The application of big data in the analysis of the impact of urban floods: A case study of Qianshan River Basin

Weiwei Shao 1, Xin Su 1, Jie Lu 2, Jiahong Liu 1*, Zhiyong Yang 1, Yongqiang Cao 2, Zhaohui Yang 1 and Kaibo Wang 1

1 State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China
2 School of Geographical Sciences, Liaoning Normal University, Dalian 116029, China
*Corresponding author’s e-mail: liujh@iwhr.com

Abstract. In recent years, with the acceleration of climate change and urban modernization, the flood risk of economic and social development has increased. The problem of urban flood disasters needs to be solved urgently. It is particularly important to conduct flood disaster loss assessment research for further flood control and disaster reduction, and emerging big data have opened a new direction for solving the problem of urban flooding. In this study, the Qianshan River Basin in Guangdong Province was used as an example. The TELEMAC model was used to simulate the inundation process under the design rainstorms of 50a and 100a in the Qianshan River Basin. Based on the heat map service, route planning service, and search service functions of Baidu Maps and Amap open platforms, this study used crawler technology to integrate multi-source data, including traffic survey data, geospatial data, population heat map data at different times (day/night) on weekdays and non-weekdays, and point of interest (POI) data from various industries. A dynamic assessment model of flood disaster loss for traffic, population, and POIs was constructed. The research results could greatly improve the timeliness and accuracy of urban flood disaster analysis, early warning, monitoring, and disaster risk assessment.

1. Introduction
Currently, China is in a period of rapid urbanization. The construction and development of cities and surrounding towns has profoundly changed the local natural environment, including local climatic conditions, underlying surface, building distribution, municipal and water conservancy facilities, etc. In highly urbanized areas, intense human activities have an important impact on the underlying surface of the watershed, runoff generation and drainage, and the relationship between watershed storage and discharge. This has changed the original natural hydrological conditions and increases the risk of flood disasters [1]. At the same time, due to the high concentration of population and industries in highly urbanized areas and high economic density, the economic loss and social impact caused by urban flooding disasters have also increased at an unprecedented rate.

The essence of flood disaster loss is the socio-economic loss caused by flood disasters, including direct and indirect losses [2]. Generally, the direct loss is the material loss directly caused by flooding, a loss of a material object. Indirect loss is the economic loss caused by the obstruction or interruption of production and service activities, and it is caused by or derived from direct economic loss. There are two types of flood disaster loss assessment methods in foreign countries: (1) the disaster loss-hazard factor relationship method, which establishes the relationship between flood disaster loss and water depth, duration, and other hazard factors. For example, the United States proposed the non-traditional...
water depth-loss curve method [3]; (2) use of extreme value theory to calculate the average unit loss method; for example, the Netherlands established the corresponding relationship between the type of inundated asset loss and statistical type and queried the maximum loss value [4]. In China, the assessment of the socio-economic impact of flood disasters has started relatively late but has gradually made a technical breakthrough in the later stage, such as establishing a stochastic model of property loss assessment, remote sensing monitoring and evaluation models based on remote sensing and spatial distribution databases, flood assessment models based on flood simulations, and models of regional flood disasters [5-7]. The assessment process of flood disasters by government departments in our country is primarily based on the industry standard of the Methodology for storm flood disaster information pre-assessment (MZ/T041-2013) [8] and the industry standard of the Flood Disaster Assessment (SL 579-2012) [9]. However, owing to the difficulty of obtaining fine socio-economic data and the high demand for timeliness of data in the chain reaction of the social impact of rainstorms and flood disasters, there is still much room for improvement in the efficiency of data acquisition and evaluation (timeliness, accuracy, etc.).

In recent years, with the rapid development of information technology, the Internet, social networks, Internet of Things, cloud computing, and other technologies have become increasingly widely used in production and life, and the total amount of information data shows an exponential growth. In recent years, water conservancy informatization construction speed in China has been very fast, and long-term data records and the application of new technologies have accumulated a large amount of long series data and grid data [10]. The development of science and technology in the field of computer and social economy, such as the application of data platforms such as Baidu, Tencent, and Gaode, has expanded data services in the field of cities. The daily application of remote sensing, GIS (Geographic Information System), GPS (Global Positioning System), and other technologies has also led to a surge in the amount of data in the field of urban hydrology and has expanded the spatial and temporal scales and element types of data. The development of big data related to urban hydrology provides a new perspective for flood disaster loss assessment and optimization of flood control and disaster reduction countermeasures.

2. Study Area
As an important part of the construction of the Guangdong-Hong Kong-Macao Greater Bay Area, the Pearl River Delta region plays an irreplaceable role in promoting the rapid economic development of the country. In recent years, the Pearl River Delta region has been affected by climate change, the rainfall pattern has changed significantly, the frequency and proportion of extreme rainfall have increased significantly, and the multiple climate risks of extreme rains, floods, and typhoons have been superimposed. As one of the typical highly urbanized areas, it faces problems such as increasing flood risk, wading public emergency risks, and increasing losses. The construction of the Bay Area will inevitably put greater demand and pressure on water resources, and the future impact of climate change on floods in the Pearl River Delta city cluster will continue to deepen.

The Qianshan River Basin is located in the crossover area of Zhongshan and Zhuhai in Guangdong Province and is a highly urbanized area (figure 1). In 2020, the total resident population of the basin was 859,054. The basin is located in the water-rich region of South China and belongs to the subtropical maritime monsoon climate, with abundant rainfall, concentrated rainfall in the flood season, and mostly extreme rain with strong intensity and short duration. In recent years, with the development of urbanization, the surface impervious area has been increasing, which has changed the conditions of slope runoff production and confluence and the confluence system of river networks. The speed of rainfall runoff production and confluence has been accelerated, the peak value is larger, the peak time is earlier, and the water volume of waterlogging increases, which aggravates the flood hazard. The selection of this area as the research area of flood risk impact under climate change is typical and representative, and it has reference significance for flood disaster loss assessment in other areas in the future.
3. Data
When constructing the urban flood model, five datasets, including pipe network, land use, DEM (digital elevation model), hydrological soil groups, and rainfall, were collected in this study. The pipeline network datasets were provided by the Zhuhai Institute of Urban Planning and Design. Land use was obtained by the remote sensing explanation method. The remote sensing images were derived from the Gaofen-1 satellite remote sensing images of the China Resources Satellite Application Center (http://www.cresda.com/CN/), with a spatial resolution of 16 m. The DEM was derived from the ASTER GDEM digital elevation data product of the geospatial data cloud (http://slt.gd.gov.cn/shzh/), with a spatial resolution of 30 m. The hydrological soil group data were established from the Chinese soil data and classification criteria provided by the USDA. The design rainstorm was calculated using the rainstorm intensity equation in Zhuhai issued by the Zhuhai Meteorological Bureau (http://weather.zhuhai.gov.cn), in which the peak coefficient was 0.4, and the rainfall duration was 120 min.

In addition, based on the heat map service, route planning service, and search service functions in open platforms such as Baidu and Amap, this study used crawler technology to obtain the distribution of population and traffic speed at different times, as well as the POI (point of interest) data of various industries, providing data support for the dynamic assessment of flood disaster loss.

4. Methods
This study aimed to dynamically assess the impact of flooding on traffic, population, and POIs, and to guide the decision-making and deployment of disaster prevention and mitigation. The overall framework of this study is presented in figure 2. Firstly, design rainstorms, pipe networks, land use, CN (curve-number) values, Manning coefficients, and DEM data were collected to provide driving data for urban flood simulation. Crawler technology was used to obtain traffic speed, population heat maps, and POI data in the study area to evaluate flood damage. Second, based on the SCS (Soil Conservation Service) rainfall runoff model, one-dimensional pipe network, and two-dimensional network, a one-dimensional and two-dimensional coupling models were established, and the inundated area and inundated depth of urban floods were simulated. The corresponding flood hazard risk areas were divided based on the inundated depth. Finally, the traffic speed in the Qianshan River Basin at different times, the affected area of different heating value areas, and the number of POIs in each risk area were calculated to complete the dynamic assessment of flood disaster loss.
5. Results

5.1. The impact of floods on urban traffic

Based on the route planning service of the Baidu Map open platform, the traffic speed of the Qianshan River Basin road network was crawled in real time. Combined with the urban flood simulation results of the TELEMAC model, the urban flood inundation simulation results were spatially integrated with the urban road network under the 50-year and 100-year extreme rainfall scenarios, and the road interruption situation is shown in figure 3. The statistical results showed that the number of road interruptions accounted for 22.68% of the total number of roads in the Qianshan River Basin under the design rainstorm scenario with a 50-year return period and for 26.77% under the design rainstorm scenario with a 100-year return period.

5.2. The impact of floods on population

Based on the simulation results of inundated depth under 50- and 100-year return periods, the inundated area was calculated for different heating value areas during weekdays and non-weekdays, daytime/night, as shown in figure 4. The results showed that the area with 0-3 heat value was more affected in the daytime, and the area with 4-7 heat value was more affected at night. Compared with the results for the 50-year rainstorm scenario, the 100-year rainstorm scenario had a higher risk and severity, and the population in the high-risk and extreme-risk areas was more affected by the disaster. Regardless of weekdays or non-weekdays, if the flooding occurred at night, the population would be more severely affected.
Note: The bold font means that the value is larger during the day than during the night. The flood risk was divided based on the inundated depth, where low risk: 0-0.1m, moderate risk: 0.1m-0.5m, high risk: 0.5m-1m, and extreme risk: > 1m.

Figure 4. Statistics of the population affected by flood disasters during the day/night and the weekdays/non-weekdays under different design storm conditions

5.3. The impact of floods on urban POIs
Based on the Amap open platform, the POI data of various industries in the study area were obtained, and the POI data of each industry were superimposed with the inundated maps under the two rainstorm scenarios simulated by the TELEMAC model. The flood disaster situations of different industries under different rainstorm scenarios were statistically analyzed. According to the disaster situation, the number of disaster-affected points of interest was used as a statistical measure, as shown in table 1. It can be seen that under the two extreme rain scenarios, the industries were mainly affected by low-and moderate-risk flood disasters. The number of POIs in the high-risk and extreme-risk areas affected by flooding under the 100-year rainstorm scenario was significantly higher than that under the 50-year rainstorm scenario, with an increase of 36.06% and 77.52%, respectively. Under the two extreme rainstorm scenarios, the flood disaster ratio of the life service was the highest at low risk, and the flood disaster ratio of government agencies was the highest at extreme risk.

Table 1. Statistics on the number of POI disasters under different rainstorm scenarios

| Industry                | Quantity | 50-year return period | 100-year return period |
|-------------------------|----------|-----------------------|------------------------|
|                         | Low risk | Moderat e risk | High risk | Extreme risk | Low risk | Moderat e risk | High risk | Extreme risk |
| Government agency       | 729      | 351 330              | 31 17      |              | 329 335 | 45 20         |          |              |
| Business residence      | 2327     | 1111 1040            | 144 31     |              | 1064 1015 | 195 52        |          |              |
| The mall                | 154      | 70 77                | 7 0        |              | 68 71 15 | 0             |          |              |
| Transportation facilities| 2449     | 1044 1221            | 165 19     |              | 990 1191 | 226 42        |          |              |
| School                  | 385      | 180 179              | 21 5       |              | 173 175 | 29 8          |          |              |
| Company                 | 7082     | 3268 3300            | 407 107    |              | 3166 3239 | 489 188       |          |              |
| Catering services       | 12468    | 5800 6019            | 570 79     |              | 5616 5851 | 864 137       |          |              |
| Domestic services       | 6394     | 3083 2935            | 327 49     |              | 3001 2883 | 412 98        |          |              |

6. Discussion
At present, population disasters are primarily determined by the affected areas of different heat value intervals. Future research is expected to establish a spatialization method of population data based on a population heat map and the population size of the study area and obtain the quantitative expression of the population disaster situation of different thermal value intervals. In addition, timeliness is an important index for flood disaster assessment. The existing hydrodynamic model requires a long
calculation cycle. Future studies should consider using the simulation results of the hydrodynamic model as training samples to build an urban flood model based on a neural network to improve the timeliness of the simulation.

7. Conclusion
With the advent of the era of big data, facing flood disaster emergency monitoring, water conservancy data sources are increasingly rich, and the data collection frequency is constantly improving, which has laid a good foundation for flood disaster emergency data. Meanwhile, artificial intelligence and cloud computing are booming, constantly enriching information mining models, enhancing big data computing capacity, improving the timeliness and accuracy of data, and providing good technical support for flood disaster loss assessment. With the diversification and rapid growth of water conservancy data, many traditional water conservancy solutions cannot meet the requirements of the rapid development of water conservancy informatization. The application of big data is becoming increasingly mature, which opens a new avenue for the informatization of the water conservancy industry. Water conservancy big data solutions will bring qualitative changes to the development of water conservancy informatization. The application of big data technology in urban flood models will also become a new idea to solve the problems of urban flood analysis and early warning.

Acknowledgments
The researchers would like to extend their thanks to the Chinese National Key Research and Development Program (2018YFC1508203) and Chinese National Natural Science Foundation (Nos. 51979285, 51739011 and 51879274).

References
[1] O’Donnell E C, Thorne C R. (2020) Drivers of future urban flood risk. Philosophical Transactions of the Royal Society A, 378(2168): 20190216.
[2] Wu Z, Shen Y, Wang H, et al. (2020) Urban flood disaster risk evaluation based on ontology and Bayesian Network. Journal of Hydrology, 583: 124596.
[3] Scawthorn C, Blais N, Seligson H, et al. (2006) HAZUS-MH flood loss estimation methodology. I: Overview and flood hazard characterization. Natural Hazards Review, 7(2): 60-71.
[4] NVrisou van Eck and M. Kok. (2001) Standard method for Predicting Damage and Casualties as a Result of Floods. Delft, The Netherlands: Ministry of Transport, public Works and water Management, 22-41.
[5] Mei C, Liu J H, Wang H, et al. (2020) Urban flood inundation and damage assessment based on numerical simulations of design rainstorms with different characteristics. Science China Technological Sciences, 63(11): 2292-2304.
[6] Neal J, Dunne T, Sampson C, et al. (2018) Optimization of the two-dimensional hydraulic model LISFOOD-FP for CPU architecture. Environmental modelling & software, 107: 148-157.
[7] Li W, Lin K, Zhao T, et al. (2019) Risk assessment and sensitivity analysis of flash flood in ungauged basins using coupled hydrologic and hydrodynamic models. Journal of Hydrology, 572: 108-120.
[8] Wang Z, Liao Y, Li B, et al. (2013) Pre-assessment method for storm-type flood disasters (MZ/T041-2013). China Standard Press, Beijing.
[9] Tian Y, Tan X, Yao W, etc. (2012) Flood Disaster Assessment Standard (SL 579-2012). China Water Resources and Hydropower Science Press, Beijing.
[10] Liu X. (2015) Research and Application of Big Data on Analysis and Early Warning of City Waterlogging Disaster. Huazhong University of Science & Technology, Thesis for Master Degree.