Assessing Urban Areas’ Vulnerability to Flood Disaster Based on Text Data: A Case Study in Zhengzhou City

Zening Wu, Yanxia Shen and Huiliang Wang *

College of Water Conservancy and Environment, Zhengzhou University, Zhengzhou 450001, China

* Correspondence: wanghuiliang@126.com; Tel.: +86-181-3567-1702

Received: 23 July 2019; Accepted: 8 August 2019; Published: 21 August 2019

Abstract: Confronting the frequent flood disasters triggered by torrential downpour, the vulnerability of urban rainstorm flood disasters was analyzed with one highly popular area of research in mind: big data. Web crawler technology was used to extract text information related to floods from Internet and popular social media platforms. Combining these text data with traditional statistical data, a flood disaster vulnerability assessment model based on Analytic Hierarchy Process (AHP) was established to evaluate rainstorm and flood disaster vulnerability, and the spatial distribution characteristics of vulnerability to pluvial flooding were analyzed based on Geographic Information System (GIS). The established model was applied in Zhengzhou, a city that often suffers from heavy rainstorms. The results show that the areas located near downtown Zhengzhou were more vulnerable to rainstorm and flooding than others, and most of the city could be at moderate and high vulnerability. Finally, the waterlogging spots extracted from various sources were used to evaluate the performance of the proposed model. The results show that most of waterlogging spots were located in very-high and high risk zones, while less waterlogging spots were found in districts with low vulnerability, which demonstrates the discriminative power of the established model based on big data sources. This study overcomes limited data in flood disaster vulnerability assessment methods and provides a basis for flood control and management in cities.

Keywords: urban flood disaster; vulnerability assessment; big data; web crawler technology; AHP; GIS

1. Introduction

Over the past few decades, changes in land use patterns, population explosion, and paving and water storage space, caused by demographic, economic, political, and/or cultural mutations, have had notable effects on rainstorms and flooding. Consequently, flood disaster has become a challenging issue, threatening the security of society and impairing economic development in cities. Flood disaster vulnerability, a function of the character, magnitude, and rate of climate variation to which a system is exposed, its sensitivity, and its adaptive capacity, contributes in a major way to the management of urban flood disaster [1,2]. Given this wide range, the difficulty of flood disaster control, and complex uncertainties, flood disaster vulnerability estimation has also become the central issue of international urban hydrology and scientific damage research. Exactly how to assess the flood disaster vulnerability is key to implementing flood management practices in cities.

Essentially, three methodologies were discussed to assess the vulnerability of rainstorm flood disasters: vulnerability assessment based on historical disaster data [3,4], evaluation based on indicator system [5,6], and scenario simulation based on hydrologic-hydraulic model [7,8]. The method of establishing an evaluation model based on indicator system is preferred by many researchers.
For example, the fuzzy logic method was adopted to perform a more objective analysis of flood risk by deriving quantitative and accurate indicators (i.e., land price, underground area index, floor area ratio, a decline of building and the material of buildings) in literature [9]. Important factors including land cover, population density, slope, soil drainage class, drainage density, elevation and rainfall were identified to predict Flood Vulnerability Index (FVI) in literature [10], and the relationships between these factors were captured and quantified using Bayesian Belief Network [10]. Ouma and Tateishi [11] applied analytic hierarchy process (AHP) to assign the weights of the flooding causative factors such as rainfall distribution, elevation and slope, drainage network and density, land use and soil type to predict urban flood vulnerability, and similar method was also used in studies of other catchments [12,13]. However, those models typically require various types of dataset, such as rainfall data, geographic data, vegetation of catch basin and socio-economic data, among others to set up and calibrate the model. Moreover, the abilities of evaluation models are limited by data and cognition limitations, such as incomplete understanding of the processes involved and inaccuracies in model formulation, invalid values of indicator parameters, and inadequate or erroneous information required for the model applications including input and calibration data. Unfortunately, there is only sparse data sampling in most urban areas, and most of these groups do not have ready access to modeling expertise and data collection [14,15]. For example, most of models are based on statistical data recorded in literature for flood vulnerability assessment, and few real-time big data sources, such as picture, video and text data collected from the Internet and social media platforms are applied [16,17]. Thus, the uncertainty for input parameters may render model prediction unreliable. The underlying question is how to properly estimate the flood disaster vulnerability in cities with real-time and abundant data availability.

In fact, the recent increase in availability of earth observation data, coupled with advances in machine learning algorithms, have expanded the possibilities and use of big data in rainstorm and flood disasters. Big data have been used extensively in rainstorms and flooding, such as urban waterlogging risk assessment [14], modeling urban coastal flood severity [18], hyper-resolution monitoring of urban flooding [19], and real-time identification of urban rainstorm waterlogging disasters [20]. This is currently significant as multiple platforms now exist for collecting multisource data regarding urban flooding. And it can be expected that, due to the nearly universal use of Internet connected devices, data and information related to flood disasters will continue to grow in volume. The development of big data serves as a first step for urban flood disaster risk assessment.

The objective of this study is to estimate flood disaster vulnerability in cities by constructing a flood disaster vulnerability assessment model based on text data to solve the uncertainty of model input. First, web crawler technology was used for extracting valuable data and information from Internet web pages and social media platforms (i.e., Weibo and WeChat). Then, a model was constructed based on these data to assess flood disaster vulnerability in Zhengzhou, a city often suffering from heavy rainstorms. Finally, the waterlogging spots extracted from Internet web pages, Weibo and WeChat using web crawler technology were compared with the results of vulnerability assessment to evaluate the performance of the established model.

2. Materials and Methods

2.1. Study Areas and Data Sets

Zhengzhou, a city in north-central Henan Province, China, located between 112°42’ and 114°14’ eastern longitude and between 34°16’ and 34°58’ northern latitude (Figure 1), comprises an area of approximately 1010 km2. It has flat terrain, small elevation fluctuations and abnormal monsoon activity, making it a potentially high-risk region for flood disaster and one of the most intensive flood control towns. The area is characterized as lowland with an average altitude of 50 m, from the southwest to northeast, stepped down. The average slope in most of the wards (70% of the total area) is below 5° and only 2.2% of the districts have slope higher than 10°. The selected region is in a temperate
continental climate with a mean annual precipitation of 625.9 mm. The flood season, a period of frequent rainstorm and flood disasters, spans from July to September every year, during which the rainfall accounts for 60~70% of the total annual rainfall. According to the statistics, Zhengzhou has suffered heavy rainstorms more than 15 times per year since 2006 and each time a flood disaster has caused more than 30 million dollars in economic losses.

The associated data sources utilized in this research include text data and statistics data. For statistics data, the DEM (Digital Elevation Model) for the case study was downloaded from the Geospatial Data Cloud Service Platform and had a 30×30 m horizontal resolution. The land use data was developed using Spot 5 imagery of 20 m resolution for the year 2018 acquired from the Data Sharing Infrastructure of Earth System Science. The population data at the end of 2018 was provided by Zhengzhou Statistical Yearbook and processed using ArcGIS. For text data related to rainstorms and floods, it can be obtained from Weibo, WeChat and web pages by using key words including Zhengzhou, flooding, rainstorm and disaster in certain search engine. Finally, the duplicate information was eliminated through rapid reading, and useful data could be extracted using web crawler technology that can be introduced in the next parts.

2.2. Web Crawler Technology

As one of the most widely applied, sourced and largest big data, text data is considered as the most common type of information storage, mainly from mainstream social media platforms as Weibo, WeChat and various Internet websites, which can be developed, processed, stored, and organized according to the specific demands of users and the corresponding internet protocol, rule and frame [14,20,21]. E-mails, Internet web pages, electronic medical cases and operation logs of various systems are all presented in the form of text, which makes text data of great commercial potential. The text data was extracted from various sources using web crawler technology, a method of automatic collection of required information from one or more pages based on certain strategy and way to access network resources via a simulated browser [22,23]. Web crawler technology can be considered as a program or bot that traverses web pages and create an index of the pages. The primary objective of web crawler is to grab web information and extract required data information through simulating browser access.
to web resources. Therefore, users can quickly search for the content they need [24]. The general architecture of the web crawler technology was detailed in Figure 2.

![Flow chart of web crawler technology.](image)

Note: URL (Uniform Resource Locator) is the location and address for information access on the Internet. Internet addresses are also known as URLs. One URL corresponds to a unique web page that contains a lot of information related to flood disaster, while a web page may have more than one URL.

As shown in Figure 2, the steps for getting text information related to flood disaster were as follows:

1. According to the expertise in flood disaster domain and the identified theme, initial seed URLs were identified and determined, which made up an original crawled queue.
2. A URL was selected from the queue to extract information related to flood disaster from the original crawled queue. Based on this URL, a satisfactory web page was obtained after checking whether the page content is in available.
3. The web page obtained in the step (2) was processed. If there is only one URL in this web page, the data information related to flood disaster was extracted and stored in certain format. When there is more than one URL in this web page, URLs related to flooding were extracted and further processed, which is added to the initial URL queue for data extraction.
4. The process was repeated until no URL is left in the original crawled queue.

2.3. The Development of Flood Disasters Vulnerability Assessment Model

2.3.1. Criteria Standardization and Weighting

Flood disaster vulnerability is the product of disaster drivers, disaster-formative environment and disaster bearers [6,10]. Disaster drivers are defined as extreme (severe and prolonged) weather events
that adversely affect human life, property, security, etc. In urban flood disasters, the disaster drivers primarily refer to heavy rain expressed by rainfall duration and accumulated rainfall. Disaster-formative environment refers to the conditions and surroundings where flood disasters occurred, determined by formative factors that mainly result from the combination of both climate variables and underlying surfaces. In this work, disaster-formative environment mainly includes elevation and average slope. Disaster bearers depend on the exposure (number of people and infrastructure impacted) and the adaptability and resiliency of the affected area. Therefore, population density and land cover are considered as measuring factors of disaster bearers. From the analysis above, according to the review of relevant literature as well as the field surveys and interviews carried out in the scope of this research, the most relevant factors for flood disaster vulnerability analysis in this study area were determined and defined. Six criteria-parameters, including rainfall and duration, elevation and slope, land use and population density were included in the evaluation model, which fit the conditions and the flood characteristics well. Main data are climatological data, topographic map, land cover and socio-economic data. All the data for this assessment are available.

The weights of various factors were defined following the Analytic Hierarchy Process (AHP), a semi quantitative method that has a partly subjective nature [13]. AHP is commonly used and widely accepted in decision making, which integrates the ideas of ranking and weighting with the knowledge of experts [13]. One of the emphases of applying AHP is to avoid the special nature of even weighting [25]. This approach requires experts’ best judgments to the relative importance of one criterion index against another in the flood disaster vulnerability model. Criterion weights are calculated using a preference matrix where all identified relevant indices are compared against each other in pairs. Then the weights can be aggregated based on the different weighted combination methods for the derivation of evaluation results [26]. The rating employs an underlying semantic scale which has values representing the member of the set: {9, 8, 7, 6, 5, 4, 3, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9}, with 9 representing absolute importance and 1/9 the absolute triviality [27].

Following the calculation of the weights of indicators using AHP, its consistency needs to be evaluated, which was calculated using the following equation:

\[
CR = \frac{CI}{RI}
\]

where CR is the consistency ratio; RI is the random index, which are given in specific tables based on the number of comparative variables, as shown in Table 1; CI is the consistency index which can be calculated using Equation (2).

\[
CI = \frac{\lambda_{\text{max}} - n}{n - 1}
\]

where \(\lambda_{\text{max}}\) is the maximum eigenvalue of the comparison matrix; \(n\) is the number of factors (\(n = 6\) in this study). According to AHP’s theory, the constructed discriminant matrix is considered as reasonable, if the CR is < 0.1.

<table>
<thead>
<tr>
<th>(n)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>RI</td>
<td>0</td>
<td>0</td>
<td>0.58</td>
<td>0.9</td>
<td>1.12</td>
<td>1.24</td>
<td>1.32</td>
<td>1.41</td>
<td>1.45</td>
</tr>
</tbody>
</table>

In this study, since flood disaster vulnerability is related to disaster drivers, disaster-formative environment and disaster bearers, these three indices were considered as the first level. Various factors corresponding to these three indices, i.e., rainfall and duration, elevation and slope, land use and population density, were regarded as the second level. The significance of different factors was determined by experts. According to the importance of different factors, the weight of each index was determined and normalized using MATLAB computational procedure.
2.3.2. Flood Disasters Vulnerability Assessment Model

In this study, flood disaster vulnerability evaluation was defined as a weighted sum of all relative indicators, which is a compensatory aggregation function in the AHP tool. Based on the indices defined above, the flood risk assessment model was established as follows:

\[ V = \sum_{i=1}^{n} w_i I_i = w_1 I_1 + w_2 I_2 + w_3 I_3 + w_4 I_4 + w_5 I_5 + w_6 I_6 \]  

(3)

where \( V \) is the vulnerability of flood disaster; \( I_i \) is the influencing index of flood disaster; \( w_i \) is the weight of the \( i \)th criterion index, and \( \sum_{i=1}^{6} w_i = 1 \); \( n \) is the number of indices applied in this study.

2.4. GIS Tools

GIS software, ArcMap 10.3 tools are used for data collection, processing and result visualization [28]. GIS is an appropriate tool for processing spatial data with attributes for deriving regional indicators on flood disaster vulnerability, which is an important step in effective risk assessment [29]. GIS-based vulnerability assessment methods are a suitable method to incorporate all relevant types of consequences [10]. Therefore, in this study, data related to flood disaster needed to be further processed before it could be used as input for the vulnerability evaluation model. Data was processed using ArcMap version 10.3. The vector datasets were rasterized. All datasets were then projected, resampled to a 200 m grid cell, clipped to the study area and registered, so all input grids accurately overlaid with the same projection, cell size and extent. The study area has been classified into 24,775 grid cells. Except for zones near the city’s boundary, all grid cells have an area of 0.04 square kilometers.

3. Results and Discussions

3.1. Distribution Characteristics of Disaster Drivers

Rainstorms and floods were caused by frequent and short rainstorms in Zhengzhou City. As the aforementioned, the dangerousness of disaster drivers was formed by accumulated rainfall and duration. Accumulated rainfall is defined as the total amount of rainfall during a flood, and rainfall duration that is the number of days from the onset of a rainstorm to the end of a flood. Therefore, these two factors were calculated by the hierarchical analysis and the weighted synthesis method to express the dangerousness of disaster drivers. The spatial distribution characteristics of the dangerousness of disaster drivers in Zhengzhou were shown in Figure 3. Specifically, the dangerousness of disaster drivers in Zhengzhou City was concentrated in the middle and eastern part of the city, such as the west of Jinshui District and Zhengdong New District, while the risk in western areas is weak, which is particularly evident for Erqi District and Zhongyuan District. Overall, the dangerousness of the disaster drivers decreases from northeast to southwest. It is concluded that the northeast areas are more vulnerable to rainstorms and floods. The distribution of risk of disaster drivers is similar to other study in this area [12].

Flood disaster vulnerability is related to the rainfall in each region. Assuming the same rainfall intensity of one rainstorm in each district which means the same dangerousness, urban flood disaster vulnerability was evaluated using the product of the disaster-formative environment and disaster bearers.
3.2. Sensibility of Disaster-Formative Environment

Topography has a significant influence on flood formation and redistribution. Therefore, elevation and slope were regarded as the influencing factors of terrain in flood disasters. Elevation is commonly represented by the vertical distance from certain surface to the reference basement. Slope is a measure of the average rate of change of elevation in a given domain. These two terrain factors were shown in Figure 4. A surface with a steeper slope has lower possibility of being inundated because the flood can be easily drained towards down-slope. As demonstrated in Figure 4a, most of Zhengzhou has a flat terrain, ranging gradually from 27 m to 340 m, decreasing progressively from southwest to northeast. The average slope of the seventy percent of the total area is below $5^\circ$, and only $2.2\%$ of the districts have slopes higher than $10^\circ$ in Figure 4b.

Sensibility of disaster-formative environment was formed by elevation and slope, which was divided into five grades: very-low, low, moderate, high and very-high, as shown in Figure 4c. Most areas of Zhengzhou City are in very-low and low risk, and high-risk zones are scattered in the eastern and northern city. Compared with other areas, south and west parts of the city have a higher elevation and slope, suggesting that these areas are not more sensitive than others in Zhengzhou City.

3.3. Exposure of Disaster Bearers

The likelihood of pluvial flood occurrence was not only related to the disaster drivers and disaster-formative environment, but also to the resilience and adaptability of this region, such as population density and land use conditions. Land use types in Zhengzhou City were mainly classified into building lands, grasslands, forest lands, roads and waters. And the area of various land types was summarized in Table 2. From this table, the area of forest and build are the largest, accounting for $39\%$ and $38\%$ of the total area in Zhengzhou City, respectively, while the grasslands, roads and waters are at less area. Buildings are mainly comprised of paved surfaces and impervious area. When heavy rains were experienced, the greater the ratio of buildings, the more disaster losses. It can be seen from Figure 5a that compared with the surroundings, areas located near downtown Zhengzhou are commercial center with many buildings, thus resulting in many impervious surfaces and low permeability. Furthermore, flood disaster vulnerability is tied to the urbanization process in the context of city-dwelling population. Therefore, population density, the number of inhabitants of per square kilometer, was used to express resilience and adaptability of different districts in the city of Zhengzhou, as shown in Figure 5b. Wards located near downtown Zhengzhou have the highest population density with more than 6000 inhabitants per square kilometer while areas near the city periphery have relatively low population density.
Figure 4. The map of sensibility of disaster-formative environment: (a) elevation; (b) slope and (c) spatial distribution of sensibility to rainstorms and floods.
Table 2. The area of different land types in Zhengzhou City.

<table>
<thead>
<tr>
<th>Land Types</th>
<th>Build Lands</th>
<th>Grass Lands</th>
<th>Forest Lands</th>
<th>Roads</th>
<th>Waters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (km²)</td>
<td>441.84</td>
<td>116.86</td>
<td>434.68</td>
<td>103.11</td>
<td>36.19</td>
</tr>
</tbody>
</table>

Figure 5. Cont.
Exposure of disaster bearers formed by land use and population density was shown in Figure 5c, which was also divided into five grades. According to this figure, most wards could be at very-high and high exposure. Districts near downtown Zhengzhou recorded high exposure about the disaster bearers, meaning that when rainfall having same intensity occurred, the damage caused by heavy rainfall in these areas was higher than that in others. The north parts of this city are new city zone with low population density, therefore, the risk is lower than others.

3.4. Comprehensive Analysis of Flood Disaster Vulnerability

In this study, the comparison of importance of different factors influencing flood disaster was built by the integration of experts’ experience, comprehensive literature review and professional expertise, as shown in Tables 3 and 4. The weights of all indictors were then identified and determined based on this, as shown in Table 5. From this table, population density was assigned the smallest weight, partly due to their almost uniform values across the case study, which has an unnoticeable impact on the vulnerability evaluation result. Slope and elevation are not only essential factors, but also reliable criteria derived from the high-resolution DEM. They can finely discriminate land units to delineate areas of different risk levels for a detailed assessment. These two indices accordingly received relatively higher weights. It is also noted that rainfall and duration received fairly high weights because they are the major drivers and contribute most to the flood disaster.

<table>
<thead>
<tr>
<th>Criterion Name</th>
<th>Disaster Drivers</th>
<th>Disaster-Formative Environment</th>
<th>Disaster Bearers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disaster drivers</td>
<td>1</td>
<td>9/7</td>
<td>9/6</td>
</tr>
<tr>
<td>Disaster-formative</td>
<td>7/9</td>
<td>1</td>
<td>7/6</td>
</tr>
<tr>
<td>environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disaster bearers</td>
<td>6/9</td>
<td>6/7</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 4. Pair-wise comparison matrix for different factors influencing flood disaster.

<table>
<thead>
<tr>
<th>Criterion Name</th>
<th>Influencing Factors</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disaster drivers</td>
<td>Rainfall</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>9/8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Disaster-formative environment</td>
<td>Elevation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>7/5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Disaster bearers</td>
<td>Land type</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Population density</td>
<td>6/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5. The weights of different factors influencing flood disaster.

<table>
<thead>
<tr>
<th>Disaster Drivers</th>
<th>Weights</th>
<th>Disaster-Formative Environment</th>
<th>Weights</th>
<th>Disaster Bearers</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>0.3065</td>
<td>Elevation</td>
<td>0.1705</td>
<td>Land type</td>
<td>0.129</td>
</tr>
<tr>
<td>Duration</td>
<td>0.2333</td>
<td>Slope</td>
<td>0.1254</td>
<td>Population density</td>
<td>0.0352</td>
</tr>
</tbody>
</table>

Based on the weights summarized and listed in Table 5, flood disaster vulnerability of each district in Zhengzhou City was calculated according to the established vulnerability assessment model in Equation (3) using ArcMap version 10.3. The results of vulnerability evaluation of flood disaster were classified into five classes: very-low, low, moderate, high and very-high, and the flood hazard vulnerability map was created and shown in Figure 6.

![Figure 6](image-url)
that were covered by rainstorms, since this criterion index received the highest weight. From this pattern, the downtown areas could be at high risk, including most areas of Jinshui District, east parts of Zhongyuan District and north of Erqi District and Guancheng District, while the southwest and northwest areas that are new city zone with low sensibility of disaster-formative environment and exposure of disaster bearers have low risk. Overall, it can be recognized that the flood disaster risk of Zhengzhou City decreases from the middle to the surroundings, which has a high fit with the historical recorded flood events. The results have highlighted the usefulness of the constructed model, because it was able to identify a high risk area in the case study. Therefore, strategies to cope with and prevent flood hazard should be proposed to reduce losses caused by heavy rainfall according to these results. This can be achieved by employing more advanced methods for forecasting rainfall, encouraging the application of new technology, such as radar and remote sensing to improve forecast accuracy and lead time. In addition, it is necessary to facilitate economic development and enhance the individual and collective adaptability to reduce the losses caused by flooding.

3.5. Model Validation

The waterlogging spots interpreted from web pages, Weibo and WeChat using web crawler technology were shown in Figure 7. From this pattern, waterlogging spots were mainly located in the downtown areas that were considered as old town center in Zhengzhou with low pipeline design standard, such as the west regions of Jinshui District, the eastern areas of Zhongyuan District, and the north of Erqi and Guancheng District. There were 23 water accumulation spots in Jinshui District, which accounted for the most in Zhengzhou City, followed by Zhongyuan District with 9 waterlogging spots. The wards located at north and south of Zhengzhou have received less waterlogging spots, for example, Huiji District only had 4 water accumulation spots, partly due to the increased green area and favorable design capacity of pipeline.

![Figure 7. The map of distribution of waterlogging spots.](image)

Waterlogging spots were also caused by many factors, such as rainfall, elevation, slope and other human factors. Therefore, the waterlogging spots were compared to the results of vulnerability assessment, which provides a partial verification of proposed modeling approach. Locations of waterlogging spots are in accordance with very-high and high risk zones obtained by vulnerability assessment model, while very-low and low risk wards record less waterlogging spots. Within the area delineated as very-high risk, 39 waterlogging spots were found, accounting for 67% of the
total water accumulation spots. Eighteen waterlogging spots were situated in the areas recording moderate risk and only 1 waterlogging spot was in low risk zones, which spreads in Guancheng District. The remaining errors could be explained by the exposure of disaster bearers introduced in the process of assessing flood disaster. Due to the non-concentrated population and less paved surface, the flood disaster vulnerability is lower than expected. Finally, it was concluded that the locations of waterlogging spots have high fit with risk zones in the case study calculated by the proposed model, suggesting that this method is more suitable for flood disaster risk vulnerability evaluation.

Through the above analysis, one explanation for the high risk in downtown Zhengzhou may be due to the higher accumulated water level. The central regions of the study area have low-relief terrain, low elevation and many impermeable areas, which make stormwater drainage problematic. Additionally, the clogged stormwater inlets and undersized stormwater pipes can increase the chance and severity of waterlogging spots. Consequently, the vulnerability of flood disasters can be mitigated by analyzing the strategy of reducing waterlogging spots, which is one of the advantages that traditional evaluation methods do not have.

3.6. Comparisons with Other Methods

Previous studies about flood disaster vulnerability assessment are mostly based on historical data, rather than an emerging big data. Lin et al. proposed a flood disaster vulnerability evaluation method based on historical data in Zhengzhou City [12]. The results show that the flood disaster vulnerability in central and eastern regions appears to be high, which is in agreement with the distribution of risk in the present study. However, there are incomplete data in this assessment with lacking effective verification methods in [12], which hampered model performance and cannot provide effective suggestions and measures for risk assessment and management of urban flood disasters. Different regions have different conditions, and the factors influencing flood disaster are also not equal. For example, the climate exposure, sensitivity and the adaption ability were considered to assess the vulnerability of each industrial park in [2]. Park and Lee [9] intended to contribute to the reduction and minimization of flood damage in the case of heavy rainfall in urban areas based on some socio-economic indicators, such as land price, underground area index and floor area ratio. However, these methods can only be carried out using historical data, rather than the multisource information regarding urban flood, which hampered the performance of the proposed model. In addition, the lack of effective analysis and verification were another limitation of these methods. In this study, the flood disaster evaluation was proposed by combining historical data and text data, and the performance of proposed model were determined based on waterlogging spots obtained from various sources in real time. As a result, the areas that are more vulnerable to rainstorm were identified, and the reasons for increased flood disaster risk were also determined, which can facilitate planners and managers to prioritize the areas with high risk and prepare responding preventive measures for more efficient flood management.

It is expected that the methodologies proposed in this paper would better predict such flood events with a larger, more complete dataset, containing more instances of similar flooding events. This is a first attempt in the long-term goal of urban flood disaster prediction to assist city managers in real-time flood adaption measures. However, there is much room for further improvement. A main limiting factor in building accurate models is the quantity and quality of the data, due to the imperfect data extraction technology. Besides, the index system established in this study was not comprehensive, and more factors influencing flood disaster must be included in the future.

4. Conclusions

In this paper, a model was constructed based on traditional statistical data and text data to evaluate flood disaster vulnerability. Web crawler technology was used to extract text data from web pages, Weibo and WeChat. The case study revealed that the flood disaster vulnerability decreases from downtown to surroundings, and the east parts of the city are more vulnerable to flooding than western. The waterlogging spots interpreted from Internet and social media platforms were used to evaluate
the performance of the proposed method. The results show that the locations of waterlogging spots coincided well with high risk zones obtained from the proposed model, while no waterlogging spot was found in very-low risk wards, which demonstrates the discriminative power of the established model. The successful application of this method in Zhengzhou provides a reference for vulnerability assessment of flood disasters in other areas, to some degree.

Since vulnerability analysis of flood disasters using big data is only a recent topic of research, there is much room for our study to be improved in the future. For example, the way and technology to extract data from Internet and social media platforms need to be improved and perfected, and more formats of big data, such as pictures, videos, could be used for flood disaster vulnerability evaluation and risk analysis in the future works.

Author Contributions: Z.W. developed the original idea and contributed to the research design for the study. Y.S. was responsible for data collecting. H.W. provided guidance and improving suggestion. All authors have read and approved the final manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 51739009.

Acknowledgments: The study is funded by the National Natural Science Foundation of China (No: 51739009). The authors thank the anonymous reviewers for their valuable comments. The authors declare that there is no conflict of interest regarding the publication of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References


27. Saaty, T.L.; De Paola, P. Rethinking design and urban planning for the cities of the future. *Buildings* 2017, 7, 76. [CrossRef]


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).