignificant correlation, and there is no significant correlation for agricultural land. The Landsat and Sentinel based urban flooding simulation results are highly consistent, with differences mainly in scattered flooded areas in flat terrain, ranging from 1% to 2%. The findings of this study can provide a reference for the Greater Bay Area to make informed decisions for urban flooding mitigation.

References

[1] Z. Yin, J. Yin, S. Xu, and J. Wen, "Community-based scenario modelling and disaster risk assessment of urban rainstorm waterlogging," *J. Geogr. Sci.*, vol. 21, no. 2, pp. 274–284, 2011, doi: 10.1007/s11442-011-0844-7.

[2] X. Wu, D. Yu, Z. Chen, and R. L. Wilby, "An evaluation of the impacts of land surface modification, storm sewer development, and rainfall variation on waterlogging risk in Shanghai," *Nat. Hazards*, vol. 63, no. 2, pp. 305–323, 2012, doi: 10.1007/s11069-012-0153-1.

[3] S. Roy, A. Bose, N. Singha, D. Basak, and R. C. Indrajit, "Urban waterlogging risk as an undervalued environmental challenge: An integrated MCDA-GIS based modeling approach," *Environ. Challenges*, vol. 4, 2021, Art. no. 100194.

[4] X. Wang, R. Jiang, J. Xie, Y. Zhao, F. Li, and J. Zhu, "Multiscale variability of precipitation and their teleconnection with large-scale climate anomalies," *J. Coast. Res.*, vol. 93, pp. 417–426, 2019.

[5] J. Yin, M. Ye, Z. Yin, and S. Xu, "A review of advances in urban flood risk analysis over China," *Stoch. Environ. Res. Risk Assess.*, vol. 29, no. 3, pp. 1063–1070, 2015.

[6] R.-S. Quan, "Rainstorm waterlogging risk assessment in central urban area of Shanghai based on multiple scenario simulation," *Nat. Hazards*, vol. 73, pp. 1569–1585, 2014.

[7] G. Wang, J. Chen, C. Zhao, X. Zhou, and X. Deng, "Exploration of the causality between area changes of green spaces and waterlogging frequency in Beijing," *Phys. Chem. Earth*, vol. 101, pp. 172–177, 2017.

[8] Y.-F. Ning, W.-Y. Dong, L.-S. Lin, and Q. Zhang, "Analyzing the causes of urban waterlogging and sponge city technology in China," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 59, no. 1, 2017, Art. no. 012047, doi: 10.1088/1755-1315/59/1/012047.

[9] M. Ma et al., "Investigation of inducements and defenses of flash floods and urban waterlogging in Fuzhou, China, from 1950 to 2010," *Nat. Hazards*, vol. 91, pp. 803–818, 2018.

[10] S. Du, A. Van Rompaey, P. Shi, and J. Wang, "A dual effect of urban expansion on flood risk in the Pearl River Delta (China) revealed by land-use scenarios and direct runoff simulation," *Nat. Hazards*, vol. 77, pp. 111–128, 2015.

[11] A. P. Cracknell, "The development of remote sensing in the last 40 years," *Int. J. Remote Sens.*, vol. 39, pp. 8387–8427, 2018.

[12] H. Huang, J. Legarsky, and M. Othman, "Land-cover classification using radarsat and landsat imagery for St. Louis, Missouri," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 1, pp. 37–43, 2007.

[13] T. Zhou, Z. Li, and J. Pan, "Multi-feature classification of multi-sensor satellite imagery based on dual-polarimetric sentinel-1A, landsat-8 OLI, and hyperion images for urban land-cover classification," *Sensors*, vol. 18, no. 2, 2018, Art. no. 373.

[14] D. L. Giandomenico, M. N. S. João, D. F. Salvatore, and M. Giuseppe, "Integrated use of sentinel-1 and sentinel-2 data and open-source machine learning algorithms for land cover mapping in a mediterranean region," *Eur. J. Remote. Sens.*, vol. 55, no. 1, pp. 52–70, 2022.

[15] B. Chen, B. Huang, and B. Xu, "Multi-source remotely sensed data fusion for improving land cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 27–39, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271616306463

[16] H. Gao et al., "A crop classification method integrating GF-3 Polsar and Sentinel-2A optical data in the Dongting Lake Basin," *Sensors*, vol. 18, no. 9, 2018, Art. no. 3139. [Online]. Available: https://www.mdpi.com/1424-8220/18/9/3139

[17] J. Zhang, "Multi-source remote sensing data fusion: Status and trends," *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 5–24, 2010, doi: 10.1080/19479830903561035.

[18] P. Koudelak and S. West, "Sewerage network modelling in Latvia, use of infoworks CS and storm water management model 5 in Liepaja city," *Water Environ. J.*, vol. 22, no. 2, pp. 81–87, 2008.

[19] J. J. Wiles and N. S. Levine, "A combined GIS and HEC model for the analysis of the effect of urbanization on flooding; the Swan Creek watershed, Ohio," *Environ. Eng. Geosci.*, vol. 8, no. 1, pp. 47–61, 2002.

[20] H. Zhou and C. Wang, "Guangdong Dongguan city urban flooding causes analysis and prevention measures," *China Flood Drought Manag.*, vol. 23, no. 4, pp. 70–71, 2013.

[21] J.-W. Jung et al., "Water management practices and SCS curve numbers of paddy fields equipped with surface drainage pipes," *Agric. Water Manage.*, vol. 110, pp. 78–83, 2012.

[22] S. Jiao, X. Zhang, and Y. Xu, "A review of Chinese land suitability assessment from the rainfall-waterlogging perspective: Evidence from the Su Yu Yuan area," *J. Clean. Prod.*, vol. 144, pp. 100–106, 2017.

[23] J. A. Reistetter and M. Russell, "High-resolution land cover datasets, composite curve numbers, and storm water retention in the Tampa Bay, FL region," *Appl. Geogr.*, vol. 31, no. 2, pp. 740–747, 2010.

[24] S. Verma, P. K. Singh, S. K. Mishra, V. P. Singh, V. Singh, and A. Singh, "Activation soil moisture accounting (Asma) for runoff estimation using soil conservation service curve number (SCS-CN) method," *J. Hydrol*, vol. 589, 2020, Art. no. 125114. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169420305746

[25] X. Zhang, F. Meng, and N. Ding, "Application of SCS model to estimating the quantity of runoff of small watershed in semi-arid or arid region," *Res. Soil Water Conserv*, vol. 10, no. 4, pp. 172–174, 2003.

[26] B. Xiao, Q.-H. Wang, J. Fan, F.-P. Han, and Q.-H. Dai, "Application of the SCS-CN model to runoff estimation in a small watershed with high spatial heterogeneity," *Pedosphere*, vol. 21, no. 6, pp. 738–749, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S100201601160177X

[27] A. Hagras, "Runoff modeling using SCS-CN and GIS approach in the Tayiba valley basin, Abu Zenima area, south-west Sinai, Egypt," *Model. Earth Syst. Environ.*, vol. 9, pp. 3883–3895, 2023, doi: 10.1007/s40868-023-01714-5.

[28] X. Liu and J. Li, "Application of SCS model in estimation of runoff from small watershed in loess plateau of China," *Chin. Geogr. Sci.*, vol. 18, no. 3, pp. 235–241, 2008.

[29] S. Xie, W. Liu, Z. Yuan, H. Zhang, H. Lin, and Y. Wang, "Integrated risk assessment of waterlogging in Guangzhou based on runoff modeling, AHP, GIS and scenario analysis," *Water*, vol. 14, no. 18, 2022, Art. no. 2899. [Online]. Available: https://www.mdpi.com/2073-4441/14/18/2899

[30] Z. Chen, K. Li, J. Du, Y. Chen, R. Liu, and Y. Wang, "Three-dimensional simulation of regional urban waterlogging based on high-precision DEM model," *Nat. Hazards*, vol. 108, no. 3, pp. 2653–2677, 2021, doi: 10.1007/s11069-021-04793-8.

[31] C. Yang et al., "Spatiotemporal evolution of urban agglomerations in four major bay areas of US, China and Japan from 1987 to 2017: Evidence from remote sensing images," *Sci. Total Environ.*, vol. 671, pp. 232–247, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969719311362

[32] M. Yan, J. C. L. Chan, and K. Zhao, "Impacts of urbanization on the precipitation characteristics in Guangdong province, China," *Adv. Atmos. Sci.*, vol. 37, no. 7, pp. 696–706, 2020, doi: 10.1007/s00376-020-9218-3.

[33] B. Liu, S. Chen, X. Tan, and X. Chen, "Response of precipitation to extensive urbanization over the Pearl River Delta metropolitan region," *Environ. Earth Sci.*, vol. 80, no. 1, 2021, Art. no. 9, doi: 10.1007/s12665-020-09299-3.

[34] D. Zhao, J. Zha, and J. Wu, "Changes in rainfall of different intensities due to urbanization-induced land-use changes in Shenzhen, China," *Clim. Dyn.*, vol. 56, no. 7, pp. 2509–2530, 2021, doi: 10.1007/s00382-020-05601-y.

[35] L. Li et al., "Review of advances in urban climate study in the Guangdong-Hong Kong-Macau greater bay area, China," *Atmos. Res.*, vol. 261, 2021, Art. no. 105759. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016980952100315X

[36] K. M. Wai et al., "Observational evidence of a long-term increase in precipitation due to urbanization effects and its implications for sustainable urban living," *Sci. Total Environ.*, vol. 599-600, pp. 647–654, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969717311105

[37] Y. Li, W. Wang, M. Chang, and X. Wang, "Impacts of urbanization on extreme precipitation in the Guangdong-Hong Kong-Macau greater bay area," *Urban Clim.*, vol. 38, 2021, Art. no. 100904. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212095521001346

[38] Z. Deng, Z. Wang, X. Wu, C. Lai, and Z. Zeng, "Strengthened tropical cyclones and higher flood risk under compound effect of climate change and urbanization across China's greater bay area," *Urban Clim*, vol. 44, 2022, Art. no. 101224. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212095522001420

[39] J. Zheng, H. Wang, and B. Liu, "Impact of the long-term precipitation and land use changes on runoff variations in a humid subtropical river basin of China," *J. Hydrol. Reg. Stud*, vol. 42, 2022, Art. no. 101136. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214581822001495

[40] P.-J. Shi, Y. Yuan, J. Zheng, J.-A. Wang, Y. Ge, and G.-Y. Qiu, "The effect of land use/cover change on surface runoff in Shenzhen region, China," *Catena*, vol. 69, no. 1, pp. 31–35, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0341816206000944

[41] B. Liu, J. Chen, W. Lu, X. Chen, and Y. Lian, "Spatiotemporal characteristics of precipitation changes in the Pearl River basin, China," *Theor. Appl. Climatol.*, vol. 123, no. 3, pp. 537–550, 2016, doi: 10.1007/s00704-015-1375-4.

[42] J. Niu and J. Chen, "Terrestrial hydrological features of the Pearl River Basin in South China," *J. Hydro-environ. Res.*, vol. 4, no. 4, pp. 279–288, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S157064431000050X

[43] H. Wu, Z. Zhai, Z. Li, and Y. Liu, "Guangdong-Hong Kong-Macao greater bay area climate bulletin," *Guangdong Meteorological Bur.*, Guangzhou, China, 2019. [Online]. Available: https://www.hko.gov.hk/en/wxinfo/pastwx/2018/files/GD_HK_Mac_GBA_2018.pdf

[44] K. Wu and R. Zhao, "Soil texture classification and its application in China," *Acta Pedologica Sinica*, vol. 56, pp. 227–241, 2019.

[45] Y. Shao and D. Shao, *New Urban Storm Lntensity Formula in China*. Beijing, China: China Architecture & Building Press, 2014.

[46] L. Breiman, "Random forests," *Mach. Learn*, vol. 45, no. 1, pp. 5–32, 2001.

[47] M. Junaid, J. Sun, A. Iqbal, M. Sohail, S. Zafar, and A. Khan, "Mapping LULC dynamics and its potential implication on forest cover in Malam Jabba region with landsat time series imagery and random forest classification," *Sustainability*, vol. 15, no. 3, 2023, Art. no. 1858.

[48] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[49] W. Boughton, "A review of the USDA SCS curve number method," *Soil Res.*, vol. 27, no. 3, pp. 511–523, 1989.

[50] J. Du et al., "Assessing the effects of urbanization on annual runoff and flood events using an integrated hydrological modeling system for Qinhuai river basin, China," *J. Hydrol.*, vol. 464-465, pp. 127–139, 2012.

[51] Q. Weng, "Modeling urban growth effects on surface runoff with the integration of remote sensing and GIS," *Environ. Manage.*, vol. 28, pp. 737–748, 2001.

[52] S. Fu, X. Wang, H. Wang, X. Wei, and A. Yuan, "Method of determining CN value in the SCS-CN method," *Arid Land Geogr.*, vol. 35, no. 3, pp. 415–421, 2012.

[53] J. Peng, H. Wei, W. Wu, Y. Liu, and Y. Wang, "Storm flood disaster risk assessment in urban area based on the simulation of land use scenarios: A case of Maozhou watershed in Shenzhen city," *Acta Ecologica Sinica*, vol. 38, pp. 3741–3755, 2018.

[54] S. K. Mishra and V. P. Singh, *SCS-CN Method*. Berlin, Germany: Springer, 2003, doi: 10.1007/978-94-017-0147-1_2.

[55] J. R. Williams, "The erosion-productivity impact calculator (epic) model: A case history," *Philos. Trans. R. Soc. Lond., B., Biol. Sci.*, vol. 329, no. 1255, pp. 421–428, 1990.

[56] M. Huang, J. Gallichand, Z. Wang, and M. Goulet, "A modification to the soil conservation service curve number method for steep slopes in the loess plateau of China," *Hydrol. Process.*, vol. 20, no. 3, pp. 579–589, 2006.

[57] H. Derakhshan and N. Talebbeydokhti, "Rainfall disaggregation in non-recording gauge stations using space-time information system," *Sci. Iran.*, vol. 18, no. 5, pp. 995–1001, 2011.

[58] J. F. O'Callaghan and D. M. Mark, "The extraction of drainage networks from digital elevation data," *Comput. Vis. Graph. Image Process.*, vol. 27, no. 2, pp. 323–344, 1984.

[59] L. Zhang, "Urban flooding analysis and modeling using GIS," *Acta Geogr. Sinica*, vol. 50, pp. 76–84, 1995.

[60] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, 2018. [Online]. Available: https://journals.lww.com/anesthesia-analgesia/fulltext/2018/05000/correlation_coefficients__appropriate_use_and.50.aspx

[61] E. C. Fieller, H. O. Hartley, and E. S. Pearson, "Tests for rank correlation coefficients. I," *Biometrika*, vol. 44, no. 3/4, pp. 470–481, 1957.

[62] G. Frapporti, L. A. M. Linnartz, and S. P. Vriend, "Spearmen-a dBase program for computation and testing of spearman rank correlation co-efficient distributions," *Comput. Geosci.*, vol. 17, no. 4, pp. 569–589, 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/009830049190115 T

[63] S. Y. Tola and A. Shetty, "Land cover change and its implication to hydrological regimes and soil erosion in Awash river basin, Ethiopia: A systematic review," *Environ. Monit. Assess.*, vol. 193, no. 12, 2021, Art. no. 836, doi: 10.1007/s10661-021-09599-6.

[64] P. E. Zope, T. I. Eldho, and V. Jothiprakash, "Impacts of urbanization on flooding of a coastal urban catchment: A case study of Mumbai city, India," *Nat Hazards (Dordr)*, vol. 75, no. 1, pp. 887–908, 2015, doi: 10.1007/s11069-014-1356-4.

[65] Y. Wang, J. Zhai, G. Gao, Q. Liu, and L. Song, "Risk assessment of rainstorm disasters in the Guangdong-Hong Kong-Macao greater bay area of China during 1990–2018," *Geomatics, Nat. Hazards Risk*, vol. 13, no. 1, pp. 267–288, 2022. doi: 10.1080/19475705.2021.2023224.

[66] Y. Li, Z. Zhang, S. Gong, M. Liu, and Y. Zhao, "Risk assessment of rainstorm disasters under different return periods: A case study of Bohai Rim, China," *Ocean Coast. Manage.*, vol. 187, 2020, Art. no. 105107.

[67] Z. Zhu, S. Zhang, Y. Zhang, R. Yao, and H. Jin, "Integrating flood risk assessment and management based on HV-SS model: A case study of the pearl river delta, China," *Int. J. Disaster Risk Reduct.*, vol. 96, 2023, Art. no. 103963. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212420923004430

[68] P. Prasad, V. J. Loveson, P. Chandra, and M. Kotha, "Evaluation and comparison of the earth observing sensors in land cover/land use studies using machine learning algorithms," *Ecol. Inform.*, vol. 68, 2022, Art. no. 101522. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954121003137

[69] M. U. Rasheed and S. A. Mahmood, "A framework base on deep neural network (DNN) for land use land cover (LULC) and rice crop classification without using survey data," *Clim. Dyn.*, vol. 61, pp. 5629–5652, 2023, doi: 10.1007/s00382-023-06874-9.

[70] M. Digra, R. Dhir, and N. Sharma, "Land use land cover classification of remote sensing images based on the deep learning approaches: A statistical analysis and review," *Arab. J. Geosci.*, vol. 15, no. 10, 2022, Art. no. 1003, doi: 10.1007/s12517-022-10246-8.

Yuqi Lei received the B.S. degree in remote sensing science and technology in 2024 from China University of Geosciences, Wuhan, China, where she is currently working toward the master's degree in photogrammetry and remote sensing.

Her research interests include change detection, image processing, laser scanning, and their applications in smart cities and digital twins.

Hui Cao received the B.S. degree in geographical information science from Central South University, Changsha, China, in 2022. He is currently working toward the master's degree in software engineering with China University of Geosciences, Wuhan, China.

His research interests include deep learning, 3-D change detection, image processing, and LiDAR point cloud processing.

Xuanyan Zhou received the B.S. degree in remote sensing science and technology from China University of Geosciences, Wuhan, China, in 2024. He is currently working toward the master's degree in resource and environment with Sun Yat-sen University, Guangzhou, China.

His research interests include related applications such as digital image processing, hydrological remote sensing, photogrammetry, etc.

Jon P. Mills received the B.Sc. (Hons.) degree in surveying science and the Ph.D. degree in digital photogrammetry from Newcastle University, Newcastle upon Tyne, U.K., in 1993 and 1996, respectively.

He is currently a Professor of geomatic engineering with the School of Engineering, Newcastle University, where he leads the UKRI Centre for Doctoral Training (CDT) in Geospatial Systems. His research interests include developing optimized geospatial solutions for measurement, mapping, monitoring, and parameterization of engineering models.

Dr. Mills is a Fellow of the International Society for Photogrammetry and Remote Sensing (ISPRS, 2022) and has previously served as an elected Member of ISPRS Council (2012–2016), as President of ISPRS Technical Commission V on close range image measurement techniques (2008-.2012), and as Chair of ISPRS Working Group I/4 on airborne digital sensor systems (2004–2008). During the period 2017–2023, he was the Chair of Commission 1, primary data acquisition, for European Spatial Data Research (EuroSDR).

Wen Xiao (Member, IEEE) received the B.S. degree in geodesy and geomatics from Wuhan University, Wuhan, China, the M.S. degree (*cum laude*) in geoinformatics from the Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands, and the Ph.D. degree in geoinformation science and technology from the National Institute of Geographical Information and Forestry (IGN), Université Paris-EST, Paris, France, in 2015.

He has been a Lecturer and a University Research Fellow with Newcastle University, Newcastle upon Tyne, U.K. He is currently a Professor with the China University of Geosciences, Wuhan, China. His research interests include 3-D mapping, laser scanning, photogrammetric computer vision, and their applications in smart cities and digital twins.

Dr. Xiao is also a Council Member of the Remote Sensing and Photogrammetry Society (RSPSoc), U.K.