



## OPEN ACCESS

EDITED BY  
Qiuming Pei,  
Southwest Jiaotong University, China

REVIEWED BY  
Himan Shahabi,  
University of Kurdistan, Iran  
Shaohua Zhao,  
Ministry of Ecology and Environment  
Center for Satellite Application on Ecology  
and Environment, China  
Salim Heddam,  
University of Skikda, Algeria

\*CORRESPONDENCE  
Shan Zhao,  
✉ 475283764@qq.com

SPECIALTY SECTION  
This article was submitted to  
Structural Geology and Tectonics,  
a section of the journal  
Frontiers in Earth Science

RECEIVED 07 December 2022  
ACCEPTED 20 January 2023  
PUBLISHED 02 February 2023

CITATION  
Guo H, Dai W, Zhang R, Zhang D, Qiao B,  
Zhang G, Zhao S and Shang J (2023),  
Mineral content estimation for salt lakes on  
the Tibetan plateau based on the genetic  
algorithm-based feature selection method  
using Sentinel-2 imagery: A case study of  
the Bieruoze Co and Guopu Co lakes.  
*Front. Earth Sci.* 11:1118118.  
doi: 10.3389/feart.2023.1118118

COPYRIGHT  
© 2023 Guo, Dai, Zhang, Zhang, Qiao,  
Zhang, Zhao and Shang. This is an open-  
access article distributed under the terms  
of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Mineral content estimation for salt lakes on the Tibetan plateau based on the genetic algorithm-based feature selection method using Sentinel-2 imagery: A case study of the Bieruoze Co and Guopu Co lakes

Hengliang Guo<sup>1</sup>, Wenhao Dai<sup>2</sup>, Rongrong Zhang<sup>2</sup>, Dujuan Zhang<sup>1</sup>,  
Baojin Qiao<sup>2</sup>, Gubin Zhang<sup>3</sup>, Shan Zhao<sup>2\*</sup> and Jiandong Shang<sup>1</sup>

<sup>1</sup>National Supercomputing Center in Zhengzhou, Zhengzhou University, Zhengzhou, China, <sup>2</sup>School of Geoscience and Technology, Zhengzhou University, Zhengzhou, China, <sup>3</sup>Henan Geological Research Institute, Zhengzhou, China

Salt lakes on the Tibetan Plateau (TP) are rich in lithium (Li), boron (B) and other mineral resources, and accurate assessment of the mineral content and spatial distribution of the brine in those salt lakes is important to guide the development and utilization of their mineral resources. There are few studies estimating the mineral content of salt lakes on the TP due to the lack of *in situ* investigation data. This study introduced an intelligent prediction model combining a feature selection algorithm with a machine learning algorithm using Sentinel-2 satellite data to estimate the Li, B, and TDS contents of Bieruoze Co and Guopu Co lakes on the TP. First, to enrich the spectral information, four mathematical transformations (reciprocal, logarithmic, reciprocal of logarithm, and first-order derivative) were applied to the original bands. Then, feature selection was performed using the genetic algorithm (GA) to select the optimal input variables for the model. Finally, prediction models were constructed by partial least squares regression (PLSR), multiple linear regression (MLR), and random forest (RF). The results showed that: 1) The spectral mathematical transformation provided rich spectral information for the mineral content estimation. 2) The performance of the estimation model constructed by the feature optimization method using GA was better than that of the estimation model constructed based on all spectral bands. Based on GA for feature optimization, the MAPE of GA-RF for estimating Li, B and TDS contents on the testing set was reduced by 77.52%, 28.54% and 36.79%, respectively. 3) Compared with the GA-MLR and GA-PLSR models, GA-RF estimated Li ( $R^2=0.99$ , RMSE=1.15 mg L<sup>-1</sup>, MAPE=3.00%), B ( $R^2=0.97$ , RMSE=10.65 mg L<sup>-1</sup>, MAPE=2.73%), and TDS ( $R^2=0.93$ , RMSE=0.60 g L<sup>-1</sup>, MAPE=1.82%) all obtained the optimal performance. This study showed that the combination of the GA-based feature selection method and the RF model has excellent performance and applicability for monitoring the content of multiple minerals using Sentinel-2 imagery in salt lakes on the TP.

## KEYWORDS

mineral content, salt lake, Tibetan plateau, Sentinel-2, random forest, genetic algorithm, feature selection

## 1 Introduction

The Tibetan Plateau (TP), known as the “Water Tower of Asia”, is rich in lake resources, most of which are saltwater lakes and saline lakes (Ma et al., 2011; Liu et al., 2021; Qiao et al., 2021). The brine of salt lakes is not only high in salinity but also rich in potassium, magnesium, lithium (Li), boron (B), uranium and other salt resources, which have high development potential and strategic value. With the advancement of Li battery technology, Li has gained notoriety as a strategic asset. Li brine resources are abundant in China, which has the third-largest Li reserves in the world. Li brines are primarily found on the TP. Due to the growing demand and the lower production cost of Li extraction from brines, the technology for Li extraction from salt lake brines has received much attention (He et al., 2020; Zhang et al., 2022). B is a crucial essential raw ingredient for the production of ceramics, detergents, fertilizers, and glass. Salt lake brines account for 80% of the world's Li and 25% of its B salt production (Kong et al., 2021). Total dissolved solids (TDS) reflect the overall content of anions and cations in salt lakes. The TP is relatively minimally influenced by humans, and changes in brine mineral content are mainly affected by natural conditions. In recent years, due to climate change and other factors, some salt lakes on the TP have shown varying degrees of lake desalination and decreases in brine mineral content (Yan and Zheng, 2015). Investigation and monitoring of the mineral content of salt lakes can provide an important theoretical basis for the study of the mineralization law, mineralization mechanism, and geological survey of salt lakes in the TP region (Ding et al., 2022).

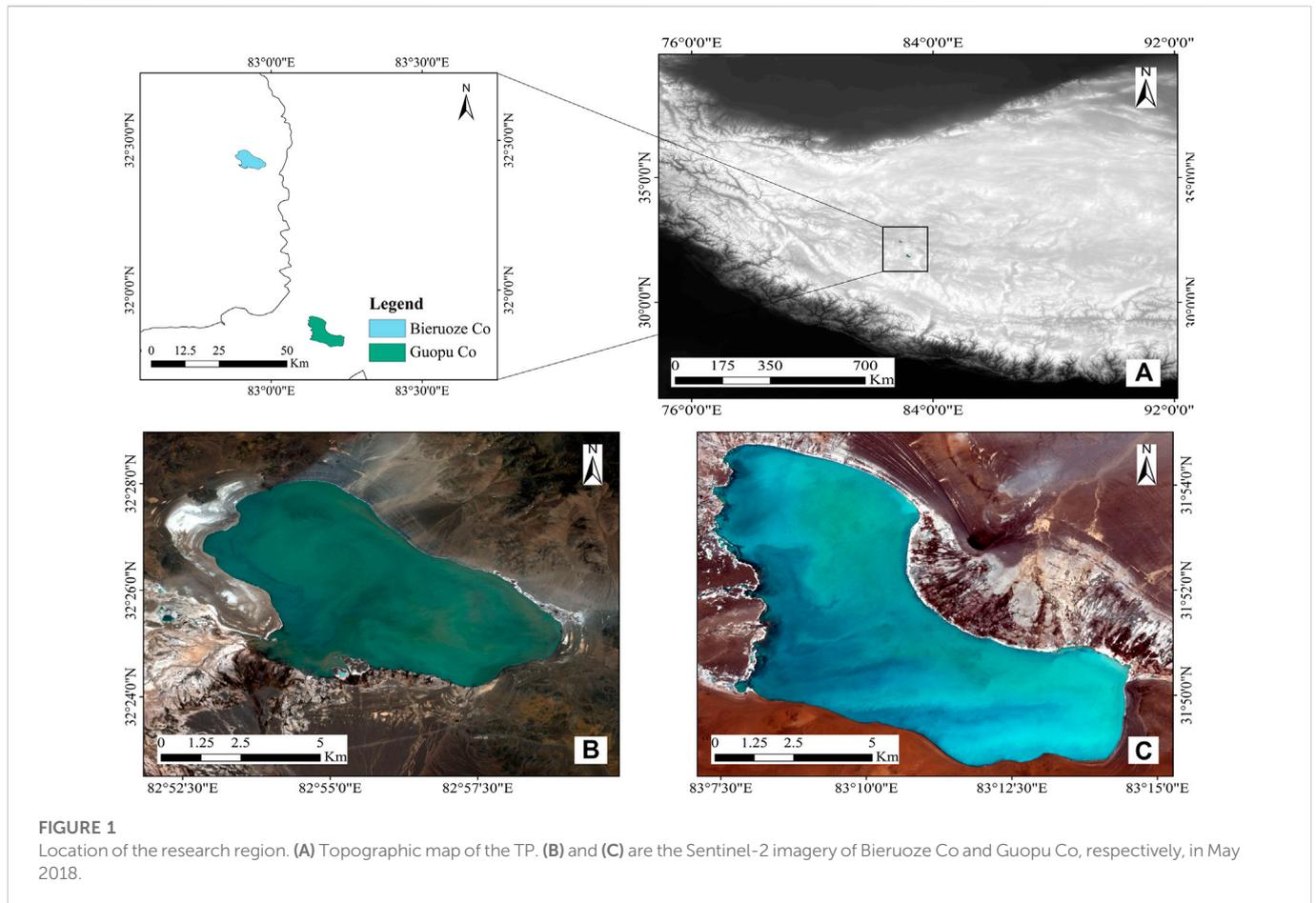
Traditional monitoring methods involve conducting field sampling and laboratory analysis. However, owing to the extreme natural environment of the TP and the poor accessibility of some salt lakes, traditional methods cannot reflect the spatial distribution of the mineral content in salt lakes as a whole. The use of remote sensing to estimate water parameters has been used in salt lakes (Wang et al., 2015; Wang et al., 2021), estuaries (Geiger et al., 2013; Fang et al., 2017), seas (Chen and Hu, 2017), inland freshwater lakes (Bayati and Danesh-Yazdi, 2021; Li et al., 2022) and other regions. Compared with traditional measures, the use of satellite remote sensing to monitor mineral content has the advantages of a large range, long duration, and periodicity (Sun et al., 2022), which overcome the shortcomings and limitations of traditional methods. Remote sensing technology can be used to quickly and accurately search for areas with high mineral content in the salt lake and to fully grasp the spatial distribution and changes in resources, thus scientifically guiding the development and production of mineral resources in the salt lake. The multispectral satellite data have high spatial resolution and are suitable for mineral content studies of salt lakes on the TP. Sentinel-2 is a multispectral satellite with publicly available free data that is commonly used for hydrological remote sensing studies (Miles et al., 2017; Marinho et al., 2021).

The models commonly used for hydrological remote sensing are multiple linear regression (MLR) (Chen and Hu, 2017), partial least squares regression (PLSR) (Song et al., 2013; Cao et al., 2018), random forest (RF) models (Hafeez et al., 2019; Cao et al., 2020; Maier et al., 2021; Sun et al., 2022), and artificial neural network models (Bayati and Danesh-Yazdi, 2021). Due to the lack of actual measurement data, there are few studies on the estimation of Li, B, and TDS contents of salt lakes on the TP. The available studies estimated the mineral content of the salt lake mainly through empirical and machine learning models. For example, Zhang et al. (2007) used the ratio

method and principal component analysis to reveal the spatial distribution pattern of boron oxide content in Zabuye Salt Lake. This method is easy to implement, but the estimation accuracy is low. The second approach used machine learning algorithms to construct estimation models. For example, Zhou et al. (2016) used an adaptive band selection method to determine the optimal band combination and a BP neural network algorithm to construct an inversion model for the ion content of the salt lake. Liu et al. (2021) used the LightGBM algorithm to invert the Li content of the Zabuye salt lake. Machine learning methods have been shown to be a better way to address complex problems without prior knowledge (Saberioon et al., 2020), and machine learning methods can address non-linear and other complicated regression issues. Therefore, machine learning models also have great potential for mineral content estimation in salt lakes.

Wang (2019) performed mathematical transformations such as logarithmic transformation and first-order differential transformation on Sentinel-2 data and predicted the Li content of Alisallo salt lake, and the results showed that spectral transformation played an important role in the prediction model. Spectral transformation has been proven to be an effective spectral preprocessing method. The spectral mathematical transformation can enrich the spectral information and extract information that is more sensitive than the original spectrum, thus improving the accuracy of the prediction model (Wang et al., 2022). In remote sensing inversion studies, input irrelevant bands can affect the accuracy of the model and even lead to overfitting. The spectral feature band selection method can improve the prediction of the model, effectively eliminate redundant information and retain valid information. The above study on mineral content estimation of salt lakes used principal component analysis for data dimensionality reduction work. The principal component analysis, as a heuristic feature selection method, is simple to operate but cannot handle the complex relationships between input and output variables. Meta-heuristic algorithms avoid these limitations. The genetic algorithm (GA) is a meta-heuristic intelligent algorithm based on natural selection and genetics (Katoch et al., 2021). Sun et al. (2022) used the GA for band selection and used the chosen bands and PLSR to estimate the soil organic matter content. Shekofteh and Masoudi (2019) designed an algorithm combining the GA with the artificial neural network (ANN) to select five soil properties that have the most influence on soil quality indicators. The GA has now been used to solve optimization problems in many fields, including feature selection (Cao et al., 2018). So far, the GA has not been applied to the study of the estimation of minerals in salt lakes. Based on this, our innovation is to use the GA for feature selection and machine learning models to estimate the mineral content.

Current studies on mineral content estimation of salt lakes on the TP have focused on Zabuye Salt Lake (Tian et al., 2005; ZHANG et al., 2007; Xu et al., 2017; Liu et al., 2021). It is necessary to investigate and survey other salt lakes on the TP using remote sensing technology and machine learning algorithms. This study investigates the feasibility of a strategy combining feature selection based on the GA and machine learning in estimating the content of multiple minerals in salt lakes, and provides an application example and theoretical support for future assessment and monitoring of mineral resources in salt lakes on the TP. In this work, we proposed an intelligent method for estimating the Li, B, and TDS contents of salt lakes on the TP using Sentinel-2 imagery and *in situ* data from two typical salt



**FIGURE 1**

Location of the research region. (A) Topographic map of the TP. (B) and (C) are the Sentinel-2 imagery of Bieruoze Co and Guopu Co, respectively, in May 2018.

lakes. 1) The feature bands were constructed by mathematical transformations, which can enhance the spectral information and raise the predictive model's accuracy. 2) Using the GA for feature band selection, the best input band combination was intelligently selected. 3) PLSR, MLR, and RF models were constructed using the optimal band combination, and the accuracy of the three models was contrasted. 4) The best estimation model was used to monitor and map the mineral content of the salt lakes.

## 2 Study area and data

### 2.1 Study area

The research area is made up of two lakes, Bieruoze Co and Guopu Co, and is situated in the southeastern Ali region of the TP (Figure 1). Based on the results of the 2018 field survey, Bieruoze Co is located at 32°24'–32°28'N, 82°52'–82°59'E, with an altitude of 4,400 m. The lake is approximately 9.6 km long from east to the west and 4.8 km wide, with a surface area of 36 km<sup>2</sup> and an average water depth of 3.3 m. Guopu Co is located at 31°49'–31°55'N, 83°7'–83°15'E, with an altitude of 4,700 m. The east–west length of the lake is approximately 14 km, it is approximately 5.9 km at its widest point, its surface area is approximately 61 km<sup>2</sup>, and its average water depth is 2.8 m.

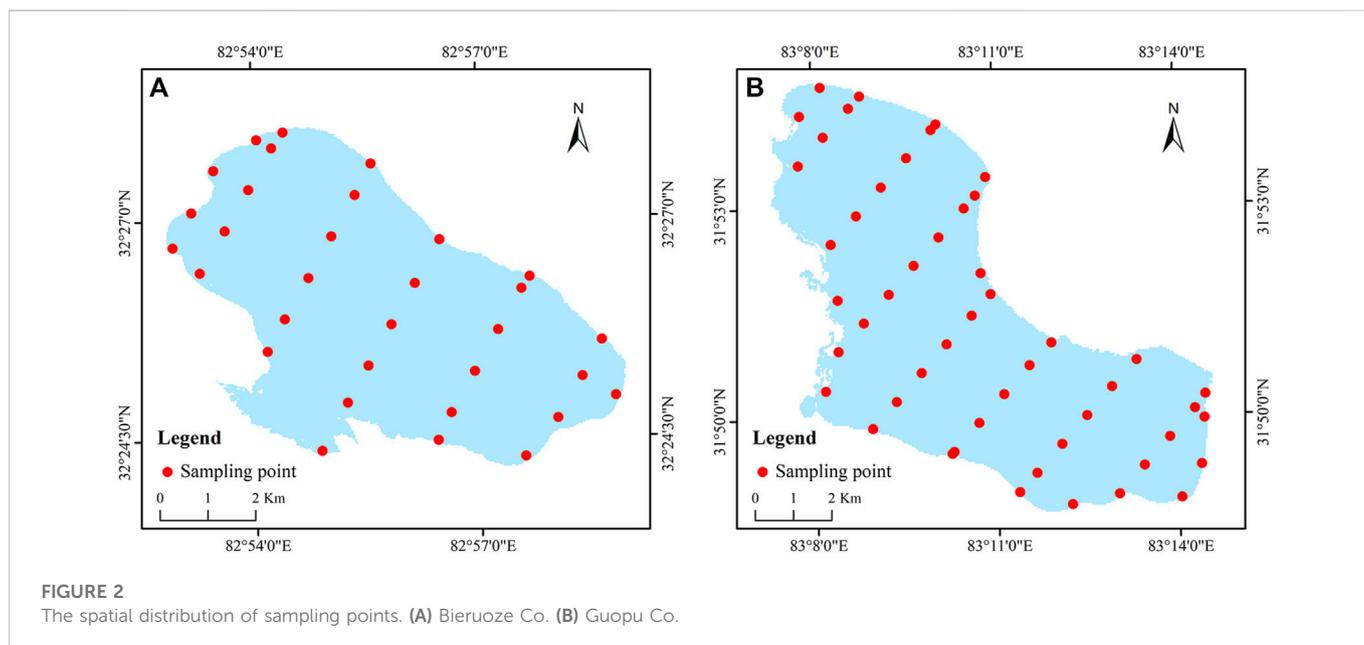
Bieruoze Co and Guopu Co are both salt lakes, blue in color, and are non-discharge lakes without outlets. The lake water is colorless, odorless, salty, and transparent. The lake contains a large number of

brine worms, and the soluble mineral salts in the area around the lake constantly converge in the lake through surface runoff. The dynamic changes in the lake are barely affected by human activities, mainly relying on atmospheric precipitation, snow and ice melt and spring recharge; the discharge relies on strong evaporation.

### 2.2 *In situ* data

In May–June 2018, field measurements of the Li, B, and TDS contents of Bieruoze Co and Guopu Co were conducted. According to the “salt lake and salt mineral geological survey specification”, the surface area of Bieruoze Co was between 10 and 50 km<sup>2</sup>, the observation network degree was 2 km, and the point distance was 1–2 km; the surface area of Guopu Co was between 50 and 100 km<sup>2</sup>, the observation network degree was 2–4 km, and the point distance was 2 km. To obtain more sample data, the observation network and sampling point spacing were established at 2 km × 1 km intervals, and the local sampling point spacing was set to 0.5 km. Figure 2 depicts the locations of the 32 sampling points in Bieruoze Co and the 50 sampling points in Guopu Co. Water samples were collected at a depth of 0.2 m from the lake surface using a 0.55 L polyethylene water bottle, and all samples were forwarded to the laboratory for analysis.

Table 1 displays the maximum, minimum, mean, standard deviation (SD) and coefficient of variation (CV) of the measured values at all sampling points. The measured data show that the two salt



**FIGURE 2**  
The spatial distribution of sampling points. (A) Bieruoze Co. (B) Guopu Co.

**TABLE 1** Statistical information on the measured values of all the sampling points of Bieruoze Co and Guopu Co.

Lake name	Measured plots	Parameters	Max value	Min value	Mean value	SD	CV
Bieruoze Co	32	Li (mg·L <sup>-1</sup> )	44.35	36.84	41.71	1.79	0.04
		B (mg·L <sup>-1</sup> )	232.48	194.23	222.66	7.70	0.03
		TDS (g·L <sup>-1</sup> )	23.43	19.61	22.21	0.92	0.04
Guopu Co	50	Li (mg·L <sup>-1</sup> )	10.06	8.18	9.37	0.37	0.04
		B (mg·L <sup>-1</sup> )	358.46	302.95	338.52	11.84	0.04
		TDS (g·L <sup>-1</sup> )	28.14	25.20	26.90	0.73	0.03

lakes, Bieruoze Co and Guopu Co, are rich in Li, B and TDS resources. The variation in Li content in Bieruoze Co Lake ranged from 36.84 to 44.35 mg L<sup>-1</sup>, with an average content of 41.71 mg L<sup>-1</sup>. The concentration of B was greater than that of Li, with the variation in B content ranging from 194.23 to 232.48 mg L<sup>-1</sup>, with an average content of 222.66 mg L<sup>-1</sup>. The variation in TDS content ranged from 19.61 to 23.43 g L<sup>-1</sup>, with an average content of 22.21 g L<sup>-1</sup>. The Li content of Guopu Co was lower than that of Bieruoze Co, and the B and TDS concentrations were higher than those of Bieruoze Co. The variation in Li content was small, ranging from 8.18 to 10.06 mg L<sup>-1</sup>, with an average content of 9.37 mg L<sup>-1</sup>. The variation in B content ranged from 302.95 to 358.46 mg L<sup>-1</sup>, with an average of 338.52 mg L<sup>-1</sup>. The variation in TDS content ranged from 25.20 to 28.14 g L<sup>-1</sup>, with an average of 26.90 g L<sup>-1</sup>.

We used 82 sampling points data from two salt lakes to construct the model, 70% were randomly selected as the training dataset, and 30% were selected as the testing dataset. The number of samples in the training dataset was 57 (including 22 from Bieruoze Co and 35 from Guopu Co); the number of samples in the test dataset was 25 (including 10 from Bieruoze Co and 15 from Guopu Co). To ensure comparability among the tested models, the training and testing sets of each model were identical.

## 2.3 Satellite data and preprocessing

Sentinel-2 is a high-resolution multispectral imaging satellite with an orbital altitude of 786 km. Sentinel-2 carries a multispectral instrument covering 13 bands from visible to shortwave infrared with a maximum spatial resolution of 10 m. The revisit period is 10 days for a single satellite and 5 days for two complementary satellites (Sentinel-2A/B). Sentinel-2 data were obtained from the Copernicus Data Centre of the European Space Agency (<https://scihub.copernicus.eu/>). To obtain more accurate experimental results, the satellite data time needs to be as close as possible to the *in situ* measurement time, and there should be no clouds covering the lake, so we selected Sentinel-2 satellite image data from May 2018. We used the Sen2Cor processing tool for atmospheric correction to obtain the Level-2A products, resampled all spectral band images to a 10 m resolution, and finally extracted the lake extent using the normalized difference water body index (NDWI) (Gao, 1996; Liu et al., 2019).

$$\text{NDWI} = (\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR}) \quad (1)$$

where Green denotes the B<sub>3</sub> band of Sentinel-2 and NIR denotes the B<sub>8</sub> band of Sentinel-2.

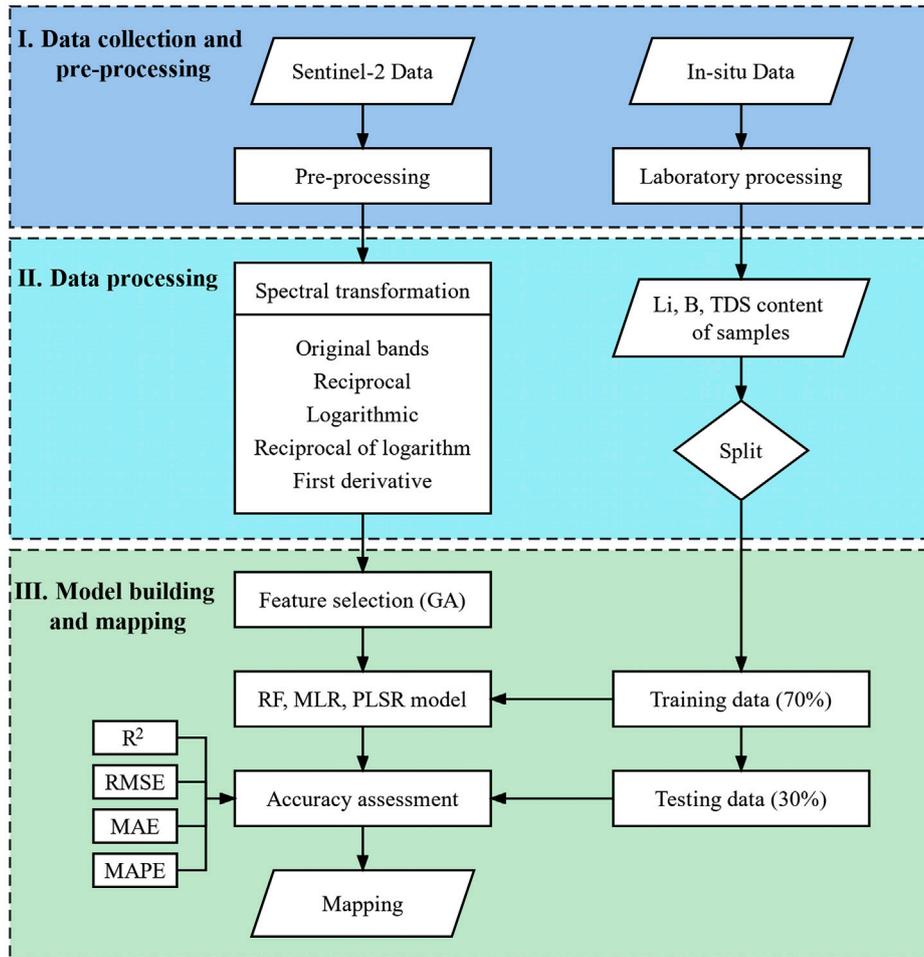


FIGURE 3  
Flow chart.

### 3 Methods

The methodology used in this study is shown in Figure 3. In the first step, we collected *in situ* data from two salt lakes on the TP, including Li, B, and TDS contents. Sentinel-2 data of the corresponding periods were obtained and pre-processed. In the second step, the spectral bands of Sentinel-2 were processed by mathematical transformations, including reciprocal transformation (RT), logarithmic transformation (LT), reciprocal of logarithm (RL), and first-order derivative (FD). The sample data were randomly divided into the training set and testing set. In the third step, firstly, the spectral bands obtained after the mathematical transformation in the second step and the original bands were put into the GA for spectral feature selection. The GA adaptively selected the feature bands according to the fitness function. Subsequently, the feature bands obtained by GA screening were used as input variables of RF, MLR, and PLSR models to construct estimation models. Finally, the performance of the three estimation models was evaluated using the evaluation function, and the optimal model was used for the mineral content estimation of the two salt lakes.

#### 3.1 Spectral feature transformation

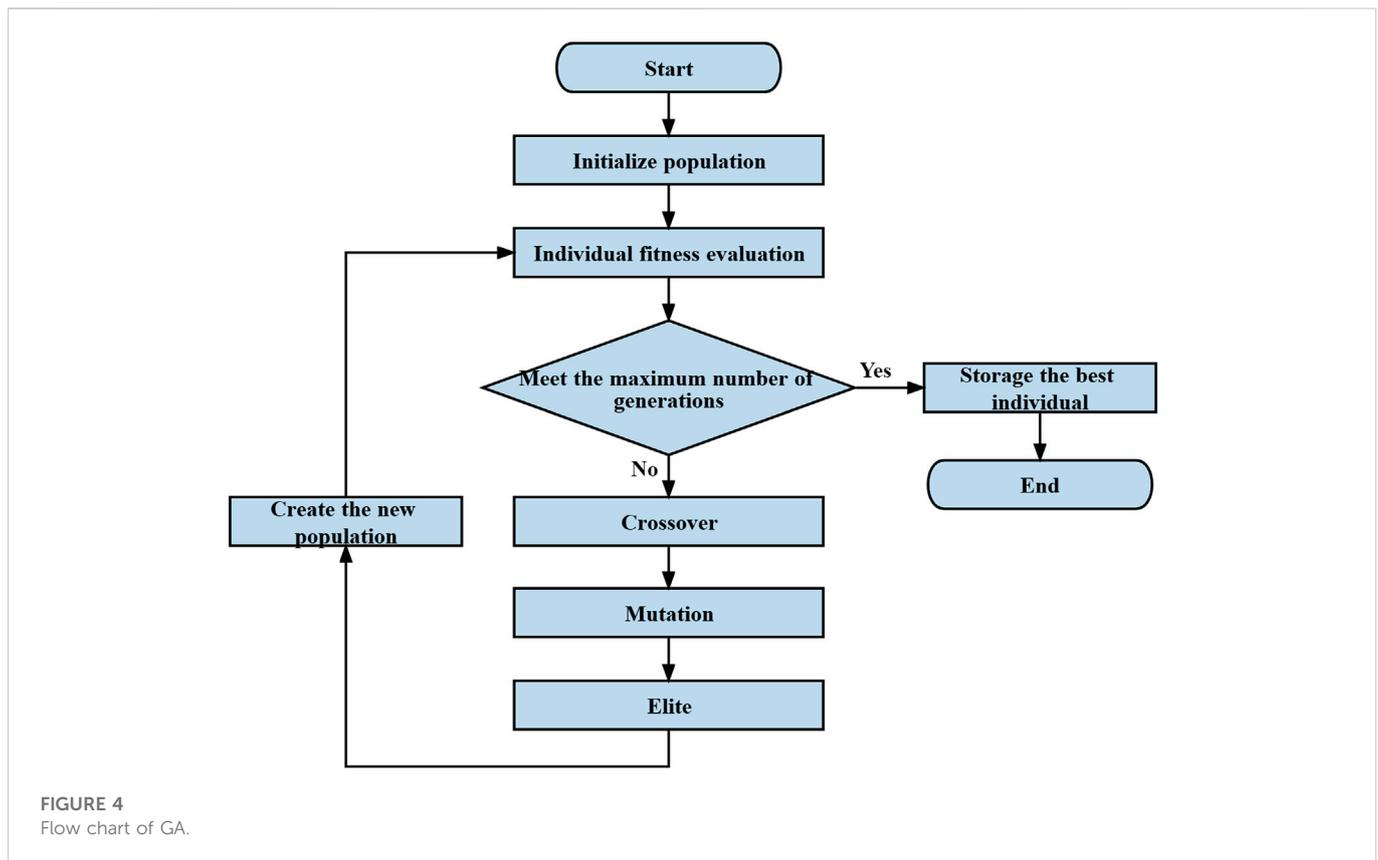
First-order derivative transformation is a common preprocessing method for hyperspectral data that is capable of extracting more delicate spectral data than the original spectrum and has been widely used for water quality parameter estimation (Wang et al., 2022) and soil parameter estimation (Wang et al., 2021). The introduction of spectral derivative transformation into multispectral images can further exploit the differences between spectral data to retrieve valuable information and thus improve the accuracy of prediction models (Wang et al., 2021). The spectral mathematical transformation allows the extraction of hidden features of water body spectra and the effective use of differences in spectral data to estimate different water body parameters.

We used the reciprocal transformation (RT), logarithmic transformation (LT), reciprocal of logarithm (RL), and first-order derivative (FD) for the Sentinel-2 spectrum.

$$RT(B_i) = 1 / (B_i) \tag{2}$$

$$LT(B_i) = \ln(B_i) \tag{3}$$

$$RL(B_i) = 1 / \ln(B_i) \tag{4}$$



$$FD(B_i) = \frac{B_{(i+1)} - B_i}{\lambda_{i+1} - \lambda_i} \quad (5)$$

where  $B_i$  denotes a single band of Sentinel-2 and  $\lambda_i$  denotes the central wavelength. In this study, the above four transformations and the untreated original spectral bands (OR) are used as input variables for the estimation model.

## 3.2 Feature optimization using GA

The number of input variables for machine learning models can have a considerable impact on how accurate the prediction model is; extraneous variables complicate the structure of the prediction model and increase the number of calibration parameters, which confounds training (Bayati and Danesh-Yazdi, 2021). Additionally, keeping useless bands raises the computational cost (Han et al., 2022). To avoid these problems, we used the GA to determine the best combination of bands for predicting mineral contents.

The GA is a metaheuristic global optimization algorithm that has been used to solve spectral subset selection problems. The GA was designed and proposed based on evolutionary laws of organisms in nature (Abba et al., 2022), and it is founded on the ideas of natural selection and evolution, applying the concepts of superiority and inferiority to find optimal solutions to optimization problems (Li et al., 2015). As a non-deterministic method of choosing variables (Tiyasha et al., 2021), the GA finds the best combination of bands by the following steps.

1. Initialize the population: Spectral variables are encoded as binary data, with 0 and 1 as individuals, and a population consists of multiple individuals.

2. Selection: According to the fitness function, suitable parents are chosen, with individuals with greater fitness levels having a larger chance of being chosen.
3. Crossover: Genetic exchange between individuals of the parental generation produces two new offspring individuals.
4. Mutation: Random variation in the individual for a particular gene value.

The above process simulates the stages of natural evolution, leading to the creation of generations that are more suited. The flow chart of the GA is demonstrated in Figure 4.

## 3.3 Model construction

### 3.3.1 RF model

RF is a machine learning technique based on decision trees for classification and regression. The basic unit of the RF is a decision tree, and several decision trees are created during the training process. Each decision tree creates a number of weak classifiers for local learning using randomly chosen samples and characteristics and then combines them to create a powerful global classifier. The final output is the mean of the predicted values for each decision tree (Wang et al., 2022). RF reflects the non-linear regression relationship between water parameters and spectral data without the need to explicitly know their functional correlation. RF has the advantage of a strong ability to handle multidimensional data and avoid overfitting.

The RF model is run in Python 3.7, and the hyperparameters are determined by a “grid search” strategy.

### 3.3.2 PLSR model

A multivariate regression technique called PLSR combines the benefits of principal component analysis with those of standard correlation analysis (Wold et al., 2001; Cao et al., 2018). PLSR compresses the input data matrix by choosing consecutive orthogonal elements to maximize the covariance between Y (water body parameters) and X (spectral bands) (Zhu et al., 2022). It successfully addresses the issue of multicollinearity between spectral data and maintains accurate predictions even with few samples (Xie et al., 2022).

### 3.3.3 MLR model

MLR is the most widely used linear regression method (Hestir et al., 2015). MLR attempts to fit the relationship between multiple independent variables and dependent variables through a linear equation (Abba et al., 2017). Although MLR has lower predictive accuracy than machine learning-based models, it can be easily interpreted (Nemati et al., 2015).

## 3.4 Accuracy evaluation

Four evaluation metrics are used to assess the estimation effectiveness of the models, namely, the coefficient of determination ( $R^2$ ), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).  $R^2$  is a metric for how well a model matches the data and illustrates the model's capacity for prediction, so the best model can be selected based on  $R^2$ . Usually, an  $R^2$  value closer to one indicates a more robust model. RMSE is used to evaluate the deviation between the estimated and true values, and MAE and MAPE metrics are used to measure the closeness of the estimated values to the true data. The smaller the RMSE, MAE, and MAPE are, the better the predictive performance of the model (Wagle et al., 2019). The following is the calculation for these metrics:

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{M}_i)^2}{\sum_{i=1}^n (M_i - \bar{M}_i)^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - P_i)^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(M_i - P_i)| \quad (8)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{M_i - P_i}{M_i} \right| \quad (9)$$

where  $M_i$  is the *in situ* data,  $P_i$  is the predicted value,  $\bar{M}$  is the average of the *in situ* data.

## 4 Results

### 4.1 Mineral content prediction model results

#### 4.1.1 Modeling results without feature selection

Before constructing the prediction model using feature selection, we used the OR and all four transformed spectral bands as the input bands and the Li, B, and TDS contents as the dependent variables to build the PLSR, MLR, and RF prediction models. Table 2 lists the modeling results

for all bands. The PLSR model estimated  $R^2$  and MAPE values of 0.81% and 38.93% for Li, 0.82% and 6.29% for B, and 0.70% and 4.11% for TDS, respectively, for the testing set. The RF model estimated  $R^2$  and MAPE values of 0.91% and 13.33% for Li, 0.88% and 3.82% for B, and 0.87% and 2.88% for TDS, respectively, for the testing set. By comparing these two models, the estimation performance of the RF model was higher than that of the PLSR model. In addition, the MLR model performed poorly in estimating the three mineral contents (MAPE >30%), possibly due to too many independent variables. Using all the bands to build the estimation model does not achieve satisfactory results. In addition, redundant bands increase the cost of model training and reduce the model accuracy. Therefore, this study used the strategy of spectral feature band selection to solve these problems and improve the model prediction accuracy.

#### 4.1.2 Results of the prediction model based on GA feature optimization

The GA-PLSR, GA-MLR, and GA-RF models were constructed based on the GA for feature selection. The number of generations of the GA was 500, and the number of populations was 50. The model accuracy evaluation is shown in Table 3. In all the models, the accuracy of the training was higher than that of the testing, which indicated that all the models were not overfitted.

From the modeling results, the performance of all three models built based on the GA for feature band selection was better than that of the models constructed based on all bands. The model performance for estimating three mineral contents was the GA-RF > GA-MLR > GA-PLSR. The plots of the predicted *versus* measured contents of the models obtained using the GA-PLSR, GA-MLR, and GA-RF are shown in Figure 5, Figure 6, Figure 7. Table 4 showed the 17, 20, and 17 feature bands selected by GA for use in the GA-RF model. The GA-RF model estimated Li ( $R^2=0.99$ , RMSE=1.15 mg L<sup>-1</sup>, MAE=0.78 mg L<sup>-1</sup>, MAPE=3.00%); B ( $R^2=0.97$ , RMSE=10.65 mg L<sup>-1</sup>, MAE=7.91 mg L<sup>-1</sup>, MAPE=2.73%); TDS ( $R^2=0.93$ , RMSE=0.60 g L<sup>-1</sup>, MAE=0.46 g L<sup>-1</sup>, MAPE=1.82%) all achieved the best performance.

Figure 8 displays the spatial distribution of the MAPE for the three mineral contents that were estimated using the GA-RF model. Overall, the sampling point errors for model training were smaller than those for model testing. The MAPE values of the three mineral contents estimated in Bieruoze Co ranged from 0.04% to 8.38% (Li), 0.01%–12.97% (B), and 0.06%–6.22% (TDS); the average MAPEs were 2.29% (Li), 1.67% (B), and 1.58% (TDS). The MAPE for the three mineral contents estimated at Guopu Co ranged from 0.02% to 8.76% (Li), 0.04%–5.56% (B), and 0.08%–5.65% (TDS); the mean MAPE were 1.74% (Li), 1.62% (B), and 1.31% (TDS).

### 4.2 Mapping of the spatial distribution of mineral content

We mapped the geographic distribution of Li, B, and TDS contents in Bieruoze Co and Guopu Co using the GA-RF model and Sentinel-2 imagery (Figure 9). The variation in the three mineral contents in Bieruoze Co ranged from 33.93 to 43.46 mg L<sup>-1</sup> (Li), 201.43–250.37 mg L<sup>-1</sup> (B), and 20.29–23.78 g L<sup>-1</sup> (TDS). The distribution of Li content in the northwest was high, and that in the southeast was low; the spatial distribution of B content was relatively uniform; and the TDS content in the west was slightly higher than that in the east. The variation in the three minerals in

TABLE 2 Results of estimation models based on all spectral bands.

		Training				Testing			
		$R^2$	RMSE	MAE	MAPE	$R^2$	RMSE	MAE	MAPE
Li	MLR	0.67	9.13	2.89	22.63	0.66	9.16	6.95	47.35
	PLSR	0.89	5.27	4.42	35.60	0.81	6.90	4.98	38.93
	RF	0.99	1.15	0.54	3.50	0.91	4.81	1.97	13.33
B	MLR	0.38	77.68	33.27	11.50	0.19	142.67	108.37	37.45
	PLSR	0.87	20.27	15.77	5.33	0.82	24.79	17.59	6.29
	RF	0.97	9.51	6.92	2.40	0.88	20.29	11.61	3.82
TDS	MLR	0.28	13.10	5.76	22.63	0.16	13.34	10.97	43.71
	PLSR	0.77	1.18	0.92	3.77	0.70	1.25	1.05	4.11
	RF	0.92	0.71	0.44	1.79	0.87	0.84	0.71	2.88

TABLE 3 Performance of the estimation model based on the feature selection method.

		Training				Testing			
		$R^2$	RMSE	MAE	MAPE	$R^2$	RMSE	MAE	MAPE
Li	GA-PLSR	0.91	4.75	3.78	28.81	0.87	5.66	3.94	28.85
	GA-MLR	0.98	2.15	1.52	9.64	0.93	4.15	2.30	13.75
	GA-RF	0.99	0.60	0.32	1.50	0.99	1.15	0.78	3.00
B	GA-PLSR	0.89	18.79	14.67	5.03	0.85	22.69	14.87	5.57
	GA-MLR	0.96	11.27	8.29	2.88	0.91	18.00	11.39	4.00
	GA-RF	0.99	4.30	3.40	1.16	0.97	10.65	7.91	2.73
TDS	GA-PLSR	0.80	1.11	0.84	3.48	0.77	1.10	0.88	3.48
	GA-MLR	0.89	0.81	0.64	2.65	0.86	0.87	0.67	2.70
	GA-RF	0.97	0.43	0.30	1.24	0.93	0.60	0.46	1.82

Guopu Co ranged from 8.51 to 9.93 mg L<sup>-1</sup> (Li), 305.01–353.23 mg L<sup>-1</sup> (B), and 24.31–27.61 g L<sup>-1</sup> (TDS); the three minerals were high in the north and east of Guopu Co.

## 5 Discussion

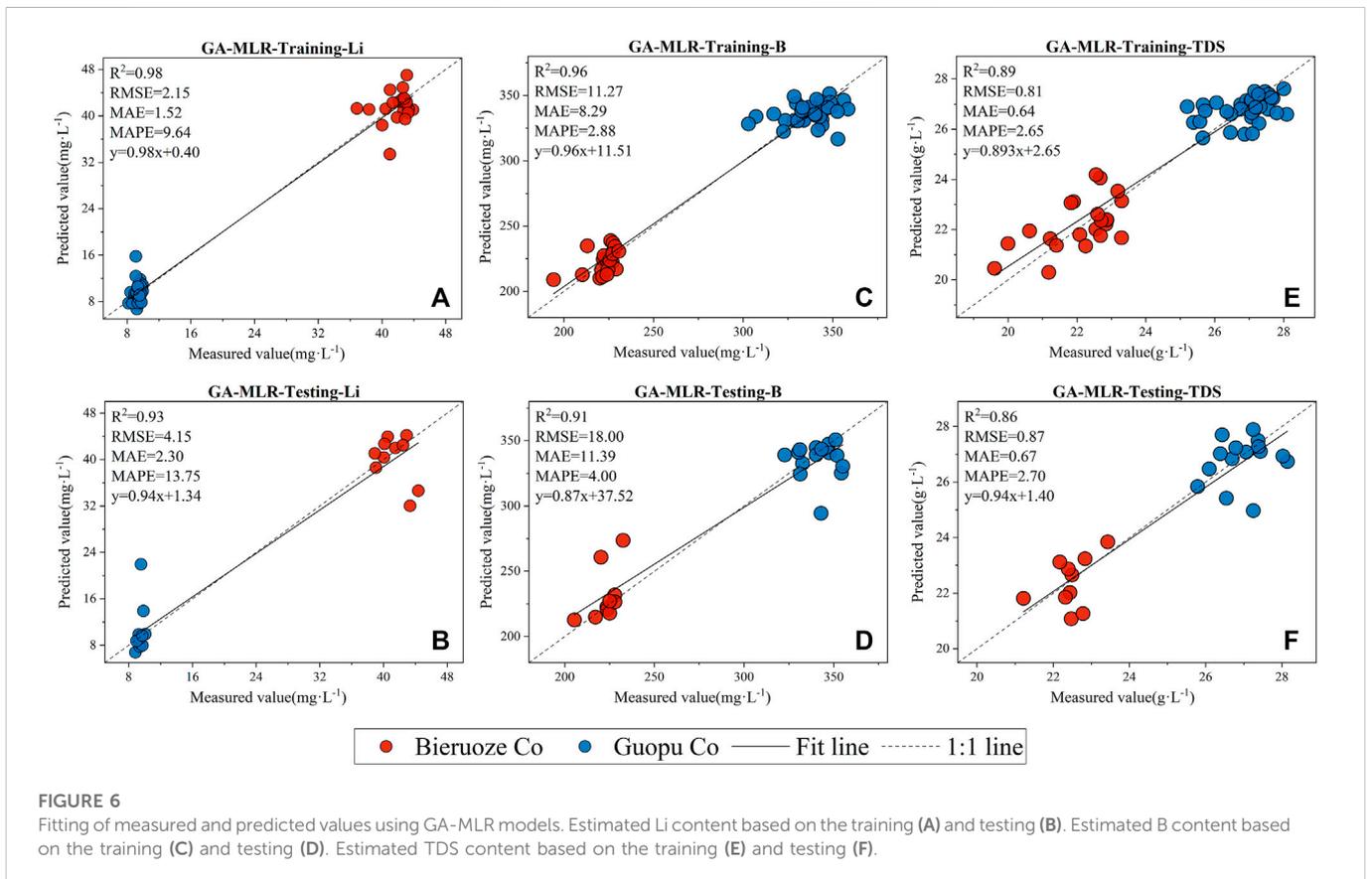
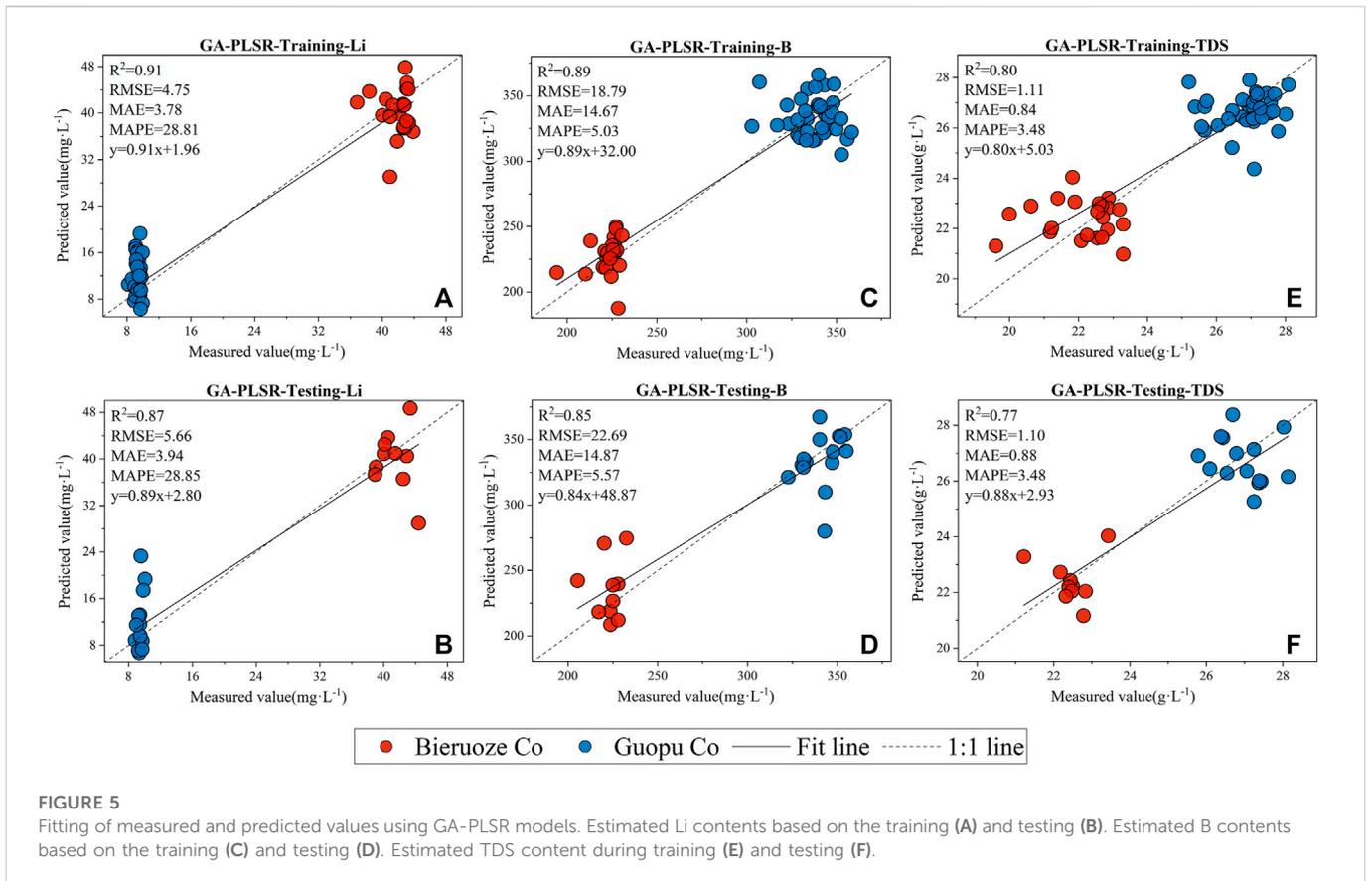
### 5.1 Advantages of spectral feature transformation

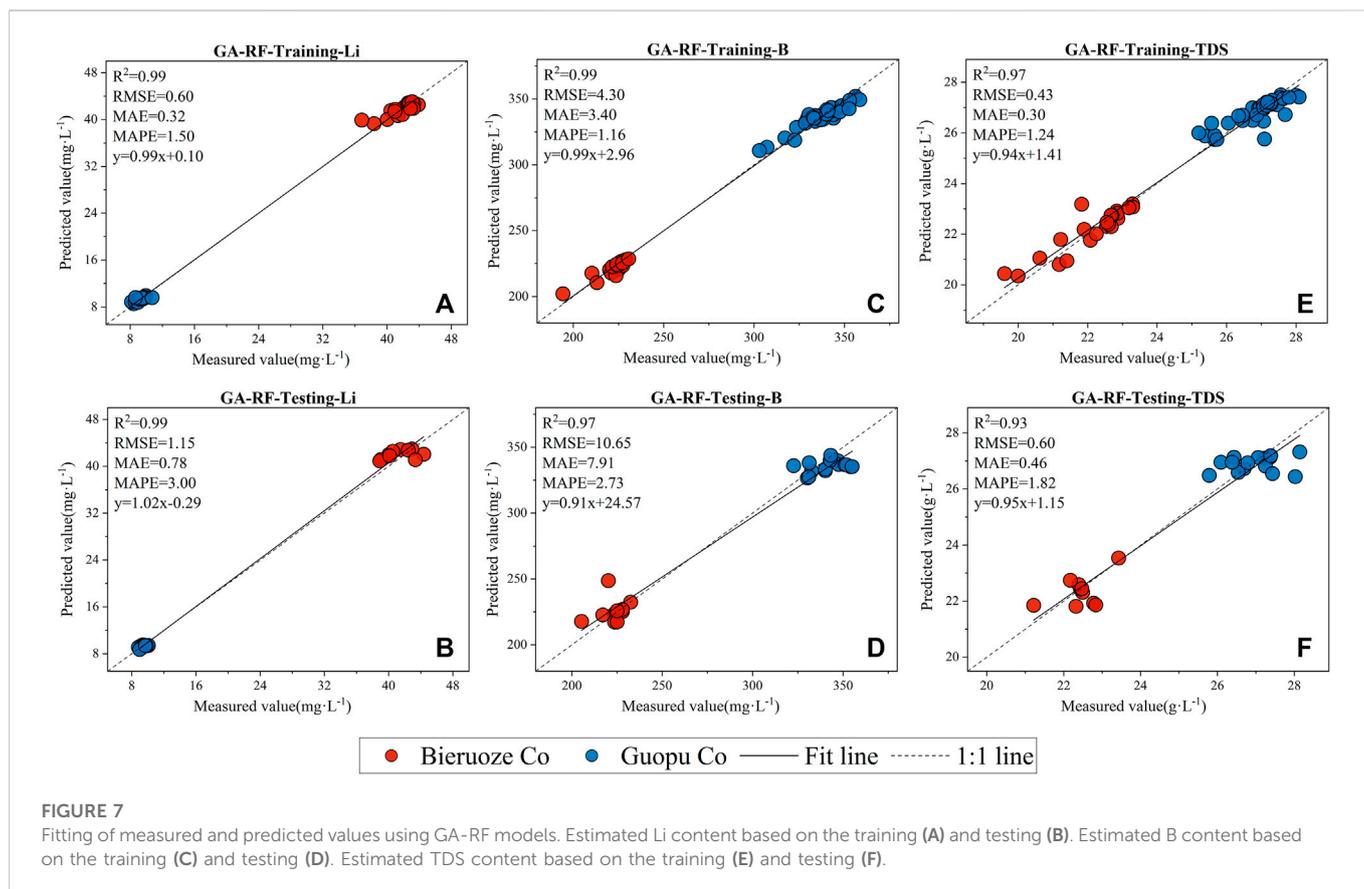
In this study, spectral transformation had a significant part in estimating the mineral content of salt lakes on the TP. Li, B, and TDS, as non-optically active substances, have more complex optical properties (Alparslan et al., 2009). We mathematically transformed the Sentinel-2 spectral band using four methods: RT, LT, RL, and FD. The correlation coefficient plots were obtained by analyzing the Pearson correlations of Li, B, and TDS contents and different spectral variables (Figure 10). For Li, the correlation coefficients of the original reflectance (OR) and LT spectra with Li content were negative, the correlation coefficients of the RL spectra with Li content were positive, and the correlation coefficients of RT and FD with Li content were alternately

positive or negative. For B, the correlation coefficients of the OR and LT bands with B were all positive, the correlation coefficients of RL with B content were all negative, and the correlation coefficients of RT and FD with Li content were alternately positive or negative. The trends of the correlation coefficients of the spectra with TDS content and B content were basically the same. The correlations of the RT-, LT-, RL-, and FD-treated spectra with Li, B, and TDS were all improved to different degrees compared to the OR, and the bands with the highest correlations were all RTB<sub>1</sub> (Li = 0.93, B = -0.93, TDS = -0.84). RT showed the highest correlations, followed by LT and OR. The results suggested that the mathematical transformation of the spectra can significantly reduce the negative effects in the spectra, enhance the small fluctuations in the reflectance spectral features, enrich the spectral details of images, highlight spectral features, and provide effective information for prediction models.

### 5.2 Advantages of the GA-RF model

The GA-RF model has two advantages, the first one is the advantage of using the GA for feature optimization extraction, and the second one is the advantage of the RF model.



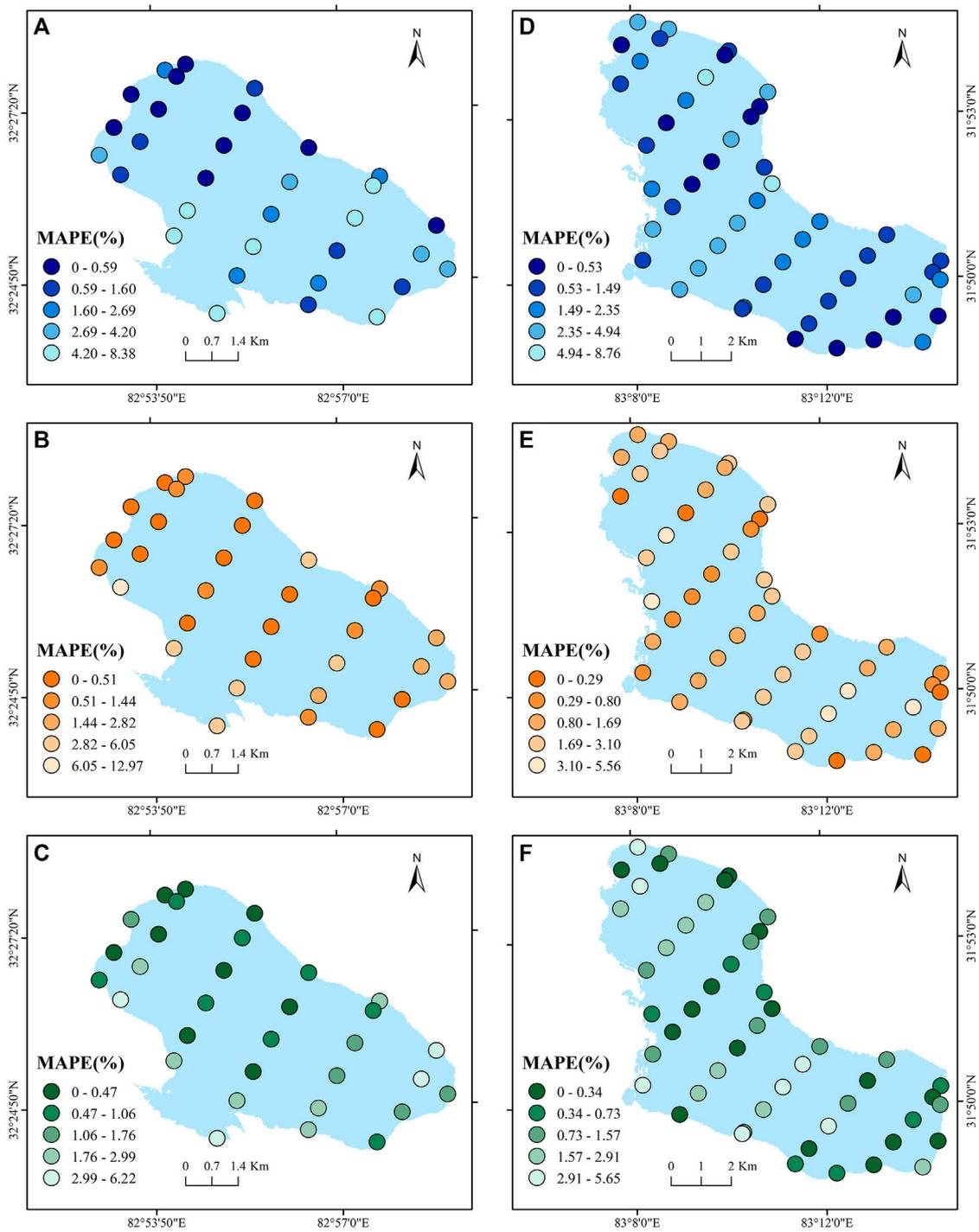


**TABLE 4 Spectral feature bands of GA-RF model obtained by GA.**

Model	Parameters	Spectral bands	Number of bands
GA-RF	Li	B <sub>1</sub> , B <sub>4</sub> , B <sub>5</sub> , RTB <sub>4</sub> , RTB <sub>5</sub> , RTB <sub>7</sub> , LTB <sub>4</sub> , RLB <sub>4</sub> , RLB <sub>5</sub> , RLB <sub>6</sub> , RLB <sub>8a</sub> , FDB <sub>2</sub> , FDB <sub>4</sub> , FDB <sub>5</sub> , FDB <sub>8</sub> , FDB <sub>8a</sub> , FDB <sub>9</sub>	17
	B	B <sub>6</sub> , B <sub>7</sub> , B <sub>8</sub> , RTB <sub>8</sub> , LTB <sub>1</sub> , LTB <sub>7</sub> , LTB <sub>8</sub> , LTB <sub>8a</sub> , LTB <sub>9</sub> , LTB <sub>12</sub> , RLB <sub>3</sub> , RLB <sub>8</sub> , FDB <sub>1</sub> , FDB <sub>2</sub> , FDB <sub>6</sub> , FDB <sub>7</sub> , FDB <sub>8</sub> , FDB <sub>8a</sub> , FDB <sub>9</sub> , FDB <sub>11</sub>	20
	TDS	B <sub>1</sub> , RTB <sub>1</sub> , RTB <sub>8a</sub> , RTB <sub>9</sub> , RTB <sub>11</sub> , RTB <sub>12</sub> , LTB <sub>1</sub> , LTB <sub>8a</sub> , LTB <sub>12</sub> , RLB <sub>2</sub> , RLB <sub>3</sub> , RLB <sub>9</sub> , FDB <sub>1</sub> , FDB <sub>2</sub> , FDB <sub>5</sub> , FDB <sub>8</sub> , FDB <sub>9</sub>	17

Table 2 displays the results of modeling based on the full spectral bands; however, the accuracy of all models was not satisfactory. This was due to the presence of invalid information in the spectral band, which affects the estimation performance of the models. In machine learning, selecting the best input variables remains a difficult task; too many input spectral variables not only complicate the operation and increase the time cost but also reduce the prediction performance of models with limited samples, which is known as a dimensional catastrophe and leads to models suffering from the “curse of dimensionality” (Bach, 2017). In contrast, using too few input spectral bands could prevent the spectrum from fully revealing its hidden information. Therefore, it is essential to select the best input variables. Common feature selection methods include both heuristics and metaheuristics. Heuristics are simple to operate but cannot handle complex relationships between input and output variables, such as principal component analysis, while metaheuristics avoid these limitations. The GA is a classical artificial intelligence algorithm among metaheuristics (Katoch

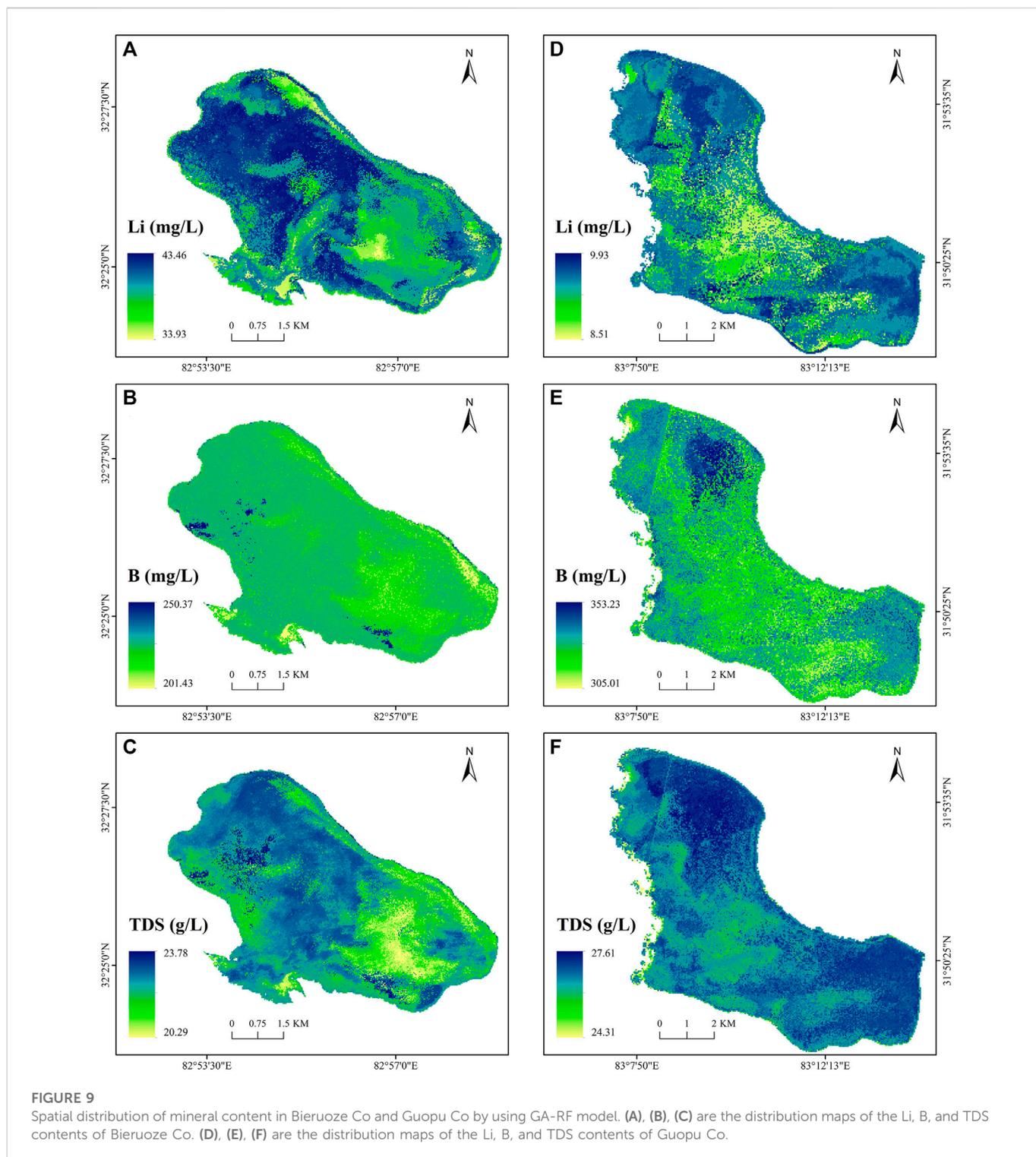
et al., 2021). In the GA, a single spectral variable is considered a gene on a chromosome, represented by a binary code: 0 means that the band is unselected, and 1 means that it is selected. Second, the genes are dynamically modified by the probability of crossover and variation to change the search process and reach the optimal solution. The GA can evaluate all individuals and output the result of optimal feature selection; therefore, the GA has better global search capability. Figure 11 exhibits the MAPE of the test set of the model constructed based on full bands and the model based on the GA for feature selection. The performance improvement of the GA-MLR and GA-RF models is large. Compared with the MLR model, the MAPE of the GA-MLR model on the testing set was reduced by 70.96% (Li), 89.32% (B), and 93.83% (TDS). Compared with the PLSR model, the MAPE of the GA-PLSR model on the testing set was reduced by 25.90% (Li), 11.47% (B), and 15.39% (TDS). Compared with the RF model, the MAPE of the GA-RF model on the testing set was reduced by 77.52% (Li), 28.54% (B) and 36.79% (TDS). In general, the accuracy of various prediction models



**FIGURE 8** Spatial distribution map of MAPE at sampling points in Bieruozhe Co and Guopu Co by using GA-RF model. (A), (B) and (C) are the MAPE of Li, B, and TDS, respectively, estimated by Bieruozhe Co. (D), (E) and (F) are the MAPE of Li, B, and TDS, respectively, estimated by Guopu Co.

with feature optimization by the GA was improved compared with the estimation models constructed based on all spectral bands. This indicates that the GA can extract the necessary information from all bands and reduce the interference of non-essential information, thus improving the accuracy of model prediction. Therefore, the GA can be used as an effective spectral band feature selection algorithm for estimating mineral content.

Zaman Zad Ghavidel and Montaseri. (2014) estimated the TDS content in the Zarinehroud Basin, Iran, using a gene expression programming algorithm with  $R=0.96$  and  $RMSE=28.99$ . Zhou et al. (2016) used a BP neural network inversion model to invert the salt lake mineral ion content of the Tajinar Salt Lake in the Qaidam Basin. The inversion accuracy was above 85%. Wang. (2019) used Sentinel-2 data and BP and RF models to invert the Li content of Alisaro Salt Lake



with  $R^2=0.731$  and  $R^2=0.771$ , and the results showed that the RF model achieved the best results. As a machine learning model, RF can train the model with less data and a lower computational cost while possessing high accuracy and generalization performance, which is suitable for mineral content inversion of salt lakes on the TP. Figure 12 illustrates the boxplots of the three models, with the first box indicating the results of the measured values in the field and the other three boxes indicating the results of the different model estimates. The results of the GA-RF model were more similar to

the results of the field measurements, and the GA-RF model estimates were the closest to the mean, median, and range of values of the field measurements of the two lakes, as well as fewer outliers. Taylor diagram allows a visual comparison of the three different statistical indicators RMSE, SD and correlation. It is obvious from Figure 13 that the predictive performance of the GA-RF model was higher than the other models. The merits of the GA-RF model can be summarized as lower errors and higher correlations. Although the accuracy of GA-MLR, GA-PLSR models was improved by the GA for feature selection.

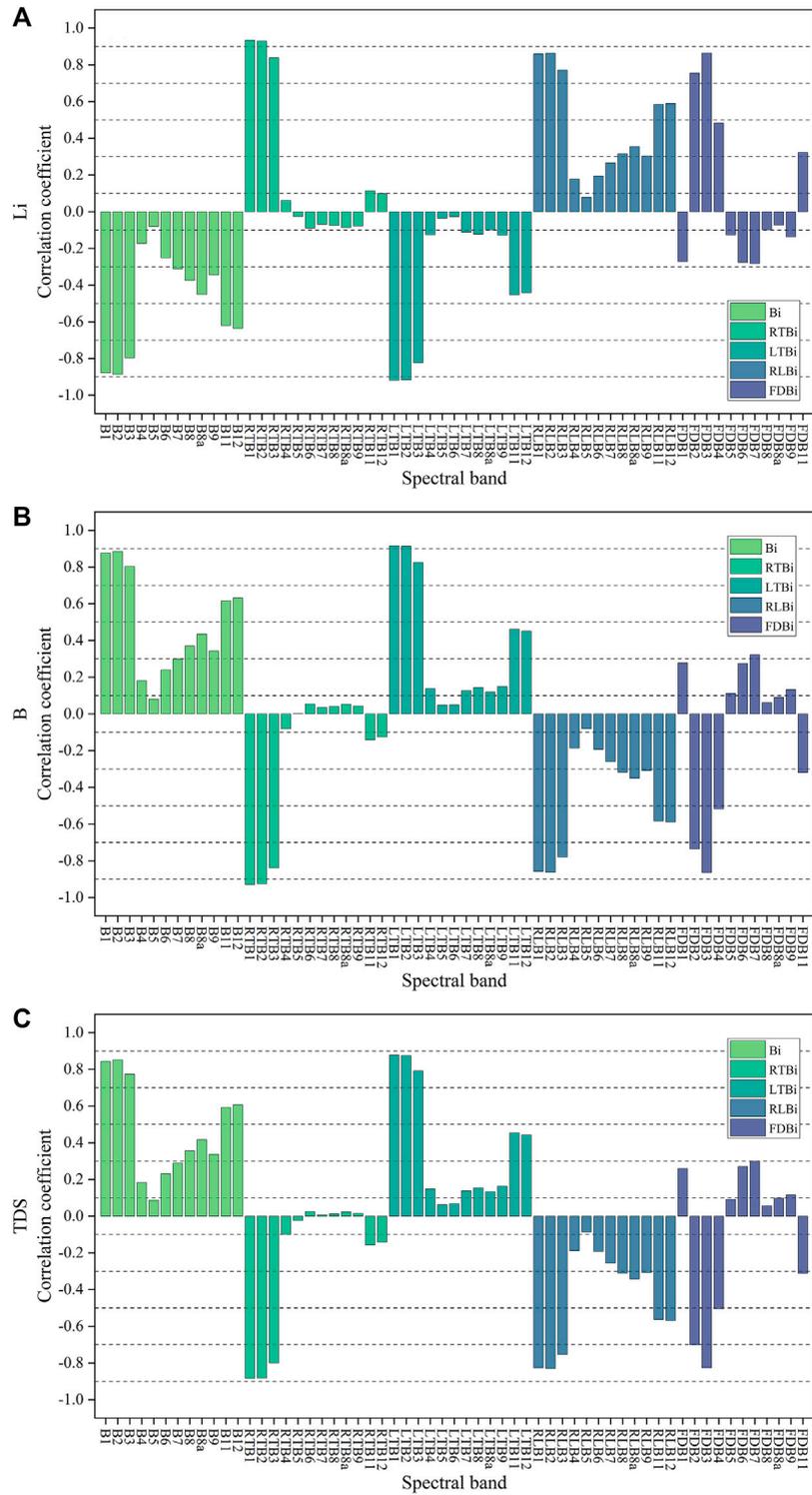
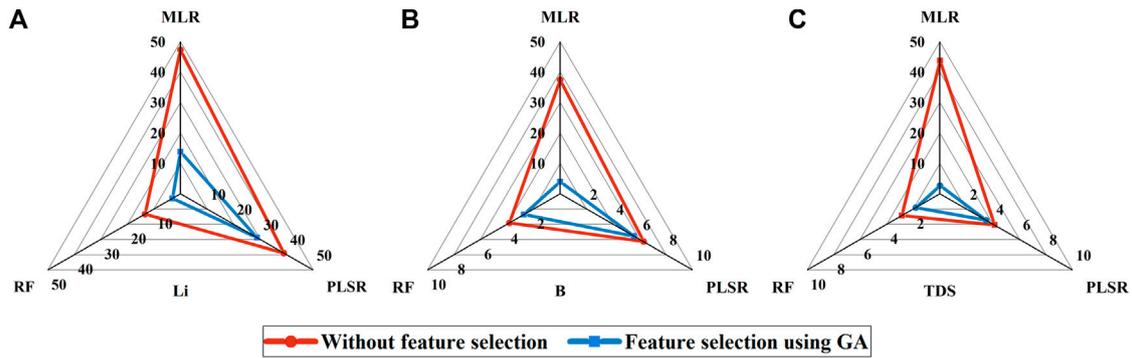


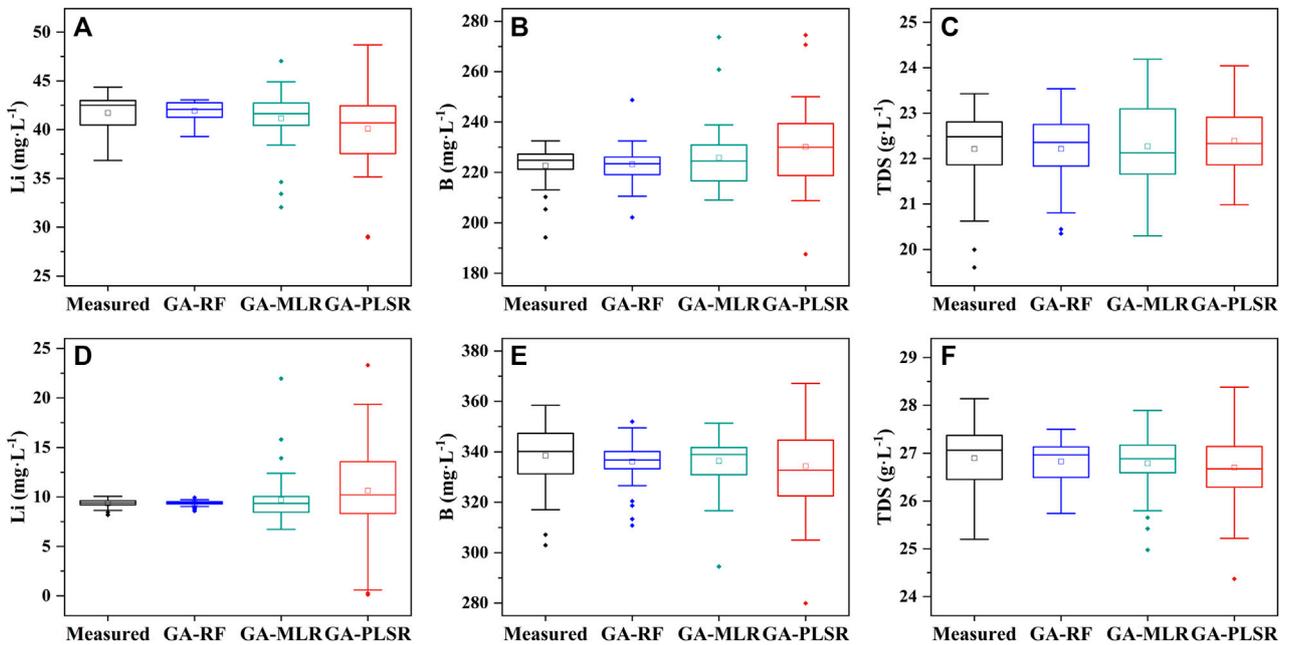
FIGURE 10 Correlation coefficients between spectral bands and Li, B, and TDS. (A) Li; (B) B; (C) TDS.

However, the performance of the two models is still lower than that of the GA-RF model because the MLR and PLSR models are weaker in dealing with non-linear complex problems. Figure 7 demonstrates that GA-RF can handle complex non-linearities between spectral variables and has a high sensitivity for predicting mineral content. In this study, four evaluation metrics were used to assess the accuracy of the model,

and the GA-RF model possessed the highest  $R^2$  and the lowest RMSE, MAE, and MAPE values. MAPE, as one of the most popularly used accuracy evaluation metrics, has the advantages of scale independence and interpretability. The GA-RF model indicated that the MAPEs were all below 3%. Therefore, the GA-RF model proposed in this study also performs well compared to previous studies.



**FIGURE 11** Radar plot of MAPE on the testing set of the three models. (A–C) are the model results for estimated Li, B, and TDS contents, respectively.



**FIGURE 12** Boxplots of field measurements and three model estimates of Li, B, and TDS contents. (A), (B) and (C) indicate the measured and model estimates of Li, B, and TDS contents of Bieruoze Co, respectively. (D), (E) and (F) indicate the measured and estimated values of Li, B, and TDS contents of Guopu Co.

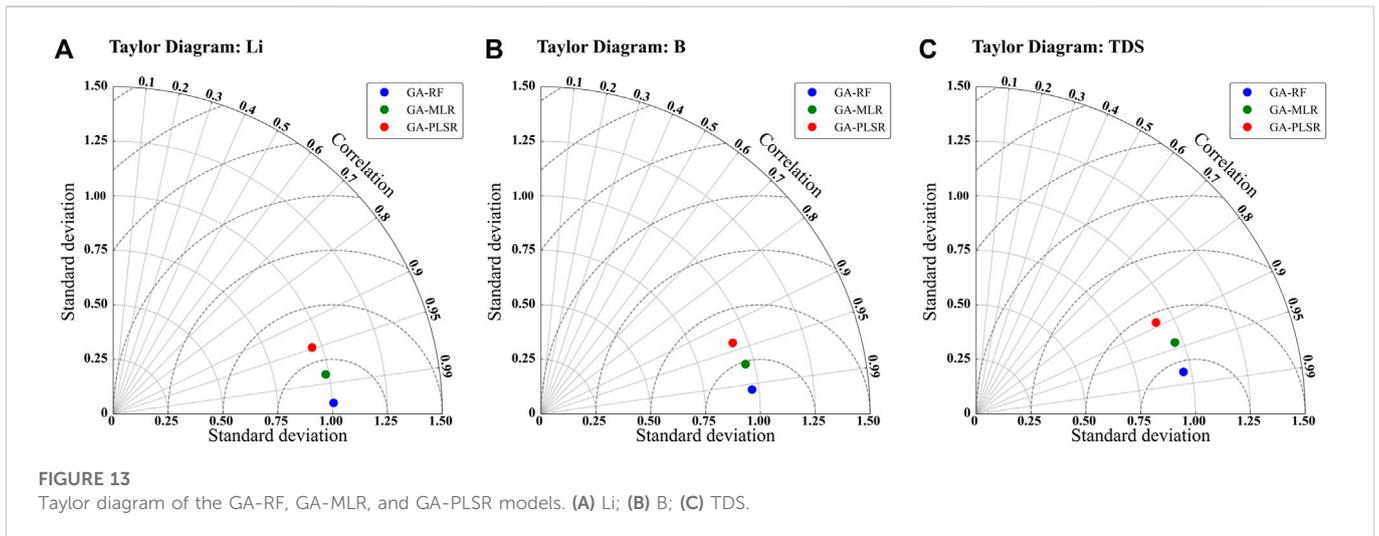
### 5.3 Importance analysis of variables for the GA-RF model

A variable importance analysis was performed on the input spectral variables of the optimal model (GA-RF) to help interpret the model results and the importance of each spectral feature in the model. The variable importance value indicates the degree of influence of that input variable on the model. Theoretically, the higher the variable importance is, the more important that input variable is to the prediction model (Wei et al., 2015). The importance of the different bands in the GA-RF is shown in Figure 13. The analysis shows that for Li, the importance of  $B_1$  is greater than 0.35 and the importance of  $FDB_2$  and  $FDB_8$  is greater than 0.15. For B,  $LTB_1$  has the highest variable importance (0.43), followed by  $FDB_8$  (0.11) and  $FDB_6$  (0.08).

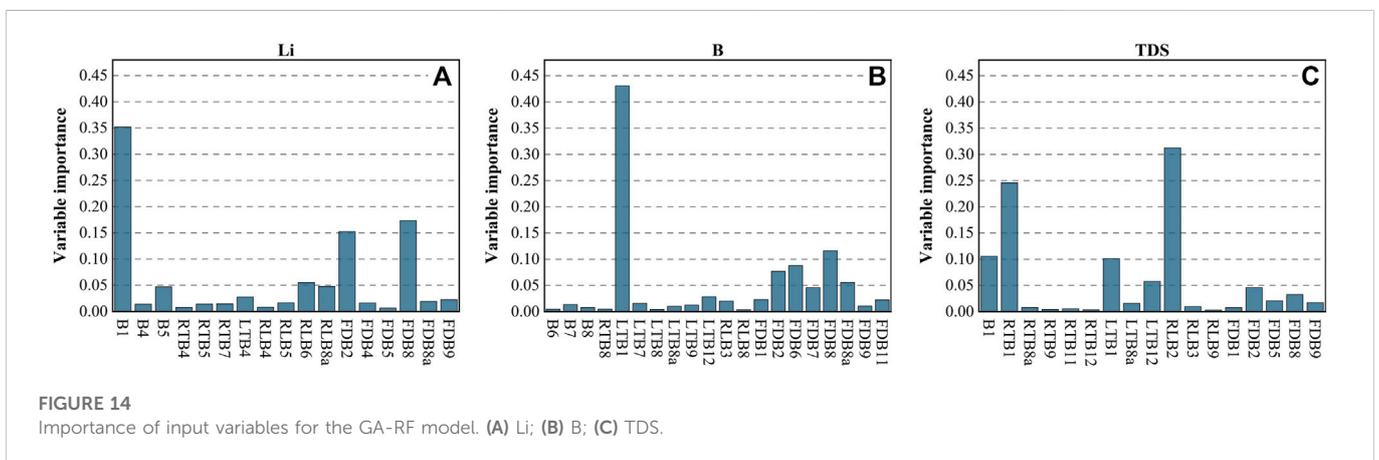
The importance of  $LTB_1$  is much higher than that of the other input bands. For TDS,  $RLB_2$  (0.31) and  $RTB_1$  (0.24) are the two bands with the highest importance. Although the importance of the other input variables is relatively low, their role in the prediction model still cannot be ignored. In addition, the bands with the highest importance are also the bands with higher correlations, so feature optimization using the GA can obtain spectral bands that are more sensitive to minerals and enhance the performance of the prediction model. Figure 14.

### 5.4 Future research

Our study showed that Sentinel-2 satellite data can estimate the brine mineral content of salt lakes on the TP with high accuracy.



**FIGURE 13**  
Taylor diagram of the GA-RF, GA-MLR, and GA-PLSR models. (A) Li; (B) B; (C) TDS.



**FIGURE 14**  
Importance of input variables for the GA-RF model. (A) Li; (B) B; (C) TDS.

According to the industrial standard in the Code for Geological Exploration of Salt Lakes and Salt Minerals (DZ/T0212-2002), the Li and B contents of Bieruoze Co and Guopu Co have reached the boundary grade, which has the potential for the development and prospecting of those minerals. The inversion method of mineral content of salt lakes on the TP based on machine learning algorithms will provide a more comprehensive, objective and accurate response to the spatial distribution of mineral content of salt lakes, and enable long-term dynamic monitoring.

Although the strategy of combining GA with RF improved the estimation accuracy, there are still some unexplained variations in information, which may be due to the influence of water body information by other environmental factors (e.g., water depth). By incorporating these environmental variables, it may be possible to enhance the estimation performance of the mineral content prediction model further. The performance of the RF model is limited by the data and the study area. If there is noise in the training data, it may cause overfitting. In addition, the generalizability of machine learning models has been a popular research topic, and it is still challenging to construct a general machine learning model. Different saline lakes on the TP vary widely, and we will collect data from different regions and periods to conduct large-scale dynamic monitoring in the future.

At the same time, the method will be tested in lakes or other aquatic systems with different salinities to check the applicability of the method as a larger-scale tool in different types of aquatic environments and at different salinity levels.

The multispectral satellites commonly used in water quality inversion are Landsat and Sentinel. Sentinel-2 satellite has 13 bands with a maximum spatial resolution of 10 m. In recent years, China has launched some multispectral satellites with similar characteristics, such as GF-1 and GF-6. The Gaofen-6 satellite is an optical remote sensing satellite in China's Gaofen series, which was launched in 2018. The GF-6 satellite combines high-resolution and wide-field-of-view (WFOV) imaging capabilities with a spatial resolution of 8 m, and has great potential for remote sensing inversion monitoring (Wang et al., 2019). In addition, feature selection algorithms have been widely used for spectral subset selection of hyperspectral satellites. In 2021, China launched the GF5-02 satellite, which carries a hyperspectral camera Advanced Hyperspectral Imager (AHSI) with 330 bands and a maximum spectral resolution of 5 nm. Compared with multispectral satellites, hyperspectral satellites have a higher spectral resolution. The emergence of new remote sensing satellites provides more possibilities and higher quality data for salt lake monitoring on the TP. In future work, we will apply the model to

new satellite data to explore the potential of new satellites for mineral content estimation and monitoring of salt lakes.

The GA-RF model proposed in this study managed to improve the accuracy of the prediction model and generated a set of modeling methods with some generalization. With the field data, the adaptive intelligent prediction model developed in this study can be used for mineral content estimation of salt lakes on the TP and prediction of other water body parameters such as total phosphorus, total nitrogen, and chlorophyll a.

## 6 Conclusion

The use of remote sensing to monitor lake parameters has proven to be a mature technique; however, there are fewer studies on brine mineral content monitoring in saline lakes. We conducted *in situ* measurements of two typical salt lakes on the TP and collected Li, B, and TDS content data. An intelligent model for estimating the mineral resource content of salt lakes on the TP was developed using Sentinel-2 high spatial resolution remote sensing data. The original spectral images were first processed by four mathematical transformations (RT, LT, RL, FD). Then, the optimal input bands were selected by feature selection through the GA. Finally, three estimation models of PLSR, MLR and RF were developed. The research conclusions were as follows.

- (1) Spectral transformations played an important role in estimating the mineral content of salt lakes on the TP. The correlation between spectral bands and Li, B, and TDS contents was increased to different degrees by spectral mathematical transformations while providing rich spectral information for the model.
- (2) The estimation model using GA for feature selection method outperforms the estimation model based on all spectral bands. Compared with the MLR model, the MAPE of the GA-MLR model on the testing set was reduced by 70.96% (Li), 89.32% B) and 93.83% (TDS), respectively. Compared with the PLSR model, the MAPE of the GA-PLSR model on the testing set was reduced by 25.90% (Li), 11.47% B), and 15.39% (TDS), respectively. Compared with the RF model, the MAPE of the GA-RF model on the testing set was reduced by 77.52% (Li), 28.54% B), and 36.79% (TDS). The genetic algorithm can be used as an effective spectral band feature selection algorithm for estimating the mineral content of salt lakes.
- (3) For all three parameters, the GA-RF model showed the best results compared with the GA-MLR and GA-PLSR models. For Li, the GA-RF model performance was  $R^2=0.99$ ,  $RMSE=1.15 \text{ mg L}^{-1}$ ,  $MAE=0.78 \text{ mg L}^{-1}$ ,  $MAPE=3.00\%$ ; for B content, the GA-RF model performance was  $R^2=0.97$ ,  $RMSE=10.65 \text{ mg L}^{-1}$ ,  $MAE=7.91 \text{ mg L}^{-1}$ ,  $MAPE=2.73\%$ ; for TDS content, the GA-RF model performance was  $R^2=0.93$ ,  $RMSE=0.60 \text{ g L}^{-1}$ ,

$MAE=0.46 \text{ g L}^{-1}$ ,  $MAPE=1.82\%$ . The GA-RF model predicted MAPE below 3% for all three mineral contents.

- (4) The combined strategy of the GA-based feature selection method and the RF showed excellent performance and applicability in mineral content prediction of salt lakes on the TP, and realized intelligent salt lake mineral search. This study provided an application example for remote sensing inversion and monitoring of mineral resources in salt lakes on the TP.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

HG: Conceptualization, Methodology, Writing–original draft, Funding acquisition. WD: Methodology, Formal analysis, Writing–Original Draft, Editing. RZ: Formal analysis, Writing–original draft, Editing. DZ: Validation, Reviewing. BQ: Validation, Reviewing. GZ: Investigation, Resources. SZ: Supervision, Funding acquisition. JS: Funding acquisition.

## Funding

This study was funded by the Major Science and Technology Special Projects in Henan Province (221100210600; 201400211000 and 201400210100); the Science and Technology Tackling Plan of Henan Province (222102320220).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Abba, S. I., Abdulkadir, R. A., Sammen, S. S., Pham, Q. B., Lawan, A. A., Esmail, P., et al. (2022). Integrating feature extraction approaches with hybrid emotional neural networks for water quality index modeling. *Appl. Soft. Comput.* 114, 108036. doi:10.1016/j.asoc.2021.108036
- Abba, S. I., Hadi, S. J., and Abdullahi, J. (2017). River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia Comput. Sci.* 120, 75–82. doi:10.1016/j.procs.2017.11.212
- Alparslan, E., Coskun, H. G., and Alganci, U. (2009). Water quality determination of küçükçekmece lake, Turkey by using multispectral satellite data. *Sci. World J.* 9, 1215–1229. doi:10.1100/tsw.2009.135
- Bach, F. (2017). Breaking the curse of dimensionality with convex neural networks. *J. Mach. Learn. Res.* 18, 1–53. doi:10.48550/arXiv.1412.8690
- Bayati, M., and Danesh-Yazdi, M. (2021). Mapping the spatiotemporal variability of salinity in the hypersaline Lake Urmia using Sentinel-2 and Landsat-8 imagery. *J. Hydrol.* 595, 126032. doi:10.1016/j.jhydrol.2021.126032

- Cao, Y., Ye, Y., Zhao, H., Jiang, Y., Wang, H., Shang, Y., et al. (2018). Remote sensing of water quality based on HJ-1A HSI imagery with modified discrete binary particle swarm optimization-partial least squares (MDBPSO-PLS) in inland waters: A case in weishan lake. *Ecol. Inf.* 44, 21–32. doi:10.1016/j.ecoinf.2018.01.004
- Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., et al. (2020). A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sens. Environ.* 248, 111974. doi:10.1016/j.rse.2020.111974
- Chen, S., and Hu, C. (2017). Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements. *Remote Sens. Environ.* 201, 115–132. doi:10.1016/j.rse.2017.09.004
- Ding, T., Zheng, M., Nie, Z., Ma, L., Ye, C., Wu, Q., et al. (2022). Impact of regional climate change on the development of lithium resources in Zabuye Salt Lake, Tibet. *Front. Earth Sci.* 10, 865158. doi:10.3389/feart.2022.865158
- Fang, Y., Chen, X., and Cheng, N. (2017). Estuary salinity prediction using a coupled GA-SVM model: A case study of the min river estuary, China. *Water Supply* 17, 52–60. doi:10.2166/ws.2016.097
- Gao, B. (1996). NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. doi:10.1016/S0034-4257(96)00067-3
- Geiger, E. F., Grossi, M. D., Trembanis, A. C., Kohut, J. T., and Oliver, M. J. (2013). Satellite-derived coastal ocean and estuarine salinity in the Mid-Atlantic. *Cont. Shelf Res.* 63, S235–S242. doi:10.1016/j.csr.2011.12.001
- Hafeez, S., Wong, M., Ho, H., Nazeer, M., Nichol, J., Abbas, S., et al. (2019). Comparison of machine learning algorithms for retrieval of water quality indicators in case-II waters: A case study of Hong Kong. *Remote Sens.* 11, 617. doi:10.3390/rs11060617
- Han, J., Pei, J., and Tong, H. (2022). *Data mining: Concepts and techniques*. Cambridge, MA, United States: Morgan Kaufmann.
- He, M., Luo, C., Yang, H., Kong, F., Li, Y., Deng, L., et al. (2020). Sources and a proposal for comprehensive exploitation of lithium brine deposits in the Qaidam Basin on the northern Tibetan Plateau, China: Evidence from Li isotopes. *Ore Geol. Rev.* 117, 103277. doi:10.1016/j.oregeorev.2019.103277
- Hestir, E. L., Brando, V., Campbell, G., Dekker, A., and Malthus, T. (2015). The relationship between dissolved organic matter absorption and dissolved organic carbon in reservoirs along a temperate to tropical gradient. *Remote Sens. Environ.* 156, 395–402. doi:10.1016/j.rse.2014.09.022
- Katoch, S., Chauhan, S. S., and Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. *Multimed. Tools Appl.* 80, 8091–8126. doi:10.1007/s11042-020-10139-6
- Kong, F., Yang, Y., Luo, X., Sha, Z., Wang, J., Ma, Y., et al. (2021). Deep hydrothermal and shallow groundwater borne lithium and boron loadings to a mega brine lake in Qinghai Tibet Plateau based on multi-tracer models. *J. Hydrol.* 598, 126313. doi:10.1016/j.jhydrol.2021.126313
- Li, L., Chen, Y., Xu, T., Liu, R., Shi, K., and Huang, C. (2015). Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm. *Remote Sens. Environ.* 164, 142–154. doi:10.1016/j.rse.2015.04.009
- Li, S., Chen, F., Song, K., Liu, G., Tao, H., Xu, S., et al. (2022). Mapping the trophic state index of eastern lakes in China using an empirical model and Sentinel-2 imagery data. *J. Hydrol.* 608, 127613. doi:10.1016/j.jhydrol.2022.127613
- Liu, C., Zhu, L., Wang, J., Ju, J., Ma, Q., Qiao, B., et al. (2021a). *In-situ* water quality investigation of the lakes on the Tibetan Plateau. *Sci. Bull.* 66, 1727–1730. doi:10.1016/j.scib.2021.04.024
- Liu, T., Dai, J., Zhao, Y., Tian, S., and Ye, C. (2021b). Remote sensing inversion of lithium concentration in salt lake using LightGBM: a case study of northern Zabuye salt lake in Tibet. *Acta Geol. Sin.* 95, 2249–2256. doi:10.19762/j.cnki.dizhixuebao.2021222
- Liu, Z., Yao, Z., and Wang, R. (2019). Automatic identification of the lake area at Qinghai-Tibetan Plateau using remote sensing images. *Quat. Int.* 503, 136–145. doi:10.1016/j.quaint.2018.10.023
- Ma, R., Yang, G., Duan, H., Jiang, J., Wang, S., Feng, X., et al. (2011). China's lakes at present: Number, area and spatial distribution. *Sci. China Earth Sci.* 54, 283–289. doi:10.1007/s11430-010-4052-6
- Maier, P. M., Keller, S., and Hinz, S. (2021). Deep learning with WASI simulation data for estimating chlorophyll a concentration of inland water bodies. *Remote Sens.* 13, 718. doi:10.3390/rs13040718
- Marinho, R. R., Harmel, T., Martinez, J., and Filizola Junior, N. P. (2021). Spatiotemporal dynamics of suspended sediments in the negro river, amazon basin, from *in situ* and sentinel-2 remote sensing data. *ISPRS Int. J. Geo-Inf.* 10, 86. doi:10.3390/ijgi10020086
- Miles, K. E., Willis, I. C., Benedek, C. L., Williamson, A. G., and Tedesco, M. (2017). Toward monitoring surface and subsurface lakes on the Greenland ice sheet using sentinel-1 SAR and landsat-8 OLI imagery. *Front. Earth Sci.* 5, 58. doi:10.3389/feart.2017.00058
- Nemati, S., Fazelifard, M. H., Terzi, Ö., and Ghorbani, M. A. (2015). Estimation of dissolved oxygen using data-driven techniques in the Tai Po River, Hong Kong. *Environ. Earth Sci.* 74, 4065–4073. doi:10.1007/s12665-015-4450-3
- Qiao, B., Ju, J., Zhu, L., Chen, H., Kai, J., and Kou, Q. (2021). Improve the accuracy of water storage estimation—a case study from two lakes in the hohxil region of north Tibetan plateau. *Remote Sens.* 13, 293. doi:10.3390/rs13020293
- Saberioon, M., Brom, J., Nedbal, V., Souček, P., and Cisar, P. (2020). Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters. *Ecol. Indic.* 113, 106236. doi:10.1016/j.ecolind.2020.106236
- Shekofteh, H., and Masoudi, A. (2019). Determining the features influencing the-S soil quality index in a semiarid region of Iran using a hybrid GA-ANN algorithm. *Geoderma* 355, 113908. doi:10.1016/j.geoderma.2019.113908
- Song, K., Li, L., Tedesco, L. P., Li, S., Duan, H., Liu, D., et al. (2013). Remote estimation of chlorophyll-a in turbid inland waters: Three-band model versus GA-PLS model. *Remote Sens. Environ.* 136, 342–357. doi:10.1016/j.rse.2013.05.017
- Sun, W., Liu, S., Zhang, X., and Li, Y. (2022a). Estimation of soil organic matter content using selected spectral subset of hyperspectral data. *Geoderma* 409, 115653. doi:10.1016/j.geoderma.2021.115653
- Sun, X., Zhang, Y., Shi, K., Zhang, Y., Li, N., Wang, W., et al. (2022b). Monitoring water quality using proximal remote sensing technology. *Sci. Total Environ.* 803, 149805. doi:10.1016/j.scitotenv.2021.149805
- Tian, S., Qin, X., Zheng, M., Hong, Y., and Kuang, S. (2005). Quantitative analysis of remote sensing on the total salinity of zhabuye Salt Lake in Tibet. *Geoscience* 19, 596–602. doi:10.3969/j.issn.1000-8527.2005.04.016
- Tiyasha, T., Tung, T. M., Bhagat, S. K., Tan, M. L., Jawad, A. H., Mohtar, W. H. M. W., et al. (2021). Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Mar. Pollut. Bull.* 170, 112639. doi:10.1016/j.marpolbul.2021.112639
- Wagle, N., Pote, R., Shahi, R., Lamsal, S., Thapa, S., and Acharya, T. D. (2019). Estimating and mapping chlorophyll-A concentration of phewa lake of kaski district using Landsat imagery. *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.* IV-5/W2, 127–132. doi:10.5194/isprs-annals-IV-5-W2-127-2019
- Wang, D. (2019). *Models for predicting the Li content in salt lake based on remote sensing: A case study of Argentina's arzaro salt lake*. Jilin, China: Jilin University.
- Wang, J., Liu, J., Li, Z., Liu, D., and Wang, D. (2015). High resolution remote sensing estimation of salinity in Salt Lake with uranium resources. *Earth Sci. - J. China Univ. Geosciences* 40, 1409–1414. doi:10.3799/dqkx.2015.126
- Wang, J., Lu, D., Zhou, M., Wu, D., and Hao, W. (2021). WV-II high resolution data based quantitative inversion of salinity content for Salt Lake: A case study of gasikule Salt Lake. *Uranium Geol.* 37, 78–86. doi:10.3969/j.issn.1000-0658.2021.37.010
- Wang, M., Cheng, Y., Guo, B., and Jin, S. (2019). Parameters determination and sensor correction method based on virtual CMOS with distortion for the GaoFen6 WFV camera. *ISPRS-J. Photogramm. Remote Sens.* 156, 51–62. doi:10.1016/j.isprsjprs.2019.08.001
- Wang, S., Peng, H., and Liang, S. (2022a). Prediction of estuarine water quality using interpretable machine learning approach. *J. Hydrol.* 605, 127320. doi:10.1016/j.jhydrol.2021.127320
- Wang, X., Song, K., Liu, G., Wen, Z., Shang, Y., and Du, J. (2022b). Development of total suspended matter prediction in waters using fractional-order derivative spectra. *J. Environ. Manage.* 302, 113958. doi:10.1016/j.jenvman.2021.113958
- Wang, Z., Zhang, F., Zhang, X., Chan, N. W., Kung, H., Arike, M., et al. (2021). Regional suitability prediction of soil salinization based on remote-sensing derivatives and optimal spectral index. *Sci. Total Environ.* 775, 145807. doi:10.1016/j.scitotenv.2021.145807
- Wei, P., Lu, Z., and Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliab. Eng. Syst. Saf.* 142, 399–432. doi:10.1016/j.res.2015.05.018
- Wold, S., Trygg, J., Berglund, A., and Antti, H. (2001). Some recent developments in PLS modeling. *Chemom. Intell. Lab. Syst.* 58, 131–150. doi:10.1016/S0169-7439(01)00156-3
- Xie, S., Ding, F., Chen, S., Wang, X., Li, Y., and Ma, K. (2022). Prediction of soil organic matter content based on characteristic band selection method. *Spectrochimica Acta Part A Mol. Biomol. Spectrosc.* 273, 120949. doi:10.1016/j.saa.2022.120949
- Xu, W., Bu, L., Kong, W., Zheng, M., and Nie, Z. (2017). Monitoring of the dynamic change of Zabuye Salt Lake: A remote sensing approach. *Sci. Technol. Rev.* 35, 89–96. doi:10.3981/j.issn.1000-7857.2017.06.011
- Yan, L. J., and Zheng, M. P. (2015). Influence of climate change on saline lakes of the Tibet Plateau, 1973–2010. *Geomorphology* 246, 68–78. doi:10.1016/j.geomorph.2015.06.006
- Zaman Zad Ghavidel, S., and Montaseri, M. (2014). Application of different data-driven methods for the prediction of total dissolved solids in the Zarehroud basin. *Stoch. Environ. Res. Risk Assess.* 28, 2101–2118. doi:10.1007/s00477-014-0899-y
- Zhang, D., Tian, S., and Luan, X. (2007). Remote sensing research on the spatial distribution of boric anhydride in the zhabuye Salt Lake of Tibet. *Remote Sens. Land & Resour.* 32–35, 48. doi:10.3969/j.issn.1001-070X.2007.01.006
- Zhang, L., Li, J., Liu, R., Zhou, Y., Zhang, Y., Ji, L., et al. (2022). Recovery of lithium from salt lake brine with high Na/Li ratio using solvent extraction. *J. Mol. Liq.* 362, 119667. doi:10.1016/j.molliq.2022.119667
- Zhou, Y., Zhang, R., Ma, H., Zhang, J., and Zhang, X. (2016). Retrieving of salt lake mineral ions salinity from hyper-spectral data based on BP neural network. *Remote Sens. Land & Resour.* 28, 34–40. doi:10.6046/gtzyy.2016.02.06
- Zhu, C., Ding, J., Zhang, Z., and Wang, Z. (2022). Exploring the potential of UAV hyperspectral image for estimating soil salinity: Effects of optimal band combination algorithm and random forest. *Spectrochimica Acta Part A Mol. Biomol. Spectrosc.* 279, 121416. doi:10.1016/j.saa.2022.121416