

# Study on temporal and spatial evolution characteristics of surface water in Taolai River Basin based on random forest algorithm

Fenhua Li, Wenju Yang\*, Dekui Kang

Water Resources Utilization Center of Taolai River Basin, Gansu Provincial Water Resources Department, Jiuquan 735000, China

**Abstract.** The Taolai River Basin is a typical arid area in Northwest China. In order to accurately grasp the spatial distribution and dynamic changes of surface water in the Taolai River Basin, the spatial and temporal evolution sources of surface water in the study area were explored. In this study, the random forest algorithm was used to classify the land use in the Taolai River Basin, and the land classification map based on remote sensing data was made. The temporal and spatial evolution characteristics of surface water in the Taolai River Basin were analyzed by confusion matrix. From 2016 to 2018, the area of surface water in the Taolai River Basin increased from 74.51 million m<sup>2</sup> to 82.35 million m<sup>2</sup>, the area of surface water transferred out was 3.66 million m<sup>2</sup>, and the area of surface water transferred in was 11.50 million m<sup>2</sup>. The main sources of area growth were bare land, grassland and farmland. From 2018 to 2020, the surface water area increased from 82.35 million m<sup>2</sup> to 96.41 million m<sup>2</sup>. Compared with 2016, the net growth rate of surface water area was 179.39%. The main sources of area growth were bare land, grassland, and snow. The surface water area in the Taolai River Basin is increasing year by year. The main sources of surface water area growth are bare land and grassland, and the transfer-in contribution rate is more than 84.87%. This study provides a scientific reference for the digital monitoring of surface water and the rational utilization of water and soil resources in the Taolai River Basin.

**Keywords:** Taolai River Basin; surface water bodies; spatio-temporal evolution of land use; random forest algorithm.

## 1. Introduction

Land use and land use conversion are critical to determine the impact of human activities on the environment [ 1 ]. In recent years, human activities such as the expansion of construction land have occupied a large amount of cultivated land, water and ecological land, resulting in major changes in land use pattern [ 2-4 ]. Through the land use dynamic change data, we can quickly understand the migration intensity and migration trend of different land types in the study area, so as to effectively manage the land resources of the basin [ 5-6 ]. Land use change matrix is an important method to describe and measure land use change. It can directly obtain the type of land use migration and the area of migration block, and then analyze the change trend of different types of land use in the study area [ 7 ]. Therefore, the study of land use in the Taolai River Basin from 2016 to 2020 and the analysis of the temporal and spatial evolution of surface water are the basis for the efficient management of water resources in the future.

At different spatial and temporal scales, remote sensing technology has become a powerful tool to obtain the earth's surface feature data [ 8 ]. The use of high-resolution remote sensing images to classify ground objects is widely used in land planning, ecological monitoring and resource protection [ 9-10 ]. For example, Zhang et al. [ 11 ] used Landsat remote sensing image data to analyze the land use data of Sanya City in 2004, 2009, 2014 and 2019, and found that the cultivated land area of Sanya City was effectively restored from 2004 to 2019. Wang et al. [ 12 ] used MODIS-NDVI remote sensing vegetation index to analyze the spatial and temporal variation characteristics of land use classification in Guangdong Province from 2000 to 2020, and found that the main reason for vegetation coverage degradation in Guangdong Province was the conversion of forest land and cultivated land to construction land. Li et al. [ 13 ] used the remote sensing data of 1982 and 2016 to analyze the land use of Longchuan River, and found that the change range of

plantation was the largest, and its area showed an increasing trend in the time series. In addition, Tobias et al. [ 14 ] explored the changes of farmland, pasture, forest and urban areas in Europe from 1990 to 2006 by using high-resolution and spatially explicit land use change indicators. Michele et al. [ 15 ] combined MODIS remote sensing land cover data set and FAOSTAT land use data to quantify the changes between land types from 2001 to 2012, providing key information for assessing the environmental impact caused by LULUC.

In this paper, the random forest ( RF ) algorithm is used to classify the remote sensing images of the Taolai River Basin in 2016,2018 and 2020 according to the types of farmland, forest, shrub, grassland, water, snow, bare land, impervious surface and wetland, and the land classification image is made. The land transfer matrix is used to analyze the evolution trend of land types in the Taolai River Basin from 2016 to 2020, to explore the temporal and spatial evolution characteristics of surface water in the Taolai River Basin, and to provide theoretical reference for land use management and sustainable development of water resources in the Taolai River Basin.

## 2. Overview of the study area and research methods

### 2.1 Overview of the study area

The Taolai River Basin is located in the middle of the Hexi Corridor. It starts from the Maying River in the east, reaches the boundary between the Taolai River and the Shushui River in the west, bounded by the Qilian Mountain Ridge in the south and Mazong Mountain in the north. It is a first-class tributary of the Heihe River system. The geographical location is between  $47^{\circ}29' - 99^{\circ}19'$  east longitude and  $38^{\circ}24' - 40^{\circ}26'$  north latitude. The basin area is  $2.7 \times 10^4 \text{ km}^2$ , of which  $19,300 \text{ km}^2$  in Gansu. The Taolai River Basin in Gansu Province mainly includes the Sunan County of Zhangye City, Jiayuguan City, Suzhou District of Jiuquan City and Jinta County.

### 2.2 Data source and processing

In order to explore the land use types and migration and transformation trends in the Taolai River Basin from 2016 to 2020, the long-term sequence remote sensing image data source used is Landsat 8 multi-spectral satellite with a data resolution of 30 meters. The data is based on Landsat surface reflectance data to calculate input characteristics, including spectrum, phenology and topography. Considering that its distribution is related to its geographical location, geographical coordinates are also selected as input data. The time span of multiple remote sensing images from 2016 to 2020 is 12 natural months, and cloudless blocks are selected for preprocessing such as fusion, clipping and splicing.

### 2.3 Research methods

#### 2.3.1 Remote sensing image classification method

In this paper, the random forest algorithm is used to classify the remote sensing images of the Taolai River Basin in 2016,2018 and 2020. Random forest is a very representative ensemble learning algorithm [ 16-19 ]. The principle of the method is shown in Figure 2, and the types are shown in Table 1. The random forest classifier is composed of a series of decision tree classifier sets. Each decision tree is defined as  $h_1(X), h_2(X), h_k(X)$ . The training samples are defined as  $T(X, Y)$ ,  $X$  is sample attribute,  $Y$  is sample categories, and the Margin Function is defined as:

$$mg(X, Y) = av_k(I(h_k(X)=Y)) - \max_{j \neq Y} av_k(I(h_k(X) = j)) \quad (1)$$

In the formula,  $I$  is the indicator function;  $av_k(\bullet)$  denotes average;  $j$  is the vector of classification error. The edge function represents the degree to which the number of correctly classified votes in the  $T(X, Y)$  exceeds the number of incorrectly classified votes. The higher the confidence of the model, the larger the edge function.

Table 1 Land use classifica

Serial number	Type
1	Cropland
2	Forest
3	Shrub
4	Grassland
5	Water
6	Sonw/Ice
7	Barren
8	Impervious
9	Wetland

2.3.2 Land use transfer matrix

The transfer matrix of land use type is derived from the quantitative description of system state and state transfer in system analysis [ 20 ]. As shown in Table 2,  $A_1$  to  $A_n$  are various land types in land use, and  $S_{nn}$  represents the conversion area of the two land types in the corresponding year ( when  $nn$  is the same, it represents the area that has not changed ). From the land transfer matrix, we can get the change of various types of area, that is, the area change from  $P_{+n}$  to  $P_n + ( P_{n+} - P_{+n} )$  from  $T_1$  to  $T_2$ .

Table 2 Land use transfer matrix

Project		$T_2$				Sum	Decrement
		$A_1$	$A_2$	...	$A_n$		
$T_1$	$A_1$	$S_{11}$	$S_{12}$	...	$S_{1n}$	$P_{1+}$	$P_{1+}-S_{11}$
	$A_2$	$S_{21}$	$S_{22}$	...	$S_{2n}$	$P_{2+}$	$P_{2+}-S_{22}$
	...	...	...	...	...	...	...
	$A_3$	$S_{n1}$	$S_{n2}$	...	$S_{nn}$	$P_{n+}$	$P_{n+}-S_{nn}$
Sum		$P_{+1}$	$P_{+2}$		$P_{+n}$		
Increment		$P_{+1}-S_{11}$	$P_{+2}-S_{22}$	...	$P_{+n}-S_{nn}$		

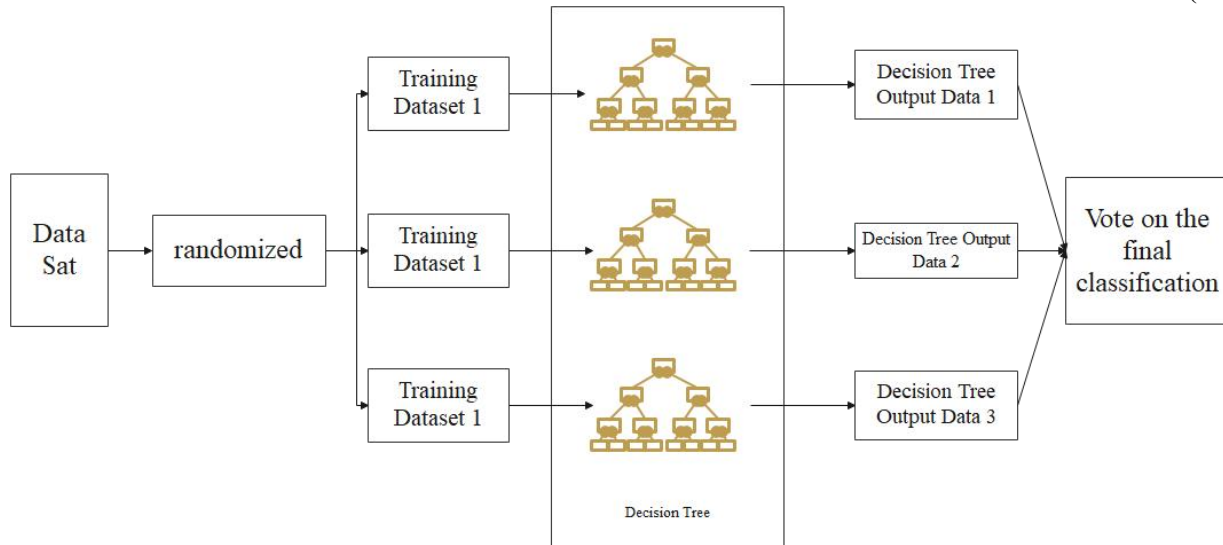


Fig 2 Principle of random forest classification algorithm

### 3. Results and analysis

#### 3.1 Remote sensing image classification results of Taolai River Basin

The RF classifier is used for land use classification, and similar plots are merged. According to Google Earth and real feature labels, training and verification are carried out to form three-year land use data sets for 2016,2018,2020, and 2020. The classification results are shown in Figure 3. It is verified that the overall accuracy of different land use types in the data set is 76.45 % -82.51 %, the average OA is  $79.30 \pm 1.99$  %, and the Kappa coefficient is 0.82. The accuracy of the data set is high and can be used for further research.

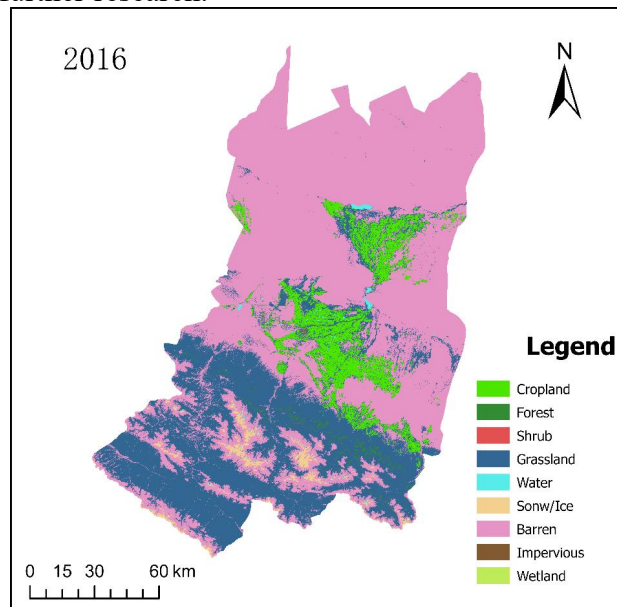


Fig. 3 Remote sensing image classification of cultivated land in Taolai River based on RF (2016)

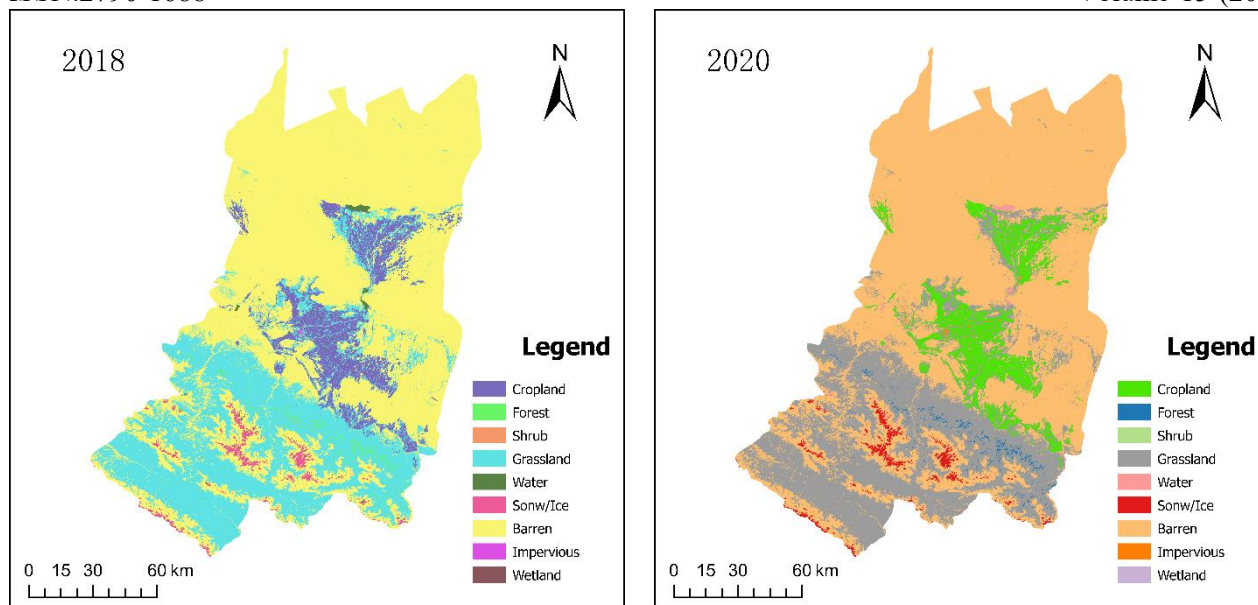


Fig. 4 Remote sensing image classification of cultivated land in Taolai River based on RF (2018-2020)

The area of farmland in the Taolai River Basin continued to grow from 2016 to 2020, and the proportion of the basin area increased from 7.36 % to 7.53 %, an increase of 0.17 %. From 2016 to 2020, the forest area remained stable, and the grassland area showed a fluctuating trend. From 2016 to 2018, the grassland area increased slightly, and the area ratio increased from 28.70 % to 28.84 %. From 2018 to 2020, it decreased significantly, from 6967.63 million m<sup>2</sup> to 6828.83 million m<sup>2</sup>, and the area ratio decreased by 0.58 %. From 2016 to 2020, the water area will remain balanced, and the proportion of the basin area will remain above 0.3 %. From 2016 to 2020, the bare land area in the Taolai River Basin fluctuated in a small range. From 2016 to 2018, the bare land area decreased by 18.79 million m<sup>2</sup>, and the bare land area increased by 65.05 million m<sup>2</sup> from 2018 to 2020. The specific data are shown in Table 3.

Table 3 Land Use Classification Area in the TaoLai River Basin, 2016-2020

Land type	2016		2018		2020	
	Area (million m <sup>2</sup> )	Proportion (%)	Area (million m <sup>2</sup> )	Proportion (%)	Area (million m <sup>2</sup> )	Proportion (%)
Farmland	1773.28	0.07	1779.30	0.07	1818.95	0.08
Forest	108.94	<0.01	113.12	<0.01	113.47	<0.01
Shrub	3.22	<0.01	4.03	<0.01	7.58	<0.01
Grassland	6933.58	0.29	6967.63	0.29	6828.83	0.28
Water	74.51	<0.01	82.35	<0.01	96.41	<0.01
Snow	297.44	0.01	262.89	0.01	277.90	0.01

Bare land	14955.89	0.62	14937.10	0.62	15002.15	0.62
Impervious ground	14.05	<0.01	14.49	<0.01	15.63	<0.01
Wetland	<0.01	<0.01	<0.01	<0.01	0.01	<0.01

### 3.2 Analysis of dynamic change of surface water area based on land use transfer matrix

In order to further explore the dynamic change characteristics of surface water in the study area in the spatial and temporal change pattern, the transfer matrix is used to directly represent the composition and area proportion of the feature types. With the help of the transfer matrix calculation results, the transfer trend of surface water and the contribution degree of the transfer-in and transfer-out plots can be obtained. The transfer matrix calculation results of various types of objects in the study area from 2016 to 2018 are shown in Table 4, and the transfer matrix calculation results of various types of objects from 2018 to 2020 are shown in Table 5.

During 2016-2018, the total area of the Taolai River Basin was 2416.10 million m<sup>2</sup>, the surface water area increased from 74.51 million m<sup>2</sup> to 82.35 million m<sup>2</sup>, and the net surface water area increased by 7.84 million m<sup>2</sup>. From 2016 to 2018, a total of 70.85 million m<sup>2</sup> of surface water in the Taolai River Basin was not converted into other land use types. The surface water transfer area was 3.66 million m<sup>2</sup>, and the surface water transfer area was 11.50 million m<sup>2</sup>. The net transfer of bare land to surface water area was 3.90 million m<sup>2</sup>, the net transfer of farmland to surface water area was 0.87 million m<sup>2</sup>, and the net transfer of grassland to surface water area was 2.75 million m<sup>2</sup>. The main sources of surface water area growth were bare land, grassland and farmland, with contribution rates of 49.78 % and 35.09 %, and 11.12 %.

During the period of 2018-2020, the total area of the Taolai River Basin was 24160.91 million m<sup>2</sup>, the surface water area increased from 82.35 million m<sup>2</sup> to 96.41 million m<sup>2</sup>, and the net increase of surface water area was 14.06 million m<sup>2</sup>. Compared with the growth rate of 2016-2018, the net growth rate of surface water area in 2018-2020 was 179.39 %. From 2018 to 2020, a total of 74.45 million m<sup>2</sup> of surface water in the Taolai River Basin was not converted into other land use types. The surface water transfer area was 7.90 m<sup>2</sup>, and the surface water transfer area was 21.96 million m<sup>2</sup>. The net transfer of bare land to surface water area was 7.81 million m<sup>2</sup>, the net transfer of grassland to surface water area was 5.23 million m<sup>2</sup>, and the net transfer of snow to surface water area was 1.40 million m<sup>2</sup>. The main sources of surface water area growth were bare land, grassland, and snow, with contribution rates of 55.52 % , 37.23 %, and 9.99 %. In addition, the conversion of farmland and water showed negative growth from 2018 to 2020, and the net conversion water area of farmland was -385.2 thousand m<sup>2</sup>.

In summary, the net growth of surface water area in the Taolai River Basin in 2016-2018 is lower than that in 2018-2020, and the source of surface water area growth during 2016-2020 is in the process of dynamic change. The main sources of surface water transfer are bare land and grassland. The contribution rate of transfer in 2016-2018 is 84.87 %, and the contribution rate of transfer in 2018-2020 is 92.75 %. The third largest source of surface water in 2016-2018 is farmland, and the contribution rate of transfer is 11.12 %. The third largest source of surface water in 2018-2020 is snow. The transfer-in contribution is 9.99 %. In the process of surface water conversion, the conversion of various land types into surface water bodies in 2016-2018 was positive, and the conversion of farmland water bodies in 2018-2020 showed negative growth.

Table 4 Transfer matrix of land use types in the study area, 2016-2018 Unit: million m<sup>2</sup>

Land type		2018									
		Farmland	Forest	Shrub	Grassland	Water	Snow	Bareland	Impervious ground	Wetland	Total
2016	Farmland	1730.93	0	0	41.16	1.13	0	0	0.06	0	1773.28
	Forest	0	108.47	0.46	0	0	0	0	0	0	108.94
	Shrub	0	0.23	2.87	0.12	0	0	0	0	0	3.22
	Grassland	28.26	4.32	0.7	6753.88	3.91	0	142.32	0.2	0	6933.58
	Water	0.26	0.01	0	1.16	70.85	0	2.23	0.01	0	74.51
	Snow	0	0	0	0.4	0.32	260.67	36.05	0	0	297.44
	Bareland	19.85	0.09	0	170.91	6.13	2.22	14756.5	0.19	0	14955.89
	Impervious ground	0	0	0	0	0.01	0	0	14.04	0	14.05
	Wetland	0	0	0	0	0	0	0	0	0	0
	Total	1779.3	113.12	4.03	6967.63	82.35	262.89	14937.1	14.49	0	24160.91

Table 5 Transfer matrix of land use types in the study area, 2018-2020 Unit: million m<sup>2</sup>

Land type		2020									
		Farmland	Forest	Shrub	Grassland	Water	Snow	Bareland	Impervious ground	Wetland	Total
2018	Farmland	1760.34	0	0	18.7	0.14	0	0	0.12	0	1779.3
	Forest	0.03	111.21	1.87	0	0	0	0	0	0	113.12
	Shrub	0	0.04	2.85	1.14	0	0	0	0	0	4.03

Grass land	45.82	2.21	2.86	6657.65	11.07	0	247.29	0.73	0	6967.63
Water	0.52	0	0	5.84	74.45	0	1.54	0	0	82.35
Snow	0	0	0	0.16	1.4	254.85	6.48	0	0	262.89
Bare land	12.23	0	0	145.35	9.34	23.05	14746.84	0.28	0	14937.1
Imperious ground	0	0	0	0	0	0	0	14.49	0	14.49
Wetland	0	0	0	0	0	0	0	0	0	0
Total	1818.95	113.47	7.58	6828.83	96.41	277.9	15002.15	15.63	0	24160.91

#### 4. Conclusions

(1) In 2016, the surface water area of the basin was 74.51 million m<sup>2</sup>, accounting for 0.31 % of the total area of the Taolai River Basin. In 2018, the area of farmland in the Taolai River Basin was 82.35 million m<sup>2</sup>, accounting for 0.34 % of the total area of the Taolai River Basin. In 2020, the area of farmland in the basin was 96.41 million m<sup>2</sup>, accounting for 0.40 % of the total area of the Taolai River Basin.

(2) From 2016 to 2020, the surface water area in the study area showed an overall upward trend, and the growth rate from 2018 to 2020 was significantly higher than that from 2016 to 2018 ; in terms of dynamic changes, the transfer relationship between various blocks in the study area is relatively stable. The conversion of bare land and grassland into surface water accounts for more than 84 %, which is the main source of surface water transfer-in. The area of surface water transferred from bare land and grassland also has a significant upward trend, from 84.87 % in 2016-2018 to 92.75 % in 2018-2020. Other important transfer-in factors change from farmland to snow. The main transfer-in sources and dynamic change sources should be comprehensively considered to formulate a more reasonable water and soil resources development and utilization plan.

#### Acknowledgements

This study was supported by the Gansu Provincial Water Resources Department 2023 the first batch of provincial water resources fee projects (Gan Water Resources Development [2023] No. 88), Gansu Provincial Water Resources Department 2022 the first batch of provincial water resources fee projects (Gan Water Resources Development [2022] No. 94).

#### References

- [1] De Rosa M, Vestergaard Odgaard M, Staunstrup J K, et al. Identifying land use and land-use changes (LULUC): a global LULUC matrix[J]. Environmental science & technology, 2017, 51(14): 7954-7962.

- [2] LUO S Q, HU X M, SUN Y, YAN C, ZHANG X. Multi-scenario land use change and its impact on carbon storage based on coupled Plus-Invest model[J]. Chinese Journal of Eco-Agriculture, 2023, 31(2): 300-314
- [3] Hu F, Zhang Y, Guo Y, et al. Spatial and temporal changes in land use and habitat quality in the Weihe River Basin based on the PLUS and InVEST models and predictions[J]. Arid. Land Geogr, 2022, 45: 1125-1136.
- [4] Liu Q, Yang D, Cao L, et al. Assessment and Prediction of Carbon Storage Based on Land Use/Land Cover Dynamics in the Tropics: A Case Study of Hainan Island, China[J]. Land, 2022, 11(2): 244.
- [5] Talebi Khiavi H, Mostafazadeh R, Asaadi M A, et al. Temporal land use change and its economic values under competing driving forces in a diverse land use configuration[J]. Arabian Journal of Geosciences, 2022, 15(20): 1-14.
- [6] Lin C, Wu C C, Tsogt K, et al. Effects of atmospheric correction and pansharpening on LULC classification accuracy using WorldView-2 imagery[J]. Information Processing in Agriculture, 2015, 2(1): 25-36.
- [7] Sun J W, Wang H Q, Zhang Y F. Analysis of land use dynamic in Genhe city based on GIS and RS[C]//Applied Mechanics and Materials. Trans Tech Publications Ltd, 2013, 256: 2298-2302.
- [8] Liang D, Zuo Y, Huang L, et al. Evaluation of the consistency of MODIS Land Cover Product (MCD12Q1) based on Chinese 30 m GlobeLand30 datasets: A case study in Anhui Province, China[J]. ISPRS International Journal of Geo-Information, 2015, 4(4): 2519-2541.
- [9] Wang Yao, Chen Ruishan, Guo Chihui, Xia Zilong. 2021. Study on the pattern change and regional differentiation of resources and environment in the Yellow River Basin and suggestions on eco-geological survey[J]. Geology in China, 48(1):1-20.
- [10] Shi Y, Wang F, Wu Z. Multi - simulation of spatial distribution of land use based on CLUE - S in Jinhe Watershed[J]. Remote Sensing for Land and Resources, 2016, 28(2):154-160.
- [11] Zhang Juan, Huang Jiajian, Zhao Runjiang, etc. Study on spatial and temporal changes of land use landscape pattern in Sanya City [J]. Journal of Northwest Forestry University, 2022,37 (6): 224-231.
- [12] Wang S, Zhang L, Lin W, et al. Study on vegetation coverage and land-use change of Guangdong Province based on MODIS-NDVI[J]. Acta Ecol. Sin, 2022, 42: 2149-2163.
- [13] Li C, Zeng H. Study on land use change in Longchuan River Basin based on area transfer matrix [J]. Yangtze River, 2018,49 (17): 39-44 + 51.
- [14] Kuemmerle T, Levers C, Erb K, et al. Hotspots of land use change in Europe[J]. Environmental research letters, 2016, 11(6): 064020.
- [15] De Rosa M, Vestergaard Odgaard M, Staunstrup J K, et al. Identifying land use and land-use changes (LULUC): a global LULUC matrix[J]. Environmental science & technology, 2017, 51(14): 7954-7962.
- [16] Qin X W, Cheng B, Yang Z P, et al. Identification of parcel-scale crop types in southwestern mountainous area based on time series remote sensing images[J]. Journal of Geo-information Science, 2023,25(3):654-668.
- [17] Chang X, Xing Y, Gong W, et al. Evaluating gross primary productivity over 9 China Flux sites based on random forest regression models, remote sensing, and eddy covariance data[J]. Science of The Total Environment, 2023: 162601.
- [18] Peng H, Zhou Y, Hu X, et al. 2023. A PM 2.5 prediction model based on deep learning and random forest. National Remote Sensing Bulletin, 27(2):430-440.
- [19] Bao F, Huang K, Wu S. The retrieval of aerosol optical properties based on a random forest machine learning approach: Exploration of geostationary satellite images[J]. Remote Sensing of Environment, 2023, 286: 113426.
- [20] [ 20 ] Yang N, Wang L, Zhu L, et al. Characteristics of NPP variation and driving factors in the Yellow River Basin in recent ten years [J].Journal of Applied Basic and Engineering Sciences, 2023,31(2): 280-295.