Water Science & Technology

 Water Science & Technology Vol 84 No 9, 2180 doi: 10.2166/wst.2021.417

Classification of sponge city construction modes based on regional features

Xin Zhao, Zhiming Zhang*, Wenhan Hu and Xiaotian Qi
Beijing Climate Change Response Research and Education Center, Beijing University of Civil Engineering and Architecture, Beijing 100044, China
*Corresponding author. E-mail: zhangzhiming@bucea.edu.cn

ABSTRACT

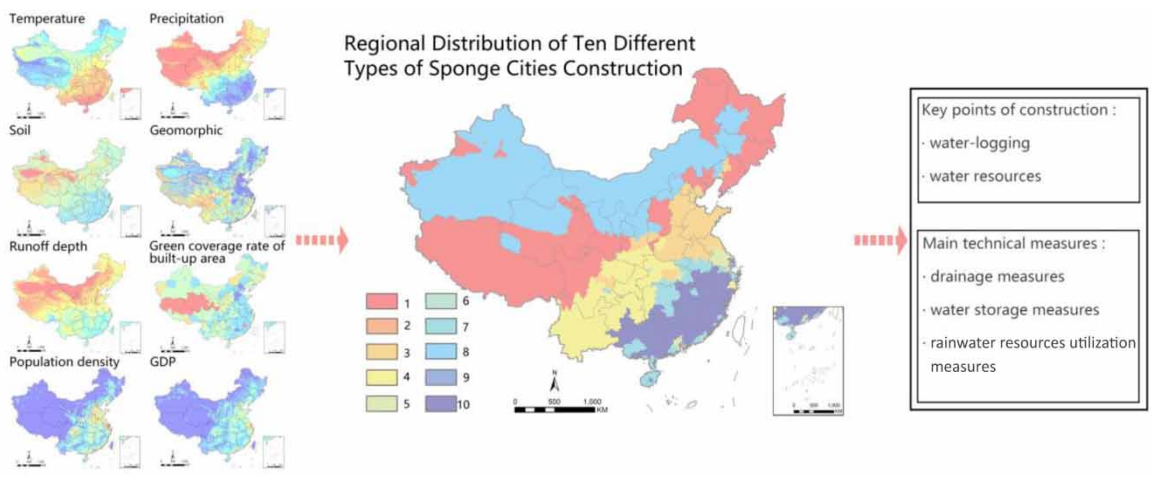
In order to make the sponge city construction model more consistent with the inherent characteristics of the city, this paper constructs classification of the sponge city construction index system. We carry out multivariate statistical and cluster analysis based on geographic information system (GIS) and classify the different regions of mainland China, putting forward key points of construction for different types of regions. The results show that climatic factor, hydrological factor and soil factor are the main factors affecting sponge city construction, followed by city scale and level of urban economic development; the third is underlying surface type and geomorphological type. The cities are classified into ten clusters, and they present a continuously zonal or flaky distribution from northeast to southwest on the whole; more than 80% of cities obviously present a continuously zonal or flaky distribution. Each of the nine clusters has at least one pilot city, which can be used as a reference for construction.

Key words: city classification, K-means clustering analysis, principal components analysis, regional features, sponge city

HIGHLIGHTS

- This paper puts forward the method of cluster analysis to classify Chinese cities.
- By using principal component analysis, it is found that the climatic, soil and hydrological factors are the most related to the results of city classification.
- Based on the construction experience of existing pilot cities, the construction emphasis of sponge cities in various types of cities is proposed.

GRAPHICAL ABSTRACT



INTRODUCTION

With the social and economic development in China and the rapid propulsion of urbanization, there are the problems of more buildings, less green space, declining regulation and storage capacity of rivers and lakes, high pavement hardening rate and poor water permeability in cities, which are causing urban water-logging, water environmental pollution and rainwater resources loss to become more and more serious (Chen & Honggang 2020). The United States put forward the rainwater management concept (referred to as low impact development (LID)) (Xing *et al.* 2011), which aimed to alleviate the negative impact of urban water-logging caused by the rainwater runoff through the rainwater source control (RSC) while storing the rainwater resources. The rainwater management models similar to LID have been widely applied in the world, such as sustainable urban drainage system (SUDS) in the UK (Thorne *et al.* 2018), water sensitive urban design (WSUD) in Australia (Morison & Brown 2010), decentralized urban design (DUD) in Germany, well balanced hydrological system (WBHS) in Japan, and healthy water circulation city (HWC2) in the Republic of Korea (Sun *et al.* 2020), which provided significant references for the concept of sponge city proposed by China in recent years. The sponge city proposed by China is a new countermeasure and development model to deal with the urban rainwater problem from a more macro perspective (Wang *et al.* 2020). Similar to the LID, the sponge city puts more emphasis on RSC compared with the traditional rainwater management model.

Due to the vast territory of China, there are great differences in natural and social conditions among different cities. In 2014, China issued the *Guideline for Sponge City Construction*, which proposed that natural factors should be taken as a part of urban runoff control strategy, such as soil and vegetation (Jia *et al.* 2017). To explore the model of sponge city construction, 30 cities were selected as pilot cities for construction (Gong *et al.* 2018). The first pilot cities of sponge city construction are focus and contiguous, which are widely distributed in Zone II, III and IV on the Zone map of Volume Capture Ratio of Annual Rainfall in China, located in humid and semi-humid areas (Han & Wu 2019), and consecutive years' average rainfall above 400 mm. This provides an important reference basis for other cities with similar conditions in the construction of sponge cities.

The traditional gray infrastructure (such as rainwater pipeline and reservoir) and green infrastructure (such as rainwater gardens and grassed swales) are important channels for the rainwater runoff control of sponge cities. Therefore, it is necessary to combine with the regional features to reach the best effect during construction. In terms of gray rainwater facilities, Yu analyzed the characteristics of urban drainage and water-logging control of plain water network, and obtained the design recurrent period of rainwater pipeline matching with the design recurrent period of water-logging control (Yu 2014). Pan *et al.* used the daily rainfall data for many years to carry out the programming simulation calculation, and took Shenzhen as an example to obtain the relationship between the volume of rainwater reservoir, rainwater utilization rate and tap water replacement rate in the engineering example (Pan *et al.* 2012). Zeng *et al.* aim at upgrading the rainwater pipe network in old urban areas, and put forward a design method for reconstruction based on the pipeline drainage load index (Zeng *et al.* 2018). In terms of green infrastructure, Jiang *et al.* studied the hydrology and pollutant removal performance of the LID (low impact development)/GI (green infrastructure) system in the arid and semi-arid areas in the United States (Jiang *et al.* 2015). Guerra *et al.* studied the soil permeability and compared the effects of high permeation soil and low permeation soil on the permeation rate and effective water-storage depth (Guerra & Kim 2020). Wei *et al.* took Alar, Xinjiang as an example; combined with the actual situation of the region, the idea of GI network construction will be put forward in oasis cities in arid regions (Wei *et al.* 2020). The research on the infrastructure of sponge cities with regional features can solve the urban rainwater problems more pertinently, and improve the effect of sponge city construction.

Undoubtedly, a better way to grasp the characteristics of the city itself from a macro perspective is more conducive to the construction of sponge cities. In the research of city classification for rainwater use, Miao classified them into three clusters and focused on the analysis of rainwater infrastructure systems in arid and semi-arid areas with the rainwater garden as core (Miao 2013). Ren *et al.* classified the Chinese cities into the northeast cold plain area, northwest arid area, northern and central warm and humid area and southern rainy area based on natural factors, and classified them into the high-density commercial and residential area, ecological leisure space, old urban area and new urban area based on the social attributes of China, in order to provide a reference for the promotion of sponge city construction (Ren *et al.* 2020). Sponge city construction by region solves the problem of large differences between cities and cannot be constructed according to the same standard.

In sponge city construction, it is important to accurately grasp the characteristics of the city. However, the previous studies did not clarify the types of cities and the pilot cities that can be referred to, and the classification relied on a single index system. Studies had insufficient grasp of the characteristics of the city, which means that the construction of cities cannot achieve the optimal. In order to provide more scientific and accurate suggestions for the construction of sponge cities for different types of cities, this paper summarizes the construction experience of existing pilot cities and the study experience of city classification, combined with the natural and social features of areas to put forward the urban classification indicator system for the sponge city construction. Based on the clustering analysis method, this paper classifies the Chinese mainland, and puts forward the macro suggestion for different sponge city construction plans. Compared with the previous research results, the classification of this paper determines the cluster of each city, which is conductive to more accurately grasping the characteristics of cities, and the pilot cities with the most similar conditions that can be referred to in the construction of each type of cities are determined. Furthermore, the classification results in this paper can provide a mutual reference for similar cities during the construction and improve the construction efficiency and effect.

DATA AND STUDY METHOD

Data sources

This study mainly focuses on 366 administrative units of Chinese mainland municipality, prefecture-level administrative region and directly administered county by province (Figure 1). The green coverage rate data of urban built-up area comes from the statistical yearbook of cities in 2019 or 2018, and the temperature, precipitation, soil, landform, gross domestic product (GDP) and population density data comes from the website of Data Center for Resources and Environmental Sciences, CAS (www.resdc.cn), which is the interpolation data of annual average temperature in 2015, the interpolation data of average annual precipitation in 2015, and the spatial distribution data of soil texture in China (Xu 2018), spatial distribution data of 1:1 million landform type in China, the spatial distribution of GDP in China based on kilometer grid data set (Xu 2017a) and spatial distribution of population in China based on kilometer grid data set (Xu 2017b), respectively. The grid precision of data is 1 km. The annual runoff depth data comes from the website of OSGeo China (https://www.osgeo.cn).

Because the formats of data obtained in various ways are different, it is necessary to carry out the calibration, projection transformation and vector processing of data and images obtained by geographic information system (GIS), so that the data can be transformed to the same projection coordinate (Wang *et al.* 2019). The processing software includes ArcGIS10.6 and

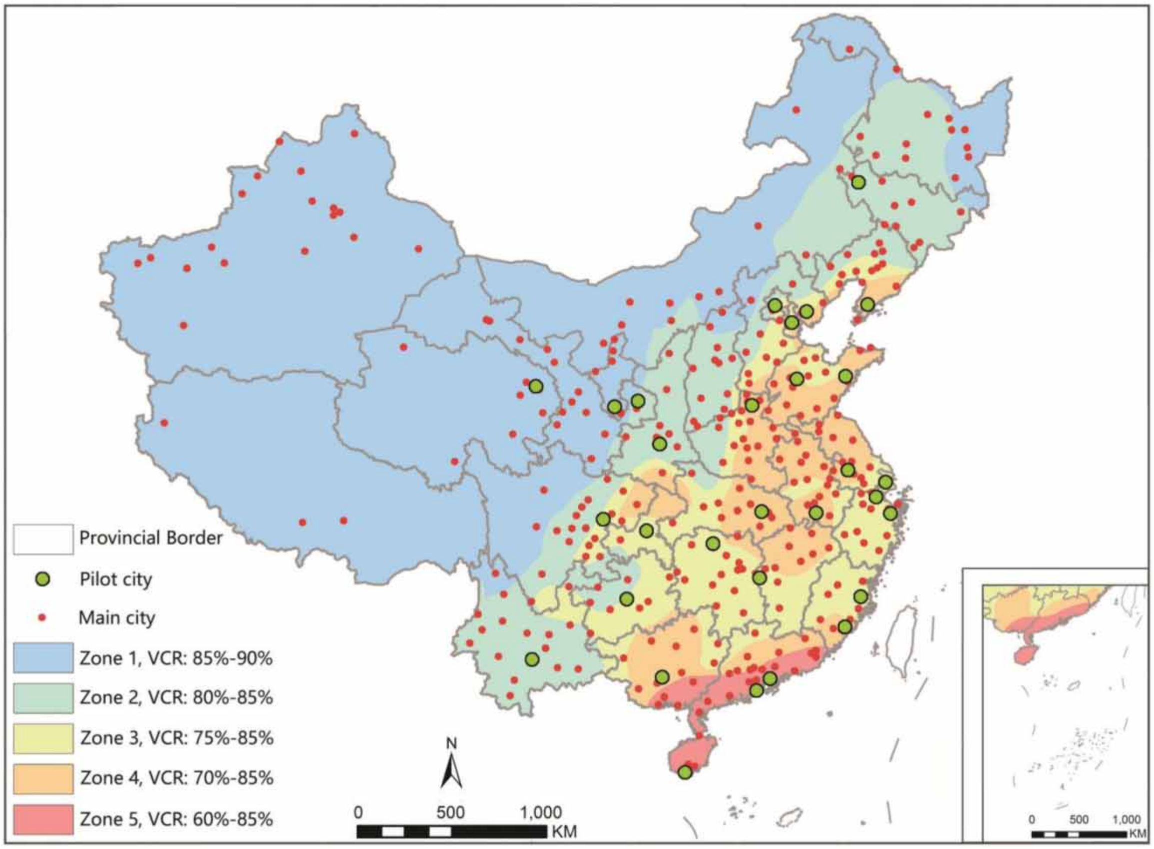


Figure 1 | The distribution of pilot cities and main cities on the Zone map of Volume Capture Ratio of Annual Rainfall (Ministry of Housing and Urban-Rural Development of the People's Republic of China 2015).

Microsoft Excel 2010. In this study, the projection grid tool in the GIS data management tool is used to unify the projection coordinate into Beijing-1954 projection coordinate system. The annual runoff depth image data is used as a reference to convert the data and image into the vector data by using the GIS mapping, and then convert them into the grid data with grid precision of 1 km by using the surface to grid tool in the conversion tool. The green coverage rate of urban built-up area is collated as the Excel data to link to the vector data of prefecture-level cities in GIS, and then convert into the grid data with grid precision of 1 km by using the surface to grid tool in the conversion tool.

STUDY METHOD

Construction of classification indicator system

The selected indicators shall not only reflect the features of cities but also be based on the needs of sponge city construction, which include a natural and a social indicator. The natural indicator includes climatic, soil, geographical and hydrological factors, while the social indicator includes economic and environmental factors. The climatic factor includes temperature and precipitation, the soil factor is soil, the geographical factor is landform, and the hydrological factor is annual runoff depth. The economic factor includes GDP and population density, and the environmental factor is the green coverage rate of urban built-up area (Table 1).

Due to the big differences in the magnitude of the evaluation indicators, the original data of the evaluation indicators must be normalized. This paper uses the linear scale transformation method to scale the data to the range (0,1]. The processed

Table 1 | Classification of impact factors on sponge city construction

Indicator Type	Indicator Factors	Specific Factors	Quantitative Method	Impact on Sponge City Construction
1 Natural indicator	1.1 Climatic factor	1.1.1 Temperature	Calculated by annual average temperature (°C)	It impacts the hydrological cycle, including evaporation, runoff and soil moisture (Li *et al.* 2019).
		1.1.2 Precipitation	Calculated by average annual precipitation (mm)	It impacts the rainfall and runoff and relates to urban water-logging (Li *et al.* 2019).
	1.2 Soil factor	1.2.1 Soil	Calculated by soil sand content (%)	The permeation and water absorption capacity of the soil are different, which mainly impacts the 'permeation' and 'detention' (Ren *et al.* 2020).
	1.3 Geographical factor	1.3.1 Landform	Calculated by altitude (m)	It reflects the topographic and water system features, such as landscape pattern, which mainly impacts on the 'permeation', 'detention' and 'storage' (Ren *et al.* 2020).
	1.4 Hydrological factor	1.4.1 Annual runoff depth	Calculated by runoff depth (mm)	The runoff depth can reflect the state of the city before development, and it can represent the construction goals of different cities, which mainly impacts on the 'permeation'.
2 Social indicator	2.1 Economic factor	2.1.1 GDP	Calculated by GDP of administrative centers (10,000 CNY/km^2)	It relates to the perfection of urban infrastructure, which mainly impacts the 'drainage' (Ren *et al.* 2020).
		2.1.2 Population density	Calculated by the population of administrative centers (person/km^2)	It represents the urban scale and relates to the urban hardening rate, which mainly impacts on the 'permeation'.
	2.2 Environmental factor	2.2.1 The green coverage rate of urban built-up area	Calculated by green coverage rate of administrative centers (%)	The green space can reduce the rainfall-runoff, delay the flood peak and purify the rainwater, which mainly impacts on the 'detention' and 'purification'.

Note: Different indicators are calculated according to equal weight, and the importance of indicators is considered to be the same.

indicator types are normalized, which can reflect the difference of the original evaluation indicators. The formula is defined as follows (Xu & Li 2020):

$$x' = \frac{x}{max(x)} \quad (1)$$

where x is the raw value, x' is the normalized value, and $max(x)$ is the maximum value of a set of data.

Technical measures for sponge city construction

The technical measures for sponge city construction include six elements (Table 2), i.e., permeation, detention, storage, purification, utilization, and drainage (Cai *et al.* 2019). Among them, the 'permeation' is to reduce the runoff from the source by reducing the hard pavement area of road surface, ground and roof and changing to the permeable pavement and green space. The 'detention' is to relieve the collection of rainwater runoff by the grassed swales and rainwater garden. The 'storage' is to collect the rainwater through the underground reservoir, to achieve the function of water regulation, storage and peak shifting. The 'purification' is to purify the rainwater by the permeation of soil, vegetation and water body. The 'utilization' means to strengthen the utilization of rainwater resources. The 'drainage' means that the sponge city construction shall be combined with the construction of urban drainage and water-logging control facilities.

City classification method

The city classification can be completed by the classification method or clustering analysis method. The common classification methods include Decision Tree, Naive Bayes and K-nearest Neighbor. The Decision Tree infers the classification rules represented by the decision tree from a group of unordered and irregular examples, which is constructed from top to bottom (Sebastiani 2002). The Naive Bayes uses the hybrid model to predict the probability of object cluster (Aggarwal & Zhai 2012), which is low accuracy. The K-nearest Neighbor measures and queries the relationship with adjacent samples by the similarity (Ertuğrul & Tağluk 2017). The clustering analysis is mainly used to study the classification of various objects or phenomena (Ke *et al.* 2011). It is the process of classifying objects into several clusters, so that the objects in the same cluster are high similarity, and objects in different clusters are low similarity. The K-means clustering and the density-based clustering are two traditional clustering methods. The classification is of the supervised machine learning classification, and the clustering is of the unsupervised machine learning classification. The classification of cities in this paper is of the latter, so the clustering analysis method is more suitable. The K-means clustering method is a practical and commonly used clustering algorithm to solve the clustering problem. This algorithm well works on the large data set clustering (Omar & Hegazy 2013), so it is the most suitable classification method in this paper.

Table 2 | Impact factors of technical measures for sponge city construction

Technical Measures for Sponge City Construction	Impact Factors	Specific Factors
Permeation	Rainfall characteristics, hardening rate, soil type, groundwater level, green coverage rate, vegetation type, temperature and water quality	Temperature, precipitation, soil, landform, annual runoff depth, population density and green coverage rate of urban built-up area
Detention	Average slope, green coverage rate, vegetation type and catchment type	Soil, landform and green coverage rate of urban built-up area
Storage	Rainfall interval, groundwater level, soil type, land use space, catchment type, sunken area and depth of green space	Soil, landform and annual runoff depth
Purification	Runoff pollutant concentration, green coverage rate, vegetation type and rainwater runoff velocity	The green coverage rate of urban built-up area
Utilization	Regional water storage, water-using structure and regional industry	GDP and population density
Drainage	Recurrent period of pipe network, slope, drainage density, water body and road	GDP

The city classification is calculated by the K-means clustering in SPSS. This algorithm classifies the data set of given cities into certain groups identified. The data set is classified into ten groups, defines ten center points, and places these ten data points in the space to represent the target of clustering, and these data points represent the center of the initial cluster. Each object is allocated to the nearest center cluster. When all objects are allocated, the positions of ten center points are recalculated. Continuously repeat the two steps of allocating objects and computing centers until the cluster center no longer changes. The minimum function of standard function is defined as follows:

$$J = \sum_{i=1}^{10} \sum_{Z \in Zi} \|z - mi\| \tag{2}$$

where 10 is the number of cluster groups, mi is the mean vector of the ith cluster, z is the score vector, and Zi is the ith cluster among 10 clusters (Pang *et al.* 2009). The technical road map of K-means clustering analysis in this study is shown in Figure 2.

Principal component analysis

Although the eight indicators selected have their own clear significance, there is a certain correlation between them. Therefore, there is some redundancy in providing information on urban features. The principal component analysis (PCA) can perform the function of dimension reduction and express the difference between cities by as few indicators as possible (Li 2019), so as to identify the main impact factors for the difference of sponge city construction.

PCA is a multivariate statistical analysis method that reduces multiple indicators into a few comprehensive indicators (Li & Hu 2017). The correlations between eight factors are examined by using PCA, and the information that can reflect the characteristics extracted from the eight factors; then the final extraction of PCA can be captured to the maximum extent of the original data. Compared with the original indicator factor, the number of principal components is lower, and the principal components are more uncorrelated, to avoid the arbitrariness of subjective judgment.

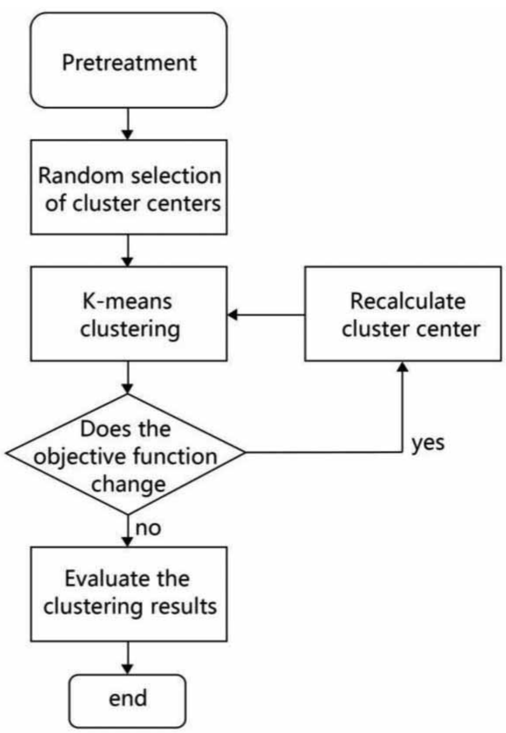


Figure 2 | Flow chart of K-means clustering algorithm.

RESULTS AND DISCUSSION

Spatial distribution of factors

The details of eight indicators for city classification are shown in Figures 3–6 respectively.

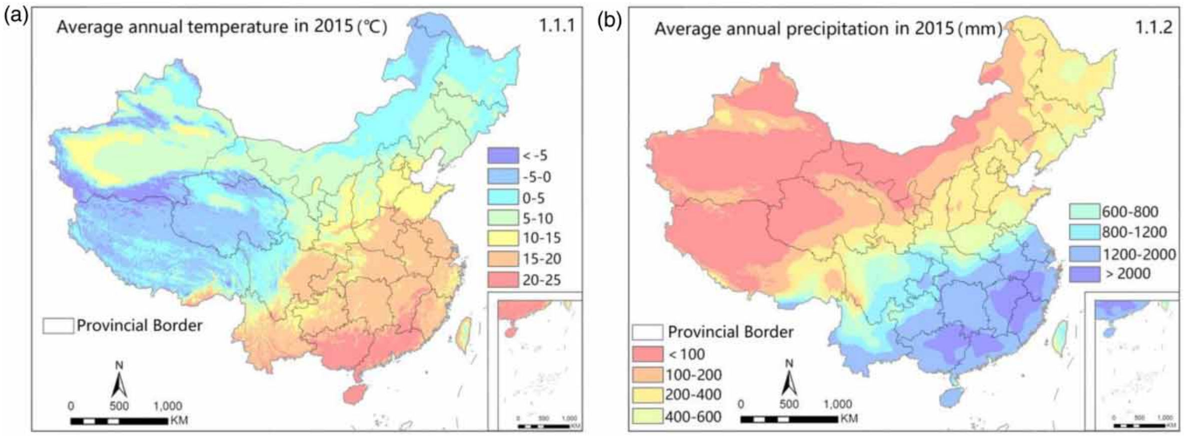


Figure 3 | Climatic Factor Map: Average annual temperature and average annual precipitation in 2015.

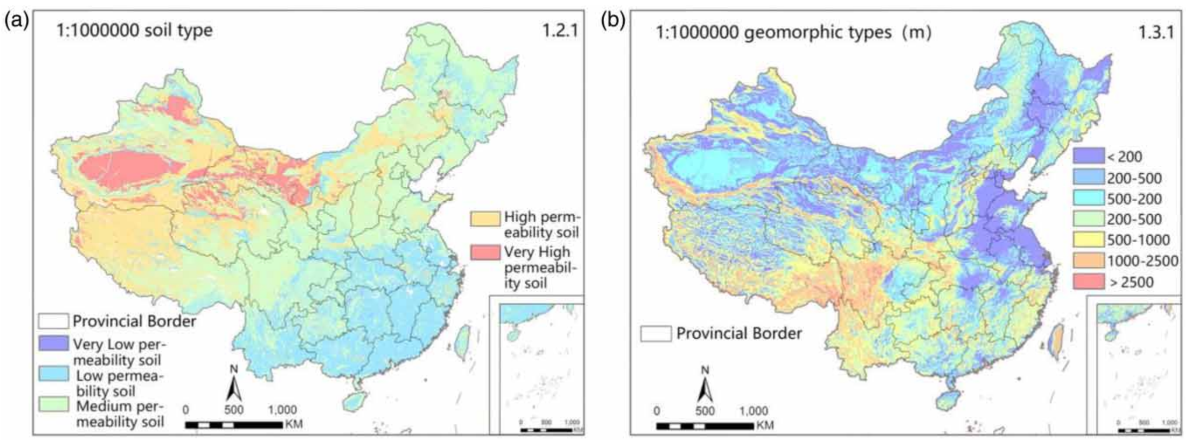


Figure 4 | Geographical Factor Map: 1:1,000,000 soil type and 1:1,000,000 geomorphic types.

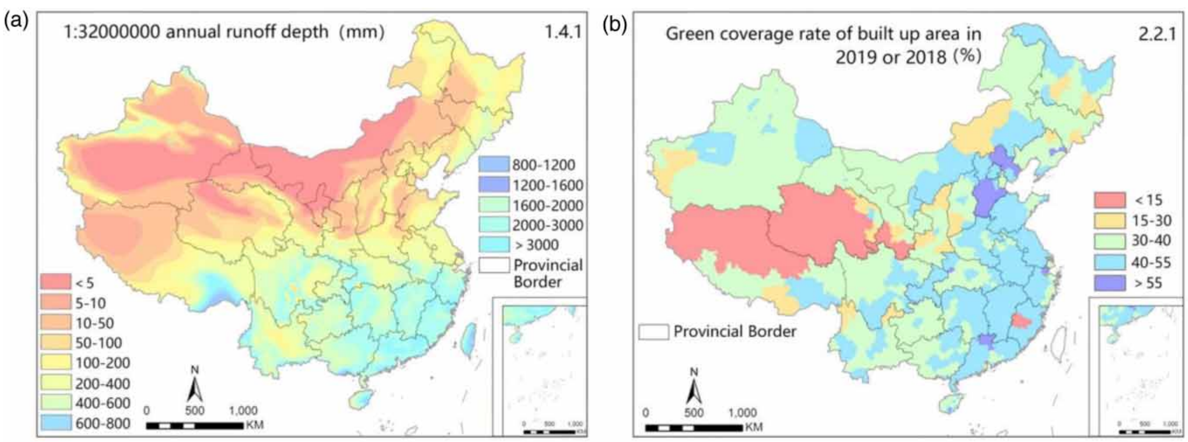


Figure 5 | Hydrological and Environmental Factors Map: 1:32,000,000 annual runoff depth and green coverage rate of built-up area in 2019 or 2018. *Note:* In (b), the color block covering the whole city only represents the green coverage rate of the urban built-up area.

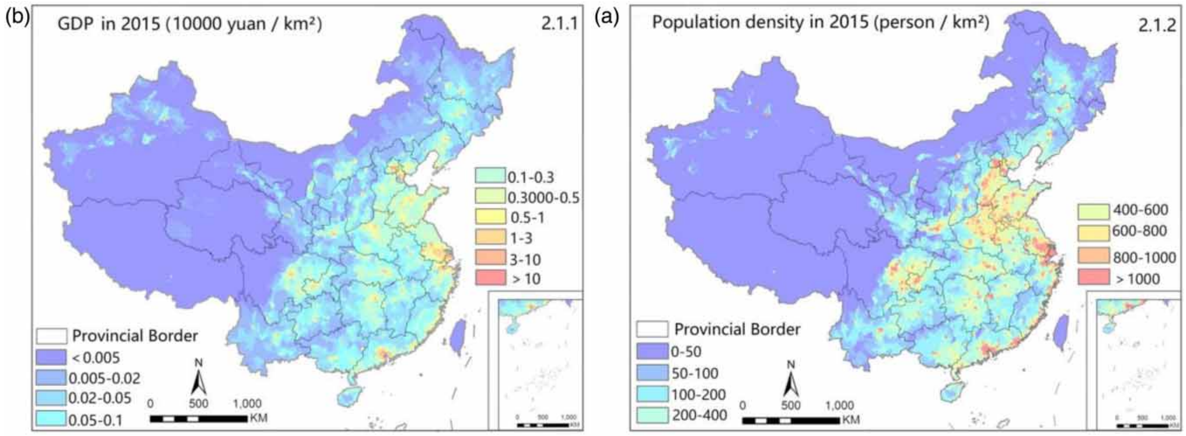


Figure 6 | Economic Factor Map: GDP and population density in 2015.

The temperature and precipitation present the trend that the temperature gradually decreases and the precipitation gradually reduces from southeast to northwest (Figure 3).

The soil sand content gradually increases and the soil permeability increases from southeast to northwest (Figure 4(a)). There are many high mountains in the southwest, many high mountains and hills in the south and a large plain area in the east (Figure 4(b)).

The annual runoff depth gradually decreases from southeast to northwest (Figure 5(a)). The green coverage rate of built-up areas in the eastern area is generally high, while that in the western and northern areas is relatively low (Figure 5(b)).

The areas with large population density and high GDP are concentrated in the east, and the population density and GDP are maximum in the east and southeast coast (Figure 5).

The changing trend of temperature, precipitation, soil and annual runoff depth is the same. The higher the temperature, the heavier the precipitation, the poorer the soil permeability, and the larger the annual runoff depth. The changing trend of population density, GDP and green coverage rate of built-up area is similar. The larger the population density, the higher the GDP and green coverage rate of built-up area. The population density and GDP relate to the landform, and the large population density and high GDP are located in the plain area (Figures 3–6).

Main impact factors of classification

The results of principal component analysis of eight factors selected on the basis of urban features and sponge city construction are shown in Table 3.

The contribution rate of the first three principal components is 81.91%. Among them, the contribution rate of the first principal component is 45.20%, that of the second principal component is 24.16%, and that of the third principal component is

Table 3 | Principal component factor loading and the cumulative variance contribution ratio table

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
1.1.2 Precipitation	0.93	0.09	0.06					
1.2.1 Soil	−0.84	−0.21	−0.07					
1.1.1 Temperature	0.81	0.16	0.31					
1.4.1 Annual runoff depth	−0.93	−0.10	0.15					
1.3.1 Landform	0.36	−0.44	−0.63					
2.1.2 Population density	0.20	0.92	0.13					
2.1.1 GDP	0.14	0.93	0.01					
2.2.1 Green coverage rate of built-up area	0.28	−0.02	0.83					
Cumulative variance contribution rate (%)	45.20	69.36	81.91	90.16	94.27	96.60	98.70	100.00

12.55%. The contribution rate of the first three principal components is much higher than that of other components, and they integrate most of the information of cities, so the first three components are the key factors to determine the city classification results. The first principal component is strongly correlated with the natural factors, such as climatic, soil and hydrological factors. Among them, the first principal component is positively correlated with the temperature and precipitation and negatively correlated with the soil and annual runoff depth. The second principal component is strongly positively correlated with the population density and GDP. The third principal component is strongly positively correlated with the green coverage rate of built-up area and a negative correlation with the landform. The results of PCA demonstrate that climatic, hydrological and soil factors are the most important factor, followed by city scale and level of urban economic development; the third is underlying surface type and geomorphological type (Table 3).

Type distribution of sponge city construction

The 366 cities are classified into ten clusters, whose clustering analysis results are shown in Figure 7.

The cities present the continuous zonal or flaky distribution on the whole from northeast to southwest, and in the first, third, fourth, eighth and tenth clusters, more than 80% of cities obviously present the continuous zonal or flaky distribution (Figure 7). The first cluster is mainly located in the northeast and Tibet, and a few cities are located in the northwest and central areas. The second and sixth clusters are located in Shenzhen and Shennongjia forestry districts. The third cluster is located in the north. The fourth cluster is mainly located in the southwest, and a few cities are located on the east coast. The fifth cluster is mainly located on the east coast. The seventh cluster is mainly located in the central and southern areas. The eighth cluster is mainly located in the northwest and north. The ninth cluster is located in Dongguan and Shanghai. The tenth cluster is mainly located on the southeast coast.

The fifth, seventh, ninth and tenth clusters are of the abundant water area, and the second, fourth and sixth clusters are of the watery area. It can suffer from water-logging, and the water safety problem is prominent in these areas. The third cluster is of the normal water area, but the precipitation is concentrated in summer, so it is easy to suffer from the water-logging in summer. The first and eighth clusters are of the water shortage area, whose average annual precipitation is very little, so

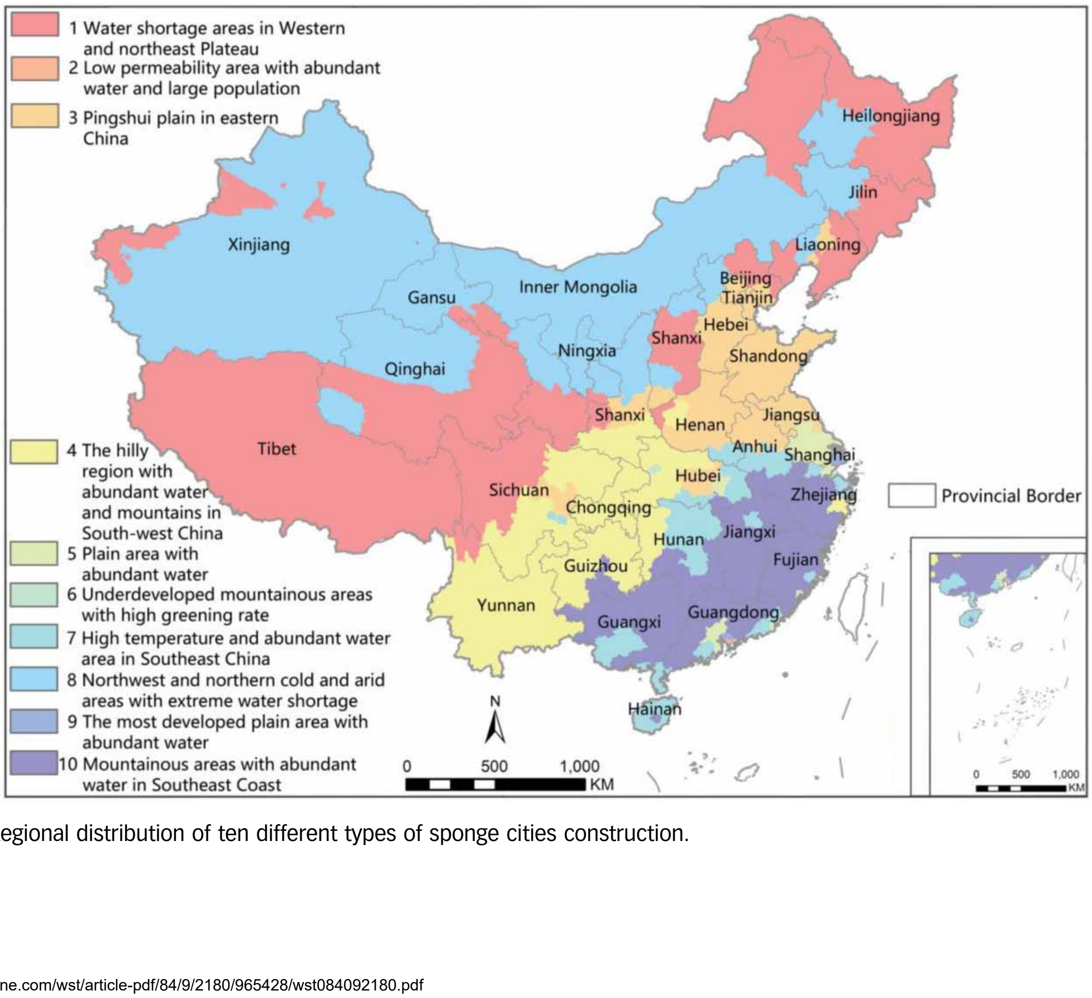


Figure 7 | Regional distribution of ten different types of sponge cities construction.

they are lacking water resources. In the abundant water area, the ninth cluster has the maximum GDP and larger population density, followed by the fifth cluster, and the seventh and tenth clusters have the lower GDP. These cities feature more precipitation, high temperature, poor soil permeability, large annual runoff depth, high degree of urbanization and high risk of urban water-logging. In the watery area, the second cluster has the maximum population density and the poorest soil permeability, the sixth cluster has the minimum population density and the minimum GDP, which is located in mountainous areas, and the fourth cluster is located in mountainous and hilly areas. The third cluster is of the normal water plain area, which has the average soil permeability, and there is an obvious water shortage. The water shortage areas (i.e., the first and eighth clusters) feature the low temperature, minimum precipitation, extremely high soil permeability, small annual runoff depth, minimum population density, low GDP, low green coverage rate of built-up area, water shortage and prominent water environment problem. The first cluster is located in plateau mountainous areas, and the eighth cluster is complex terrain (Figures 3–7).

The P (significance) is less than 0.05 by using the variance to analyze the difference of various factors for clustering clusters (Table 4), which indicates that the difference between ten clusters obtained by the clustering analysis is large (Zhou *et al.* 2007), and the clustering clusters are significant for all study items. It is concluded that the factors have obvious differences for clustering clusters.

Analysis on key points of sponge city construction

According to the city classification results and by referring to the construction of pilot cities in each cluster (Table 5), this paper puts forward different sponge cities construction strategies as follows.

According to Table 5, cities were grouped into the four broad categories, Cities in the first category have less rainfall, complex geological conditions, high soil permeability and low GDP; the main problem is water shortage, and the emphasis is on collecting the rainwater and the resource utilization of rainwater. Cities in the second category have abundant rainfall, low soil permeability, high GDP; the main problem is water-logging and water environment, and the emphasis is on using the rainwater detention and rainwater purification measures, repairing the urban drainage system. Cities in the third category have abundant rainfall or are rainy in summer; the main problem is water shortage and water-logging, and the emphasis is on collecting and using the rainwater, repairing the urban drainage system. Cities in the fourth category have more rainfall, low GDP, and are located in mountainous areas; the main problem is soil erosion, and the emphasis is on collecting and using the rainwater, and reduce using the permeation facilities. Clustering the ten cities, nine clusters have at least one pilot city; the other category is Shennongjia forestry district. These nine types of cities can use the pilot cities as a reference during construction.

It is concluded from different sponge cities' construction strategies that the main problems of Chinese cities are the water-logging and water resources. The main technical measures taken include drainage measures, water storage measures and rainwater resource utilization measures. For cities that need to solve the problem of water-logging, it is necessary to focus on the drainage, and the rainwater detention measures shall be taken for some cities. For cities with water resources problems, some cities have water shortage, while other cities have the overuse of groundwater. These cities need to focus on water storage and use. For cities with a high degree of urbanization, it is necessary to focus on water purification (Table 5).

Table 4 | Comparison results of variance analysis for clustering clusters

	Clustering		Error			
	Mean Square	Degrees of Freedom	Mean Square	Degrees of Freedom	F	Significance
2.1.1 GDP	0.33	9	0.00	356	127.80	0.000
1.3.1 Landform	0.92	9	0.01	356	67.10	0.000
1.1.2 Precipitation	2.30	9	0.00	356	322.20	0.000
2.1.2 Population density	0.35	9	0.00	356	149.82	0.000
1.2.1 Soil	0.38	9	0.00	356	71.20	0.000
1.1.1Temperature	1.84	9	0.01	356	176.72	0.000
1.4.1Annual runoff depth	1.24	9	0.01	356	156.10	0.000
2.2.1 Green coverage rate of built-up area	0.10	9	0.01	356	12.66	0.000

Table 5 | Category and cluster of cities, sample number of cities, natural and social conditions, pilot city, main problem and key points of construction in each cluster

Category	Cluster	Sample Number of Cities	Natural and Social Conditions	Pilot City	Main Problem	Key Points of Construction
1	1	65	Plateau mountainous area, dry climate and water shortage, low temperature, high soil permeability, low GDP and low green coverage rate	Dalian and Xining	Water shortage	1. Reduce using the permeation facilities. 2. Collecting the rainwater and the resource utilization of rainwater.
	8	60	It is of drought and water shortage, collapsible loess area with water shortage and complex geological conditions.	Baicheng, Qingyang and Guyuan	Water shortage	3. The sponge city construction in such areas focuses on 'storage' and 'utilization'.
2	2	1	Abundant rainfall, poor soil permeability, high GDP, large population density and prominent water environment problem	Shenzhen	Water-logging and water environment	1. Use the rainwater detention and rainwater purification measures. 2. Repair the urban drainage system.
	7	47	It is of abundant rainfall and high-temperature areas with a high degree of urbanization.	Ningbo, Changde, Nanning and Sanya	Water-logging and water environment	3. The sponge city construction in such area focuses on 'detention', 'purification' and 'drainage'.
	9	2	Abundant rainfall, poor soil permeability, scarce groundwater and developed economy.	Shanghai	Water-logging, water shortage and water environment	1. Use rainwater detention facilities, collect rainwater, purify and reuse the resources. 2. Repair the urban drainage system. 3. The sponge city construction in such area focuses on 'detention', 'storage', 'purification', 'utilization' and 'drainage'.
	10	54	Abundant rainfall, poor soil permeability, mountainous areas, high groundwater level and prominent water environment problems.	Chizhou, Fuzhou and Pingxiang	Water-logging and water environment	1. Reduce using the permeation facilities. 2. Use the rainwater detention and the rainwater purification facilities. 3. Repair the urban drainage system. 4. The sponge city construction in such area focuses on 'detention', 'purification' and 'drainage'.
3	3	68	Plain area, obvious water shortage, seasonal rainfall and rainy in summer	Beijing, Tianjin, Jinan, Qingdao, Hebi, Suining Qianan and Xixian New Area	Water shortage and seasonal water-logging	1. Collection and utilization of rainwater resources, use the rainwater permeation measures. 2. Repair the urban drainage system. 3. The sponge city construction in such areas focuses on 'permeation', 'storage', 'utilization' and 'drainage'.

	4	52	Mountainous and hilly area, which is greatly affected by the terrain and has abundant rainfall. However, the city itself has a water shortage.	Chongqing, Yuxi and Guian New Area	Water shortage and water-logging	1. Reduce using the permeation facilities. 2. Collection and utilization of rainwater resources. 3. Repair the urban drainage system. 4. The sponge city construction in such area focuses on 'storage', 'utilization' and 'drainage'.
	5	16	It is located in the plain river network-intensive area with abundant rainfall, low soil permeability, and scarce groundwater level.	Zhenjiang, Jiaxing, Xiamen, Wuhan and Zhuhai	Water shortage and water-logging	1. Collection and utilization of rainwater resources, use the rainwater detention measures. 2. Repair the urban drainage system. 3. The sponge city construction in such area focuses on 'detention', 'storage', 'utilization' and 'drainage'.
4	6	1	It is located in mountainous areas with a high green coverage rate, abundant rainfall, and an underdeveloped economy.	None	Soil erosion	1. Reduce using the permeation facilities. 2. Storage and utilization of rainwater. 3. Sponge city construction in such areas focuses on 'storage' and 'utilization'.

CONCLUSION

According to the temperature, precipitation, soil, landform, annual runoff depth, GDP, population density and green coverage rate of built-up area, 366 cities are classified into ten clusters, and then the variance analysis and principal component analysis are carried out for the results to conclude as follows:

(1) It can be seen from the clustering analysis and variance analysis that there is an obvious difference among ten clusters of cities obtained by the K-means clustering analysis, which presents a continuous zonal or flaky distribution from northeast to southwest on the whole; more than 80% of cities obviously present a continuously zonal or flaky distribution.
(2) Through principal component analysis, the eight specific factors can be reduced to three principal components, explaining 81.91% of the total variance. The results of PCA demonstrate that climatic, hydrological and soil factors are the most important factors with the percentage contributed 45.2%, followed by city scale and level of urban economic development with the percentage contributed 24.265%, and the third is underlying surface type and geomorphological type with the percentage contributed 12.55%.
(3) According to the clustering results, the major cities in China can be classified into ten clusters, and each of the nine clusters has at least one pilot city, which can be used as a reference for construction.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Aggarwal, C. C. & Zhai, C. X. 2012 *A Survey of Text Classification Algorithms in Mining Text Data*. Springer Science Business Media, 163–222. https://doi.org/10.1007/978-1-4614-3223-4_6.

Cai, X., Cheng, J. & Wong, P. 2019 Study on urban water system ecological restoration and management under the background of sponge city. *Jianghuai BBS* **06**, 75–80. https://doi.org/10.16064/j.cnki.cn34-1003/g0.2019.06.010.

Chen, H. & Honggang, C. 2020 Research on sponge city construction. *IOP Conference Series: Earth and Environmental Science* **568** (1). https://doi.org/10.1088/1755-1315/568/1/012048.

Ertuğrul, Ö. F. & Tağluk, M. E. 2017 A novel version of k nearest neighbor: dependent nearest neighbor. *Applied Soft Computing* **55**, 480–490. https://doi.org/10.1016/j.asoc.2017.02.020.

Gong, Y., Yin, D., Fang, X. & Li, J. 2018 Factors affecting runoff retention performance of extensive green roofs. *Water* **10** (9). https://doi.org/10.3390/w10091217.

Guerra, H. B. & Kim, Y. 2020 Understanding the performance and applicability of low impact development structures under varying infiltration rates. *KSCE Journal of Civil Engineering* **24** (1), 1430–1438. https://doi.org/10.1007/s12205-020-2274-5.

Han, X. & Wu, Y. 2019 Construction and application of 'the sponge city' in different precipitation regions: case studies in Pingxiang and Ji'nan, China. *Energy Procedia* **159**, 207–212. https://doi.org/10.1016/j.egypro.2018.12.052.

Jia, H., Yu, S. L. & Qin, H. 2017 Low impact development and sponge city construction for urban stormwater management. *Frontiers of Environmental Science & Engineering* **11** (4), 20. https://doi.org/10.1007/s11783-017-0989-4.

Jiang, Y., Yuan, Y. & Piza, H. 2015 A review of applicability and effectiveness of low impact development/green infrastructure practices in arid/semi-arid United States. *Environments* **2** (2), 221–249. https://doi.org/10.3390/environments2020221.

Ke, W., Cheng, H. P., Yan, D. & Lin, C. 2011 The application of cluster analysis and inverse distance-weighted interpolation to appraising the water quality of Three Forks Lake. *Procedia Environmental Sciences* **10**, 2511–2517. https://doi.org/10.1016/j.proenv.2011.09.391.

Li, K. 2019 Research on the evaluation of my country's rural ecological civilization construction level based on the principal component method. *Agriculture and Technology* **39** (17), 153–157. https://doi.org/10.19754/j.nyyjs.20190915061.

Li, X. & Hu, B. 2017 Evaluation of water resources carrying capacity in Shaanxi Province based on principal component analysis groundwater. **39** (05), 161–163. https://doi.org/10.3969/j.issn.1004-1184.2017.05.059

Li, Y., Guo, F., Mao, K. & Chen, F. 2019 Response strategy for drought and flood in sponge city construction risk under the background of climate change. *IOP Conference Series: Earth and Environmental Science* **252** (4). https://doi.org/10.1088/1755-1315/252/4/042015

Miao, Z. T. 2013 Urban rainwater infrastructure planning in arid and semi-arid regions based on LID. *Advanced Materials Research* **671–674**, 2410–2413. https://doi.org/10.4028/www.scientific.net/AMR.671-674.2410.

Ministry of Housing and Urban-Rural Development of the People's Republic of China 2015 *Sponge City Construction Technical Guide-Low Impact Development Rainwater System Construction (Trial)*. China Architecture & Building Press, Beijing.

Morison, P. J. & Brown, R. R. 2010 Understanding the nature of publics and local policy commitment to water sensitive urban design. *Landscape and Urban Planning* **99** (2), 83–92. https://doi.org/10.1016/j.landurbplan.2010.08.019.

Omar, B. A. & Hegazy, A. E.-F. 2013 Hybrid ant-based clustering algorithm with cluster analysis techniques. *Journal of Computer Science* **9** (6), 780–793. https://doi.org/10.3844/jcssp.2013.780.793.

Pan, Z., Lu, M., Wang, L., Peng, S. & Li, J. 2012 Study on volume design and calculation of rainwater reservoir in Shenzhen city. *Water Supply and Drainage* **48** (10), 43–46. https://doi.org/10.13789/j.cnki.wwe1964.2012.10.015.

Pang, Y., Liu, M. & Li, Y. 2009 Application of K-means clustering analysis in process improvement. *Journal of Huazhong University of Science and Technology (Natural Science Edition)* **37** (S1), 245–247. https://doi.org/10.13245/j.hust.2009.s1.006.

Ren, N., Huang, H. & Wang, Q. 2020 The regional classification construction paradigm of sponge city. *Environmental Engineering* **38** (04), 1–4. https://doi.org/10.13205/j.hjgc.202004001.

Sebastiani, F. 2002 Machine learning in automated text categorization. **34** (1), 1–47. https://doi.org/10.1145/505282.505283.

Sun, Y., Deng, L., Pan, S.-Y., Chiang, P.-C., Sable, S. S. & Shah, K. J. 2020 Integration of green and gray infrastructures for sponge city: water and energy nexus. *Water-Energy Nexus* **3**, 29–40. https://doi.org/10.1016/J.WEN.2020.03.003.

Thorne, C. R., Lawson, E. C., Ozawa, C., Hamlin, S. L. & Smith, L. A. 2018 Overcoming uncertainty and barriers to adoption of Blue-Green Infrastructure for urban flood risk management. *Journal of Flood Risk Management* **11** (52), S960–S972. https://doi.org/10.1111/jfr3.12218.

Wang, H., Wang, L. & Wei, B. 2019 Research on rainwater collection and storage technology based on GIS and LID – A case study of Yingqiu Town, Shandong Province. *Bulletin of Soil and Water Conservation* **39** (02), 155–160. https://doi.org/10.13961/j.cnki.stbctb.2019.02.025.

Wang, Z., Lou, Y., Zhang, B. & Shi, B. 2020 Thoughts and strategies of sponge city construction planning: a case study of Pingxiang, Jiangxi Province. *Open Journal of Civil Engineering* **10** (1), 43–54. https://doi.org/10.4236/ojce.2020.101005.

Wei, J., Zhang, Y., Liang, J. & Zhou, X. 2020 Research on green infrastructure network construction in oasis city in arid regions – a case study of Alar city. *Chinese Landscape Architecture* **36** (05), 24–29. https://doi.org/10.19775/j.cla.2020.05.0024.

Xing, W., Zhao, D., Chen, J. & Wang, H. 2011 Sustainable urban rainwater systems based on low impact development (LID). *Water and Drainage in China* **27** (20), 13–16. https://doi.org/10.19853/j.zgjsps.1000-4602.2011.20.005.

Xu, X. 2017a *China's GDP Spatial Distribution km Grid Data set*. Data Registration and Publication System of Data Center for Resources and Environmental Sciences, CAS (http://www.resdc.cn/DOI). https://doi.org/10.12078/2017121102.

Xu, X. 2017b *Spatial Distribution of Population in China Based on Kilometer Grid Data set*. Data Registration and Publication System of Data Center for Resources and Environmental Sciences, CAS (http://www.resdc.cn/DOI). https://doi.org/10.12078/2017121101.

Xu, X. 2018 *Spatial Distribution of Soil Texture in China*. Data Registration and Publication System of Data Center for Resources and Environmental Sciences, CAS (http://www.resdc.cn/DOI). https://doi.org/10.12078/2018060501.

Xu, L. & Li, M. 2020 Research on data preprocessing method in dynamic comprehensive evaluation. *Chinese Journal of Management Science* **28** (01), 162–169. https://doi.org/10.16381/j.cnki.issn1003-207x.2020.01.014.

Yu, X. 2014 Study on the design recurrent period of urban rainwater pipeline in plain water network. *Water Supply and Drainage* **50** (06), 25–28. https://doi.org/10.13789/j.cnki.wwe1964.2014.0118.

Zeng, X., Li, E., Rao, S., Zou, H., He, P., Shi, Y., Luo, F., Tao, T. & Wang, H. 2018 Reconstruction design of rainwater pipe network in old urban district based on drainage load of pipeline. *China Water & Wastewater* **34** (08), 68–72. https://doi.org/10.19853/j.zgjsps.1000-4602.2018.08.015.

Zhou, F., Guo, H., Liu, Y., Luo, D. & Wang, Z. 2007 A new approach for water quality assessment based on multivariate statistical analysis and radial basis function neural networks. *Acta Scientiae Circumstantiae* **05**, 846–853. https://doi.org/10.13671/j.hjkxxb.2007.05.023.

First received 25 July 2021; accepted in revised form 9 September 2021. Available online 22 September 2021