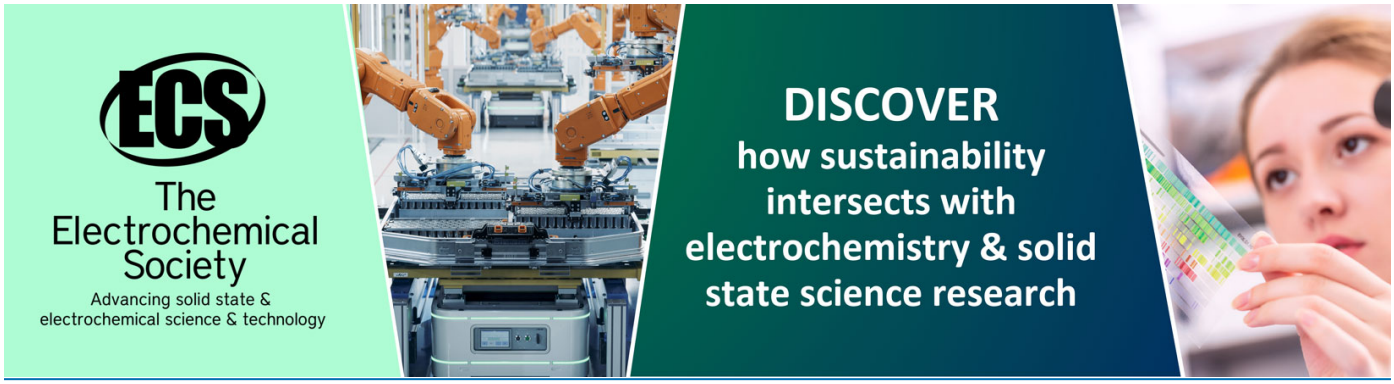
A systematic review of water resources in Yellow River Basin of China



A systematic review of water resources in Yellow River Basin of China

T Pan[1], R G Jiang[1,2], J C Xie[1], Y Y Liu[1] and Y P Wang[1]

[1] State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an 710048, China

[2] E-mail: jrengui@163.com

Abstract. This paper reviews the literature on the Yellow River Basin's water resources from 2000 to 2019 based on the Web of Science. The annual publication, source country and source journal are analyzed using SATI4.0. The nonhomogeneous linear difference equation, comprehensive score analysis and Bradford's law are used for fundamental analysis. CiteSpace is used for co-occurrence analysis and burst detection of noun phrases. Finally, Carrot2 is used to cluster the literature to get a knowledge framework for the research. Results showed that: (1) The research process of water resources in the Yellow River Basin (YRB) can be divided into three stages. This topic will continue to flourish in the future. (2) The research covers a wide range of contents, which mainly includes runoff changes, runoff-sediment relationship, water resources allocation, disaster response, the impact of human activities and climate change, as well as ecological protection and sustainable development. Human activities and the impact of climate change are the focus of research. (3) In the future, scholars should adhere to the belief of green and sustainable development of YRB and focus on the study of the Loess Plateau.

1. Introduction

The Yellow River Basin (YRB) is an important birthplace of Chinese civilization [1]. The situation of water resources in the YRB is related to the ecological security and sustainable development of water resources. In September 2019, a symposium on ecological protection and high-quality development of the YRB was held in Zhengzhou, China. The meeting points out that the amount of water in the YRB is limited. We should develop water-saving industries and technologies, promote water-saving in agriculture, and take actions to save water for the whole society. The ultimate goal is to realize the transformation of water use from extensive to economical and intensive [2].

In recent years, runoff reduction and water environment pollution have appeared in the YRB. Promoting the ecological protection and high-quality development of the YRB becomes a hot topic. Many studies focus on the water resources in the YRB. For example, Zheng et al. (2009) used the concept of climate elasticity to evaluate the impact of climate and land use rate changes on river flow. The results showed that the impact of land use rate changes on river runoff reduction accounted for more than 70% in the 1990s [3]. Based on the hydrological and meteorological data of six regions in the YRB, Abubaker Omer et al. (2020) quantitatively analyzed the impact of human activities and nature on drought [4]. Zuo et al. (2016) studied the effects of land use and climate change on water and sediment yields in the Huangfuchuan watershed [5]. Wang et al. (2020) studied the occurrence level, source and potential toxicological significance of polycyclic aromatic hydrocarbons (PAHs) in the surface sediment of the Yellow River Estuary. The results showed that the PAHs in the sediment mainly came from coal

and biomass combustion, oil pollution and vehicle emission [6]. Di et al. (2020) put forward a game model to solve the problem of water rights transaction, and the YRB was used as an example to verify the applicability and effectiveness [7].

In this paper, the periodical literature on the water resources in the YRB was analyzed using bibliometrics based on the paper published in the Web of Science. The research hotspots and frontiers were analyzed using CiteSpace and SATI for fundamental analysis, co-occurrence analysis and burst detection. The Carrot2 was used to summarize the research framework of the study.

2. Material and methodology

2.1. Data collection

The data of this paper comes from the Science Citation Index Expanded database in the Web of Science Core Collection Database. The retrieval theme is "Yellow River" or "Huanghe River" and "Water resource*". The time span is from 2000 to 2019. After retrieval, there are 574 records. In order to ensure the accuracy of the literature, the retrieval results are screened one by one. The reports, news, meeting notices and irrelevant items are removed. SATI4.0 is used to clean and remove duplication, and 562 results are retained.

2.2. Methods

SATI is a document information statistical analysis tool based on. Net platform and C# language. CiteSpace is a knowledge visualization software developed by professor Chen [8]. It can present the structure, laws and distribution of scientific knowledge. It can explore the hotspots and the frontiers of research. Carrot2 is a clustering tool that does not require presorting. It provides various components to obtain search results, including Google Application Programming Interface (Google API), Bing Application Programming Interface (Bing API), eTools Meta search, Lucene, Solr, etc. It integrates clustering algorithms such as lingo, Suffix Tree Clustering (STC), and provides visual display forms such as circle chart and bubble tree [9]. Figure 1 shows the analysis flow of this paper.

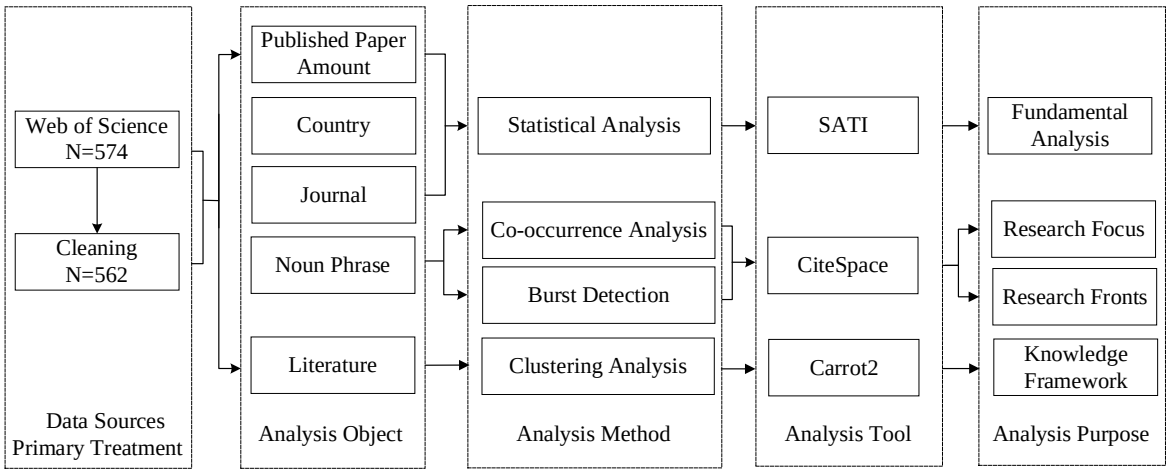


Figure 1. The analysis process of water resources in the YRB.

3. Results and discussion

3.1. Fundamental analysis

3.1.1. Analysis of annual frequency of literature. The distribution map of annual publication frequency of literature with time can reflect the degree of attention and development trend of this research field. It

can provide a reference for evaluating the value of future research. The frequency is calculated by year using SATI4.0. The annual frequency distribution of literature is shown in Figure 2.

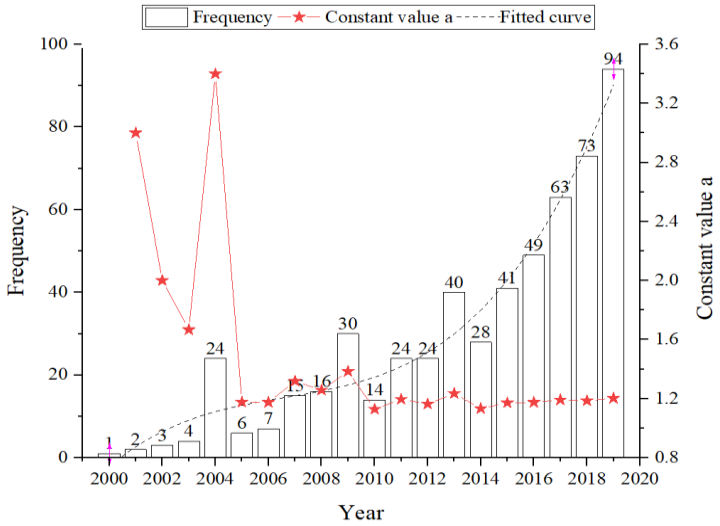


Figure 2. Annual frequency distribution of literature.

The polynomial is used to fit the trend line, and the fitting curve (dotted line in the Figure 2) is:

$$y=0.0328x^3-0.7508x^2+6.8114x-8.2378 \quad (1)$$

where x is the n-th year since 2000. $R^2=0.9449$. It indicates a good fitting effect. It can be seen that the number of literatures is on the rise, and the growth rate is accelerating. We can predict that the number of literatures will exceed 100 in 2020. In order to further study its development process, the nonhomogeneous difference equation is used to divide it into stages:

$$P_t=a\times P_{t-1} \quad (2)$$

where t is the year, P_t is the cumulative volume of literatures, a is the stage dividing point.

The research process from 2000 to 2019 can be roughly divided into the following three stages:(1) Initial period of fluctuation (2000-2009): The value of a in this period is between [1.18, 3.40], and there is a large fluctuation. It shows that the research is in the initial stage. In 2004, the literature volume increased significantly. In January 2004, the Yellow River Conservancy Commission put forward the theoretical framework of "1493" to control the Yellow River. It has caused wide concern among scholars. (2) Slow growth period (2010-2014): The average value of a in this period is about 1.17. The fluctuation range is significantly reduced. The fastest growth is from 2010 to 2011, reaching 71%. The second fastest growth is from 2012 to 2013, reaching 67%. (3) Stable development period (2015-2019): In this period, the value of a is stable at about 1.19, and the variance is 0.0001. The research enters a stable development period. This shows that scholars are paying more and more attention to this field. In the future, this topic will still be in the stable development period. The study of water resources in the YRB will still be a valuable hot topic.

3.1.2. Analysis of national comprehensive competitiveness. Within the scope of statistics, we select the top five countries with the largest number of publications for comprehensive competitiveness evaluation. First, the number of literature publications, total citation times and h-index are selected as indicators. Then, the index standardization formula (3) is used to standardize each index. Finally, the comprehensive score algorithm (4) is used to calculate the comprehensive score of each country.

$$A_{ij}=\left(x_{ij}-\bar{x}_j\right)\Big/\sqrt{\frac{\sum_i (x_{ij}-\bar{x}_j)^2}{T}} \quad (3)$$

$$A_j=\sum_j A_{ij} \quad (4)$$

where A_{ij} is the standard score of the j-th indicator of the i-th country, x_{ij} is the value of the j-th indicator of the i-th country, x_j is the average value of the j-th indicator of all countries, T is the total number of countries, and A_j is the comprehensive score of the i-th country. The scores for each country are shown in Table 1.

Table 1. Comprehensive competitiveness analysis of countries.

Country	Number of Literature	Score1	Total Citation Times	Score2	h-index	Score3	Comprehensive Score
China	533	5.702	13024	5.388	58	4.596	15.686
USA	107	0.929	4071	1.447	35	2.542	4.918
Australia	44	0.223	2791	0.883	24	1.559	2.666
Japan	19	0.057	1184	0.176	12	0.487	0.720
Canada	15	0.102	167	0.272	8	0.130	0.504

The following points can be found from the Table 1: (1) China has the strongest comprehensive competitiveness. All of its scores are higher than those of other countries. This shows that China is in a clear leading position in this field. (2) The United States is second and Australia is third. Although they have fewer publications, their total citation times and h-index are relatively high. This shows that their results have attracted a lot of attention from scientists. Their literatures have high quality and reference value. (3) Both Japan and Canada have a comprehensive score of about 0.5. Their comprehensive competitiveness is similar. They may become the rising stars in this field.

3.1.3. Analysis of core journals. We use Bradford's law, Bradford's discrete coefficient calculation method (5) and core area quantity calculation method (6) to make an analysis on the journals.

$$m=\sqrt[R]{(e^{E}\cdot Y)} \quad (5)$$

$$P=2\ln(e^{E}\cdot Y) \quad (6)$$

where m is the Bradford discrete coefficient, R is the number of divisions, Y is the number of journals with the largest number of papers, P is the number of core areas, E is the Euler coefficient, E= 0.5772. By substituting R=3 and Y=39 into the formulas, we can get m=4 and P=8. Therefore, we get journal statistical data as shown in Table 2.

Table 2. Bradford's law of journal statistics.

Area	Number of Journals	Proportion	Number of Literature	Proportion	Average Document Density
Core area	8	4.76%	190	33.81%	23.75
Related area	32	19.05%	219	38.97%	6.84
Discrete area	128	76.19%	153	27.22%	1.20
Total	168	100.00%	562	100.00%	

Core journals are the key to journals. To some extent, core journals represent the development level and the trend of a certain field. The core area of journals includes eight journals: *Water*, *Journal of Hydrology*, *Hydrological Processes*, *Water Resources Management*, *Science of the Total Environment*, *Hydrology and Earth system Sciences*, *Environmental Earth Sciences* and *Water International*. Although this area only accounts for 4.76% of the total number of journals, the amount of literature accounts 33.81% of the total literature. The average document density is as high as 23.75. This shows a strong core effect.

The source journals of highly cited literature usually involve many contents related to the theme. By comparing them with the core journals, we can expand and supplement the core journals. At the same

time, we can explore the most important points of the core journals. The frequency of core journals and source journals of highly cited literature (top 100) are shown in Figure 3.

The journals marked in red include *Journal of Hydrology*, *Hydrological Processes*, *Water Resources Management*, *Science of the Total Environment* and *Hydrology and Earth system Sciences*. They are not only from the core area of journals, but also in the top eight journal sources of highly cited literature. These journals mainly cover hydrology and environmental science. They play a key role in the study of water resources in the YRB.

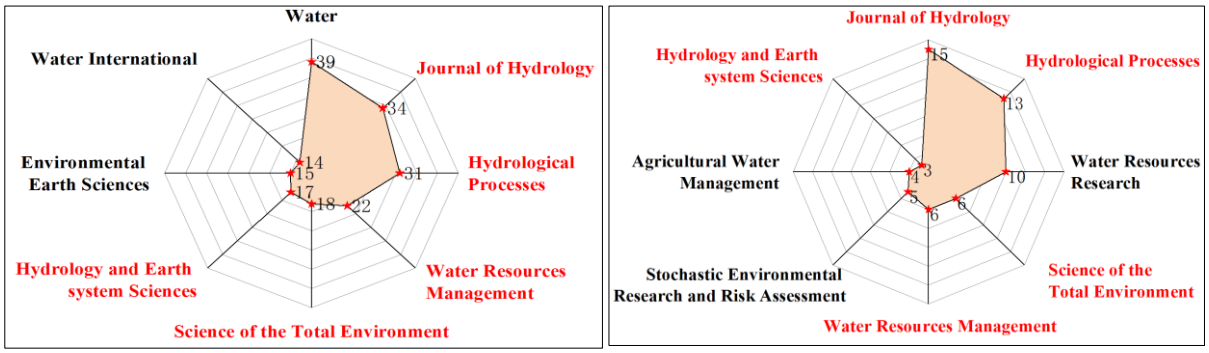


Figure 3. The frequency of core journal (left) and journal source of cited literature (right).

3.2. Hotspot analysis

The noun phrases in CiteSpace are extracted from titles, keywords, auxiliary keywords and abstracts. In CiteSpace, we use "Noun Phrase" to extract the noun phrase. Then, we set the node type as "Term" the time slice as "2 year" and the Top N as "20%". At last, we combine synonymous noun phrases, such as "human activity" and "human activities". Table 3 shows high-frequency noun phrases.

Table 3. High-frequency Noun Phrases.

Noun Phrase	Count	Centrality	Noun Phrase	Count	Centrality
water resources	222	0.09	loess plateau	35	0.08
climate change	202	0.12	hydrological model	33	0.22
Yellow River	190	0.08	potential evapotranspiration	29	0.10
Yellow River Basin	167	0.09	abrupt change	29	0.02
human activity	125	0.22	upper reaches	26	0.10
water resources management	86	0.09	decreasing trend	24	0.16
annual runoff	46	0.06	runoff change	23	0.01
hydrological process	36	0.41	river basin	20	0.22

Except for subject-related noun phrases such as "water resource", "Yellow River" and "Yellow River Basin", the most frequent noun phrases are "climate change" and "human activity". In recent years, the YRB is being severely tested by climate change and human activities. How to effectively solve these problems has become the focus of scholars' attention. The noun phrases "annual runoff" and "runoff change" show that the study of runoff change occupies a certain position. The emergence of "hydrological model" indicates that hydrological model is one of the commonly used methods in this field. The models mainly include System Hydrologic European (SHE), Soil and Water Assessment Tool (SWAT), etc.

3.3. Research framework

Literature clustering analysis can summarize the knowledge framework and clear your mind quickly. We use the data conversion function in the CiteSpace software to convert the literature into the format of .HML. Then we use the Carrot2 software to perform cluster analysis. We chose the clustering algorithm as Lingo. Figure 4 shows the literature clustering map. The bigger the bubble is, the more important the clustering will be [9].

Based on the basic analysis and hotspot analysis, the clustering map can be summarized from four aspects: research scope, research object, research method and research content. Figure 5 is the knowledge framework obtained through summary and arrangement. The research scope of this topic is mainly concentrated in the middle and upper reaches of the YRB. The research content includes surface water resources and groundwater resources. The research methods include model research and experimental research. The research content covers a wide range, mainly including: (1) runoff changes; (2) runoff-sediment relationship; (3) water resources allocation; (4) disaster response; (5) the impact of human activities and climate change; (6) ecological protection and sustainable development. These contents are not independent. For example, human activities and climate change will have an impact on runoff changes; the research on the relationship between water and sediment is conducive to strengthening ecological protection; the rational allocation of water resources and the prevention and control of disasters are important directions to promote the sustainable development of the YRB.

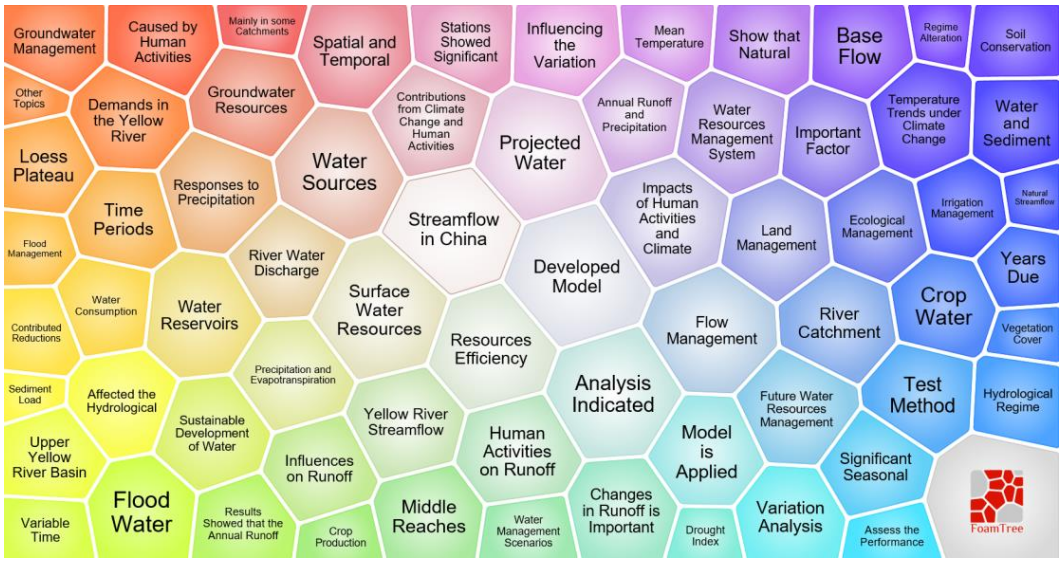


Figure 4. Literature clustering map.

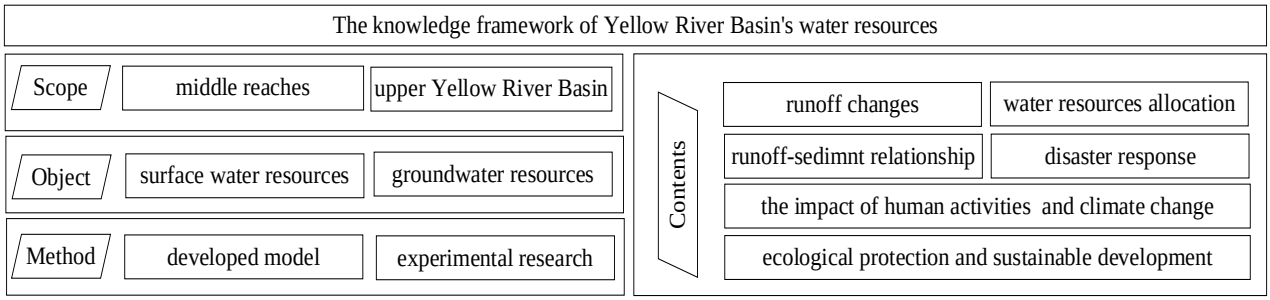


Figure 5. The knowledge framework of water resources in the YRB.

3.4. Frontier analysis

Burst detection can find things that have received particular attention. These represent the frontiers of research. After extracting noun phrases, the parameter settings are as follows: the node type is "Term", the time slice is "2 years", the clipping mode is "Pathfinder", and the "Top N" is 30. Table 4 is the noun phrase burst information table. The red line represents the duration of the burst.

Table 4. Noun phrase bursts analysis.

Noun Phrase	Strength	Begin	End	2000 - 2019
Yellow River Basin	8.022	2000	2008	
river basin	3.922	2002	2009	
Yellow River	4.236	2004	2007	
significant change	3.346	2012	2014	
annual runoff	5.111	2015	2016	
decreasing trend	3.411	2016	2017	
loess plateau china	3.519	2016	2019	

The highest strength of the burst is "Yellow River Basin (8.022)". The period of emergence is from 2000 to 2008. The "river basin (3.922)" and "Yellow River (4.236)" appeared during this period. The second highest strength of the burst is "annual runoff (5.111)". Due to the influence of climate change and human activities, the runoff of the YRB has changed significantly. Scholars pay more and more attention to the causes and trends of runoff fluctuation [10]. They tried to improve the system of water resources management. The ultimate goal is to promote the comprehensive development and effective utilization of water resources in the YRB. The "Loss plateau china (3.519)" is the latest noun phrase. The period of emergence is from 2016 to 2019. It will become the research frontier in the future. The high sediment concentration of the Yellow River is largely caused by erosion on the Loess Plateau. This is also the reason why scholars focus on the study of the Loess Plateau. In recent years, the research on the Loess Plateau also include the study on soil erosion [11], the study on the project of returning farmland to forest and grass [12] and the study on the construction of warping dams [13].

4. Conclusions

- The research process of water resources in the YRB is divided into three stages: the initial period of fluctuation (2000-2009), the slow growth period (2010-2014) and the stable development period (2015-2019). The literatures are mainly published in journals related to hydrology and environmental sciences. At the national level, China has the strongest comprehensive competitiveness.
- The scope of the study is mainly concentrated in the middle and upper reaches of the Yellow River. It mainly uses models and experiments to study surface water and groundwater. The research covers a wide range, including runoff changes, runoff-sediment relationship, water resources allocation, disaster response, the impact of human activities and climate change, as well as ecological protection and sustainable development. The impacts of human activities and climate change are the research focus.
- In the future, we should aim to promote the ecological protection and high-quality development of the YRB. Scholars should adhere to the belief of green and sustainable development and focus on the research of the Loess Plateau. We should solve the practical problems existing in the YRB and make the Yellow River truly a happy river for the people.

Acknowledgments

This paper was partly funded by the National Key Research and Development Program of China (Grant No. 2016YFC0401409), National Natural Science Foundation of China (Grant No. 51679188,

51509201), Young Technology Star in Shaanxi Province of China (Grant No. 2020KJXX-092). The authors thank the editor and anonymous reviewers for their valuable comments and suggestions.

References

[1] Liu Q, Yang Z F and Cui B S 2008 *J. Hydrol.* **361** 330-8
[2] Bai C L 2020 *Bull. Chin. Acad. Sci.* **35** 138-44
[3] Zheng H X, Zhang L and Zhu R R 2009 *Water Resour. Res.* **45** W00A19
[4] Omer A, Ma Z G and Zheng Z Y 2020 *Sci. Total Environ.* **704** 135428
[5] Zuo D P, Xu Z X and Yao W Y 2016 *Sci. Total Environ.* **544** 238-50
[6] Wang C Y, Zou Y M and Yu L J 2020 *Environ. Sci. Poll. Res.* **27** 9780-89
[7] Di D Y, Wu Z N and Wang H L 2020 *Water Resour. Manag.* **34** 245-62
[8] Chen C M 2010 *J. Am. Soc. Inform. Sci. Technol.* **61** 1386-09
[9] Jin Y R and Li X 2019 *Multimed. Tools Appl.* **78** 1289-13
[10] Miao C Y and Ni J R 2009 *Int. J. Environ. Res. Public Health* **6** 2849-64
[11] Cai J Y, Zhou Z H and Liu J J 2019 *J. Hydrol.* **578** 124005
[12] Yang B, Wang Q J and Xu X T 2018 *Environ. Earth Sci.* **77** 304
[13] Yuan S L, Li Z B and Li P 2019 *J. Water* **11** 834