

# Quantitative assessment of urban flood disaster vulnerability based on text data: case study in Zhengzhou

Zening Wu, Yanxia Shen, Huiliang Wang and Meimei Wu

## ABSTRACT

Planning to evaluate flood disaster vulnerability is a crucial step towards risk mitigation and adaptation. In this study, the vulnerability curve model was established 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 social media platforms. Based on the three indicators of rainfall intensity, duration and coverage area, the heavy rainfall index was calculated, while the comprehensive disaster index was calculated based on the affected population, area and direct economic loss. Taking the heavy rainfall index as an independent variable and comprehensive disaster index as a dependent variable, the vulnerability curve of flood disasters was established, and the performance of this model was validated by comparing it with real-life situations. The results show that the relationship between rainfall and disaster is significant, and there is exponential correlation between the heavy rainfall index and comprehensive disaster index. This model is more than 65% accurate, which demonstrates the discriminative power of the established curve model. The results provide some basis for flood control and management in cities.

**Key words** | big data, flood disasters, vulnerability curve, web crawler technology

Zening Wu  
Yanxia Shen  
Huiliang Wang (corresponding author)  
Meimei Wu  
College of Water Conservancy and Environment,  
Zhengzhou University,  
Zhengzhou, Henan, 450001,  
China  
E-mail: wanghuiliang@126.com

## 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 (Yin *et al.* 2015; Ryu *et al.* 2016). 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 (Yoon *et al.* 2015). Exactly how to assess flood

disaster vulnerability is key to implementing flood management practices in cities.

Essentially, three methodologies were discussed to assess the vulnerability to flood disasters: vulnerability assessment based on historical disaster data (Boudou *et al.* 2016), evaluation based on an indicator system (Chen *et al.* 2015), and scenario simulation based on a hydrologic–hydraulic model (de Moel & Aerts 2011). Ouma & Tateishi (2014) applied the analytic hierarchy process (AHP) to assign the weight for attributes of decision-making parameters, and similar methods have been used in studies of other catchments (Chen *et al.* 2015; Lin *et al.* 2016). The method of flood disaster evaluation based on hydrologic and hydraulic models can derive the flood range of urban waterlogging, submerged depth, and submerged duration, caused by different scenarios of rainfall, by using the watershed runoff model and

numerical simulation of the flood routing model (Gori *et al.* 2018). However, in general, it is evident that those methods typically require various types of dataset, such as rainfall data and socio-economic data, among others, to set up and calibrate the method, and the abilities of those techniques are limited by data and cognition limitations, such as incomplete understanding of the processes involved and inaccuracies in model formulation, invalid values of model parameters, and inadequate or erroneous information required for model applications including input and calibration data. Unfortunately, there is always only sparse data sampling in most urban areas, and most of these groups do not have ready access to modeling expertise and data collection (Lin *et al.* 2018). For example, most of the models are based on statistical data recorded in the 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 (e.g. Weibo and WeChat), are applied (Ahmad *et al.* 2017; Zhang *et al.* 2018). Thus, the uncertainty for input parameters may render model prediction unreliable. The underlying question is how to properly estimate flood disaster vulnerability in cities with real-time and abundant data availability.

The objective of this study is to quantitatively estimate flood disaster vulnerability by constructing a flood disaster vulnerability curve model based on text data to solve the uncertainty of model input. First, web crawler technology was used to extract valuable data and information from distributed heterogeneous platforms. Secondly, a heavy rainfall index and comprehensive disaster index were calculated by combing this data and statistical data. Then, a vulnerability curve model was constructed based on these two indices to assess flood disaster vulnerability in Zhengzhou, a city often suffering from heavy rainstorms. Finally, the model was validated based on statistical analysis of historical flood disasters.

## MATERIALS AND METHODS

### Study area and datasets

Zhengzhou, a city in north-central Henan Province, China, is located between 112°42' and 114°14' eastern longitude

and between 34°16' and 34°58' northern latitude. 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 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 traditional statistics. For traditional data, the flood-affected area was developed using Spot 5 imagery of 20 m resolution acquired from the Data Sharing Infrastructure of Earth System Science between 2010 and 2018. The population and economic data were provided by the Zhengzhou Statistical Yearbook from 2010 to 2018. Text data related to rainstorm and flood can be obtained from Weibo, WeChat and the Internet by using key words, including Zhengzhou, flooding, rainstorm and disaster in certain search engines. Finally, the duplicated information was eliminated through rapid reading and useful data was extracted using web crawler technology that will be introduced in the following sections.

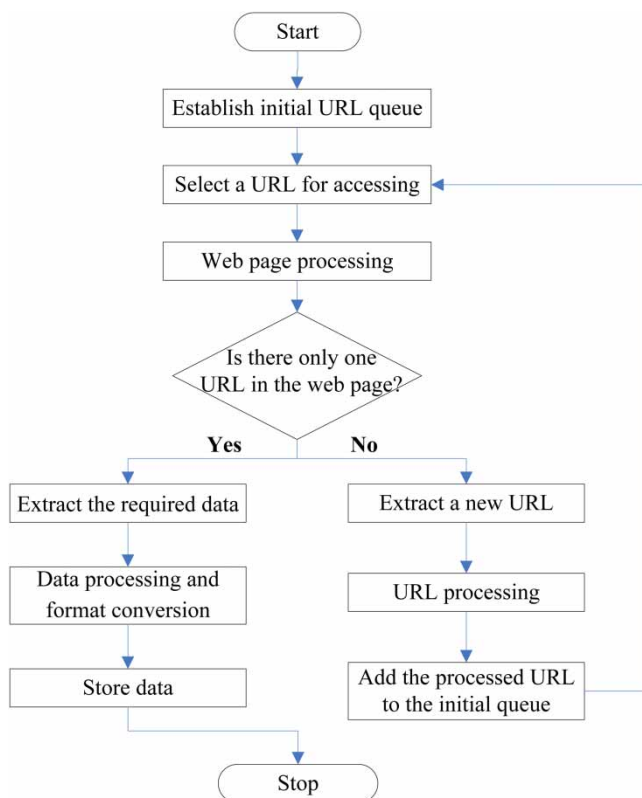
### Web crawler technology

As one of the most widely applied, sourced and the largest types of 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 (Eilander *et al.* 2016; Lin *et al.* 2018; Xiao *et al.* 2018). 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 a certain strategy and way to access network resources via a simulated browser, according to the principle which is detailed in Figure 1 (Weng et al. 2019).

As shown in Figure 1, the steps for getting text information are as follows:

- (1) According to the expertise in the flood disaster domain and the identified theme, initial URLs were identified and determined, which made up an original crawled queue.
- (2) A URL was selected to extract information related to flood disaster from the original crawled queue. Based on this URL, a satisfactory web page was obtained.
- (3) The web page obtained in step (2) was processed. If there was only one URL in this web page, the data information related to flood disaster was extracted and stored in a certain format. When there was more than one URL in this web page, URLs related to flooding were extracted and processed, which were added into the initial URL queue for data extraction.



**Figure 1** | Flow chart of web crawler technology. Note: URL (Uniform Resource Locator) is the location and address for information access on the Internet. A unique URL corresponds to one web page that contains a lot of information related to flood disaster, while a web page may have more than one URL.

- (4) No different steps were conducted until the execution reached the last URL address or satisfied the established requirements.

### Vulnerability curve model of flood disasters

A vulnerability curve, also known as disaster loss rate curve, is used to describe the relationship between the disaster intensity and the losses of the disaster-bearing body, which is expressed as:

$$L = f(D) \quad (1)$$

where  $D$  denotes the disaster drivers: in this paper, a heavy rainfall index (HRI) was calculated to express this variable.  $L$  is the losses caused by hazards, which is expressed using a comprehensive disaster index (CDI).

In Equation (1), HRI was calculated based on three factors: rainfall intensity, duration, and coverage. Rainfall intensity was expressed as the average daily precipitation; duration was defined as the time from the beginning to the end of heavy rainfall. The distribution of rainfall stations plays an important role in rainfall monitoring and verification of rainfall data extracted from different sources based on web crawler technology. Therefore, the coverage is expressed by the proportion of monitoring stations with rainfall intensity reaching a certain intensity to the total monitoring stations. According to the classification of rainfall intensity and its index in the Rainfall Intensity Grade, combined with the actual rainfall situation in Zhengzhou City, these indices were divided into five levels, as shown in Table 1.

Finally, the HRI was calculated based on these three factors:

$$HRI = I \times C \times T \quad (2)$$

where  $I$ ,  $C$  and  $T$  are respectively the intensity index, coverage index, and duration index. According to the real rainfall situation of Zhengzhou and the relevant standards, the heavy rainfall process was classified into five classes, as presented in Table 2.

The proportion of affected population, affected area, and direct economic losses were selected to express the comprehensive disaster caused by rainfall. Then the CDI

**Table 1** | Value-determined criteria of various indices characterizing heavy rainfall process

Rainfall intensity (mm·d <sup>-1</sup> )	Duration (d)	Coverage (%)	Index value
≥100	4	≥80	1
[50, 100)	3	[60, 80)	2
[25, 50)	2	[40, 60)	3
[10, 25)	1	[20, 40)	4
[0, 10)	0.5	[0, 20)	5

**Table 2** | Grade division of heavy rainfall index

Heavy rainfall index	Grade	Severity
$1 \leq H \leq 25$	I	Particular
$25 < H \leq 50$	II	Severe
$50 < H \leq 75$	III	Relative
$75 < H \leq 100$	IV	Moderate
$100 < H \leq 125$	V	Slight

was calculated using grey correlation analysis to analyze any correlations between these three indicators. Grey correlation analysis is a methodology to measure the degree of correlation between indicators employing the degree of similarity or dissimilarity of development trends among factors (Yue *et al.* 2018). More about this method can be found in references by Deng (2019) and Khalaj *et al.* (2019). Given the rapid urbanization process and dense population in Zhengzhou, the same precipitation and intensity could cause more serious losses than in other regions. Therefore, based on the results calculated using grey relational analysis and the actual flood disaster situations in Zhengzhou, and according to recent works, the CDI was divided into five levels, as shown in Table 3.

The vulnerability curve model was evaluated using historical disaster data. The objective function used in

**Table 3** | Grade division of comprehensive disaster index

Levels	CDI
Severe disaster	(0.9, 1.0]
Heavy disaster	(0.8, 0.9]
Moderate disaster	(0.7, 0.8]
Small disaster	(0.6, 0.7]
Slight disaster	[0.5, 0.6]

model calibration was the probability of being completely consistent (CCP) and basically consistent (BCP), which were defined as follows:

$$CCP = \frac{n_1}{N} \times 100\% \quad (3)$$

$$BCP = \frac{n_2}{N} \times 100\% \quad (4)$$

where *CCP* means that there is no difference between simulation results and actual results, and *BCP* denotes that the difference between the simulated value and the actual value is within one level;  $n_1$  is the number of samples whose assessment results are consistent with the realistic results; and  $n_2$  is the number of samples in which the difference between the modelled value and the actual value is equal to or less than one level. For example, for a rainstorm, the class of simulated rainfall intensity is grade two, but the actual class is grade three, showing that these are not completely consistent, which should be recorded as one of the samples to be collected.  $N$  represents the total number of samples.

## RESULTS

### Heavy rainfall process

The accumulated precipitation, rainfall intensity and duration of 19 heavy rainfalls from 2010 to 2018 were compared, as presented in Figure S1 (Supplementary Data). As demonstrated in this picture, the precipitation and rainfall intensity in Zhengzhou has gradually increased in recent years, which is tied directly to the acceleration of the urbanization process in the context of the modern social economy and the increase of the city-dwelling population, which increases the frequency of sudden strong rain (Lin *et al.* 2016). The HRI is exhibited in the second column of Table 4, where the rainfall events are ranked through the CDI with reference to the studies proposed by Julien *et al.* (2010). It can be observed that among 19 rainstorms collected in Zhengzhou during 2010–2018, the majority of the comprehensive evaluation grades are II and III, whose corresponding severity grades are ‘Severe’

**Table 4** | The rank of the grade division of comprehensive disaster for 19 rainfalls during 2010–2018 in Zhengzhou

Rank	Rainfall event	Rainfall grade	CDI	Disaster grade
1	4 July 2012	I	0.8062	Heavy disaster
2	17 August 2018	I	0.7632	Moderate disaster
3	3 August 2018	II	0.7128	Moderate disaster
4	23 June 2015	I	0.7108	Moderate disaster
5	19 August 2012	II	0.7067	Moderate disaster
6	18 July 2016	II	0.7067	Moderate disaster
7	15 May 2018	II	0.7035	Moderate disaster
8	8 July 2013	II	0.6908	Small disaster
9	28 August 2017	II	0.6813	Small disaster
10	7 August 2017	II	0.6735	Small disaster
11	29 July 2014	II	0.6705	Small disaster
12	13 September 2014	III	0.6694	Small disaster
13	13 September 2011	III	0.6645	Small disaster
14	3 July 2018	III	0.6525	Small disaster
15	6 September 2010	III	0.6387	Small disaster
16	18 July 2010	IV	0.6225	Small disaster
17	25 July 2017	IV	0.6056	Small disaster
18	18 August 2017	V	0.5976	Slight disaster
19	27 May 2013	V	0.5768	Slight disaster

and ‘Relative’ respectively, indicating that the rainfall intensity of Zhengzhou has increased over the last several years. There were three unexpected rainfalls with a comprehensive level of heavy rainfall of I, which received a severity grade equivalent to ‘Particular’. In addition, eight rainstorms had a comprehensive class of II that received a severity grade equivalent to ‘Severe’, while there were only two strong rainfalls with a comprehensive grade of level V, which indicates that Zhengzhou was seriously threatened by heavy rainfall. Therefore, strategies to improve the resilience of the city should be proposed to reduce the economic losses and casualties caused by heavy rainfall.

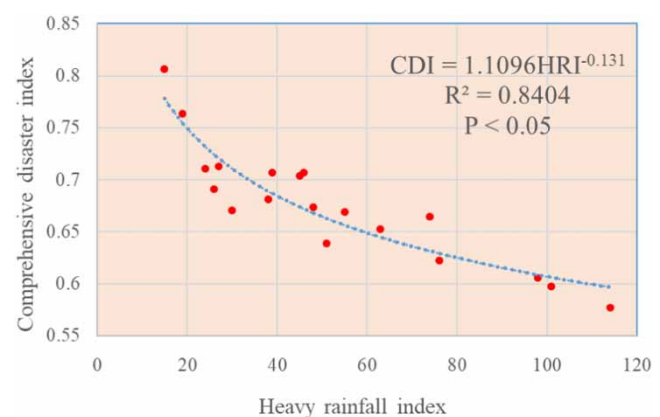
### Comprehensive disaster assessment

As can be seen from the third and fourth columns of Table 4, in which the CDI is 0.55–0.81, the flood disasters triggered by heavy rainfalls were mostly ‘Small disaster’

and ‘Moderate disaster’, accounting for 52.63% and 31.58%, respectively. One heavy rainfall whose comprehensive disaster index was 0.8062 occurred on 4 July 2012, in which the proportion of affected population, affected area and direct economic loss were also the highest among the 19 rainfall events studied, influencing people’s lives, properties and economic development seriously, followed by a heavy rainfall that occurred on 17 August 2018. From 2017 to 2018, the relatively high comprehensive disaster assessment index was thought to be attributed to rainfall having occurred frequently and heavily, which may be the case in serious disasters.

### The development of the vulnerability curve model of flood disasters and its performance

The developed vulnerability curve model of flood disasters is shown in Figure 2, which demonstrates that the determination coefficient ( $R^2$ ) of the curve model is greater than 0.8, together with a correlation coefficient that passes the significance level test of 0.05. Therefore, it was concluded that the relationship between rainfall and disaster is significant. As defined in Table 2, the greater the HRI, the lower the severity caused by rainfall, while the CDI is directly proportional to the degree of disasters according to Table 3. From this figure, the CDI shows an exponential relationship with HRI. With the increase of HDI, the severity of heavy rainfall decreases, and the CDI also decreases. The results show no difference from empirical knowledge and other studies (Lin *et al.* 2016).

**Figure 2** | Vulnerability curve of rainstorm flood disaster.



**Table 5** | The evaluation results of vulnerability curve model

	The number of samples for which simulated level is consistent with actual level	CCP (%)	The quantity of samples in which the difference between the simulated and the actual is within one level	BCP (%)
Rainfall	15	78.95	19	100
Comprehensive disaster	13	68.4	18	94.7

The performance of this model was evaluated and the results are listed in Table 5. Some conclusions can be drawn from this table. For the heavy rainfall process, of 19 heavy rainfalls that had occurred in Zhengzhou, 15 of the model simulation grades are consistent with the actual CCP reaching 78.95% and the BCP reaching 100%. In addition, for the comprehensive disaster assessment, 13 rainfalls have a similar class, with an accuracy of 68.4%, and the BCP is more than 90%. The accuracy of comprehensive disaster assessment is lower than that of the heavy rainfall assessment, given that the damage induced by heavy rainfall is not only related to accumulated precipitation, but also other factors, such as population, buildings, road conditions, etc. These factors are highly uncertain and vary from district to district, making it difficult for them to be accurately quantified in the process of estimation.

## DISCUSSION

### Strengths and limitations of data sources and methodology

The vulnerability curve was constructed to assess the vulnerability to flood disasters based on text data extracted from various platforms using web crawler technology, which effectively solves the poor assessment effect and low accuracy caused by lack of data in the vulnerability assessment of flood disasters. The evaluation results have a high fit with the historical flood disaster events. Moreover, a method of coupling geographical information system and analytical hierarchy process was proposed by Lin *et al.* (2016) to assess flood risk in Zhengzhou, showing that the risk of flood disaster in Zhengzhou had increased in recent years, which was basically consistent with the results

of the conclusions drawn in this paper. Successful application of this method in Zhengzhou provides a reference for vulnerability assessment of flood disasters in other areas, to some degree.

Since the vulnerability analysis of flood disasters using big data is only a recent topic of research, it is prone to loopholes and lacks intuitive information. For example, given this study's requirements for permission for information collection across multiple network platforms, only relevant information from authorized institutional and individual users was collected and analyzed. Therefore, the number of research samples should be enlarged in future studies. In addition, rainstorm flood disaster losses are manifested in life, production, service, and other aspects, but the classification of the disaster-bearing body was not complete and detailed enough in the process of assessment, which can reduce the precision of rainstorm and flood disaster vulnerability assessment.

### Implications for flood risk management and policy

Urban areas included in the same disaster classes share similar characteristics, so it can help decision-makers to develop more successful flood risk management strategies. Different vulnerability reduction strategies can be proposed for each class. For example, in 'Small disaster' cases, vulnerability reduction should be targeted towards individual protection, especially the elderly and children, while for the 'Moderate disaster', strategies should be focused on creating a municipal system of incentives to encourage inhabitants to carry out mitigation measures at the household level. When it comes to 'Heavy disaster', measures should be targeted towards building municipal economic support funds to help affected inhabitants (after a flash-flood event) during the recovery phase. Moreover, as for flood management in Zhengzhou, it is necessary to strengthen the rainstorm

forecast in the future to prevent the occurrence of disasters in advance. 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, etc. Furthermore, it is important to facilitate economic development and enhance individual and collective adaptability to reduce the losses caused by flooding.

## CONCLUSIONS

In this paper, the vulnerability curve model of flood disasters was developed based on text data extracted from different sources using web crawler technology to estimate flood disaster vulnerability. In the vulnerability curve model established, the HRI was considered as an independent variable and the CDI was considered as a dependent variable. Based on the vulnerability curve, it is proven that the relationship between rainfall and disaster is significant and the CDI decreases with the increase in the HRI. Based on the actual situation, the CCP and BCP were calculated to be above 65% and 90%, respectively, which demonstrates the discriminative power of the vulnerability curve model established. It was concluded that flood disaster was affected by many factors, and therefore implementing control measures to reduce disaster vulnerability is crucial. Finally, big data should be considered as a useful tool for vulnerability analysis. If more formats of big data, such as pictures and videos, are extracted and applied to assess flood disaster, it will be more conducive to the accuracy and comprehensiveness of flood disaster vulnerability evaluation.

## ACKNOWLEDGEMENTS

The study is funded by the Key Project of 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.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this paper is available online at <https://dx.doi.org/10.2166/ws.2019.171>.

## REFERENCES

- Ahmad, A., Khan, M., Paul, A., Din, S., Rathore, M. M., Jeon, G. & Chio, G. S. 2017 *Toward modeling and optimization of features selection in Big Data based social Internet of Things. Future Generation Computer Systems* **82**, 715–726. <http://doi.org/10.1016/j.future.2017.09.028>.
- Boudou, M., Danière, B. & Lang, M. 2016 *Assessing changes in urban flood vulnerability through mapping land use from historical information. Hydrology and Earth System Sciences* **20** (1), 161–175. <http://doi.org/10.5194/hess-20-161-2016>.
- Chen, Y., Liu, R., Barrett, D., Gao, L., Zhou, M., Renzullo, L. & Emelyanova, I. 2015 *A spatial assessment framework for evaluating flood risk under extreme climates. Science of The Total Environment* **538**, 512–523. <http://doi.org/10.1016/j.scitotenv.2015.08.094>.
- Deng, X. J. 2019 *Correlations between water quality and the structure and connectivity of the river network in the Southern Jiangsu Plain, Eastern China. Science of The Total Environment* **664**, 583–594. <http://doi.org/10.1016/j.scitotenv.2019.02.048>.
- de Moel, H. & Aerts, J. C. J. H. 2011 *Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates. Natural Hazards* **58** (1), 407–425. <http://doi.org/10.1007/s11069-010-9675-6>.
- Eilander, D., Trambauer, P., Wagemaker, J. & van Loenen, A. 2016 *Harvesting social media for generation of near real-time flood maps. Procedia Engineering* **154**, 176–183. <http://doi.org/10.1080/10095020.2014.988199>.
- Gori, A., Blessing, R., Juan, A., Brody, S. & Bedient, P. 2018 *Characterizing urbanization impacts on floodplain through integrated land use, hydrologic, and hydraulic modeling. Journal of Hydrology* **568**, 82–95. <http://doi.org/10.1016/j.jhydrol.2018.10.053>.
- Julien, P. Y., Ghani, A. A., Zakaria, N. A., Abdullah, R. & Chang, C. K. 2010 *Case study: flood mitigation of the Muda River, Malaysia. Journal of Hydraulic Engineering* **136** (4), 251–261. [http://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000163](http://doi.org/10.1061/(ASCE)HY.1943-7900.0000163).
- Khalaj, M., Kholghi, M., Saghafian, B. & Bazrafshan, J. 2019 *Impact of climate variation and human activities on groundwater quality in northwest of Iran. Journal of Water Supply: Research and Technology - AQUA* **68** (2), 121–135. <https://doi.org/10.2166/aqua.2019.064>.
- Lin, L., Hu, C. & Wu, Z. 2016 *Assessment of flood hazard based on underlying surface change by using GIS and analytic hierarchy process. In: Geo-Spatial Knowledge and*

- Intelligence: 4th International Conference on Geo- Informatics in Resource Management & Sustainable Ecosystem* (H. Yuan, J. Geng & F. Bian, eds), Springer, Singapore, pp. 589–599. [http://doi.org/10.1007/978-981-10-3966-9\\_65](http://doi.org/10.1007/978-981-10-3966-9_65).
- Lin, T., Liu, X. F., Song, J. C., Zhang, G. Q., Jia, Y. Q., Tu, Z. Z., Zheng, Z. H. & Liu, C. L. 2018 Urban waterlogging risk assessment based on internet open data: a case study in China. *Habitat International* **71**, 88–96. <https://doi.org/10.1016/j.habitatint.2017.11.013>.
- Ouma, Y. O. & Tateishi, R. 2014 Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment. *Water* **6** (6), 1515–1545. <http://dx.doi.org/10.3390/w6061515>.
- Ryu, J. E., Lee, D. K., Park, C., Ahn, Y., Lee, S., Choi, K. & Jung, T. 2016 Assessment of the vulnerability of industrial parks to flood in South Korea. *Natural Hazards* **82** (2), 811–825. <http://doi.org/10.1007/s11069-016-2222-3>.
- Weng, Y., Wang, X., Hua, J., Wang, H., Kang, M. & Wang, F. 2019 Forecasting horticultural products price using ARIMA model and neural network based on a large-scale data set collected by web crawler. *IEEE Transactions on Computational Social Systems* **6** (3), 547–553. <http://doi.org/10.1109/TCSS.2019.2914499>.
- Xiao, Y., Li, B. Q. & Gong, Z. W. 2018 Real-time identification of urban rainstorm waterlogging disasters based on Weibo big data. *Natural Hazards* **94** (2), 833–842. <http://doi.org/10.1007/s11069-018-3427-4>.
- Yin, J., Ye, M. W., Yin, Z. N. & Xu, S. Y. 2015 A review of advances in urban flood risk analysis over China. *Stochastic Environmental Research and Risk Assessment* **29**, 1063–1070. <http://doi.org/10.1007/s00477-014-0939-7>.
- Yoon, D. K., Kang, J. E. & Brody, S. D. 2015 A measurement of community disaster resilience in Korea. *Journal of Environmental Planning and Management* **59** (3), 436–460. <http://doi.org/10.1080/09640568.2015.1016142>.
- Yue, C. F., Wang, Q. J. & Li, Y. Z. 2018 Evaluating water resources allocation in arid areas of northwest China using a projection pursuit dynamic cluster model. *Water Science and Technology: Water Supply* **19** (3), 762–770. <http://doi.org/10.2166/ws.2018.120>.
- Zhang, Q., Yang, L. T., Chen, Z. & Li, P. 2018 A survey on deep learning for big data. *Information Fusion* **42**, 146–157. <http://doi.org/10.1016/j.inffus.2017.10.006>.

First received 27 February 2019; accepted in revised form 1 November 2019. Available online 18 November 2019