Journal of Geochemical Exploration 197 (2019) 228–237

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

# A combined approach using spatially-weighted principal components analysis and wavelet transformation for geochemical anomaly mapping in the Dashui ore-concentration district, Central China

Xinhua Wang[a,b,c], Jiajun Liu[a,b,*], Emmanuel John M. Carranza[d], Jiaxin Wang[a,b], Gongwen Wang[a,b], Degao Zhai[a,b], Qingzheng Yuan[b], Wenquan Shi[e], Jiangsu Zhang[e]

[a] State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Beijing, China
[b] School of Earth Science and Resources, China University of Geosciences (Beijing), Beijing, China
[c] No. 7 Gold Geological Team, China Armed Police Force, Yantai, China
[d] Geological Geosciences, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Durban, South Africa
[e] No.3 Geology and Exploration Team, Gansu Bureau of Geology and Mineral Exploration and Development, Lanzhou, China

ARTICLE INFO

*Keywords:*
Geochemical anomaly
Wavelet
Spatially-weighted principal components analysis
Receiver operating characteristics
Epithermal gold

ABSTRACT

Sometimes geochemical anomalies linked with buried mineral deposits are too weak to be recognized by conventional methods. In this study, element concentration data from a stream sediment survey were subjected to a combined method of SWPCA (spatially-weighted principal components analysis) and WT (wavelet transformation) to derive a geochemical anomaly model for epithermal Au deposits in the Dashui ore-concentration district. The SWPCA was applied as a data integration method to extract information related to mineralization in the geochemical data. The WT was applied as a powerful tool for recognizing mineralization related anomalies in a complex geochemical field and for enhancing weak anomalies from background. The SWPCA–WT geochemical anomaly model indicated high favorability values for the known mineral occurrences and out-performed PCA–WT and SWPCA models. The SWPCA–WT geochemical model generated in this study provides a robust guide for further gold exploration in the Dashui ore-concentration district.

## 1. Introduction

Geochemical and geophysical anomalies may reflect the presence of mineral deposits (Archibald et al., 1999; Carranza, 2008). Recognizing geochemical anomalies related to mineral deposits is an important task of geochemical mineral exploration (Carranza, 2009; Zuo, 2014). Exploration of concealed mineral deposits in covered areas has drawn increasing attention in economic geology in recent years (Cheng, 2012). However, surface geochemical anomalies caused by concealed mineral deposits can be very weak such that the anomalies may not be easily recognized by traditional analytical mapping methods. Cheng (2012) argued that just considering magnitudes of surface geochemical anomalies is not robust enough when used in covered areas due to the complex origins and subtleness of such anomalies. Identification of complex and weak anomalies has brought crucial challenge to mineral exploration in covered areas. However, over the past decades, exploration geologists have realized that the multi-fractal nature of geochemical anomalies must be considered when dealing with geochemical exploration data (Carranza, 2010; Cheng and Zhao, 2011; Arias et al., 2012; Yousefi et al., 2012, 2013; Zuo et al., 2015; Zuo and Wang, 2016).

Stream sediment geochemical surveys are significant for mineral exploration to recognize geological processes, specifically mineralization. The recognition and elimination of background geochemical field to distinguish geochemical anomalies linked to mineralization is essential for mineral exploration (Cheng, 2007; Afzal et al., 2010). The traditional statistical analyses include the “mean plus two standard deviations” to define background threshold values (Levinson, 1974; Rose et al., 1979; Howarth, 1983). However, the multifractal/fractal theory introduced by Mandelbrot (1983) has been proposed in modern methods that have been found to be more robust for enhancing and demarcating geochemical anomalies from background (Agterberg and Cheng, 1999; Cheng, 1999; Afzal et al., 2010; Carranza, 2011a; Zuo, 2011a; Sadeghi et al., 2015; Wang and Zuo, 2016).

The task of recognizing geochemical anomalies from background is akin to recognizing signals of interest from image data. In this regard,

* Corresponding author at: State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Beijing, China.
*E-mail address:* liujiajun@cugb.edu.cn (J. Liu).

https://doi.org/10.1016/j.gexplo.2018.12.008
Received 27 January 2018; Received in revised form 27 October 2018; Accepted 9 December 2018
Available online 12 December 2018
0375-6742/ 

wavelet transformation (WT) is robust for frequency-scale processing of images and signals. Techniques for wavelet analysis enable us to separate a complex signal into a number of simple ones for individual analysis (Goswami and Chan, 2011). In the broad field of geochemistry, multi-resolution wavelet analysis has been shown to be robust for processing of geochemical data, e.g., to process data on oxygen isotope compositions of marine sediments (Bolton et al., 1995), to remove surface interference in geochemical data for hydrocarbon exploration (Pei, 1998; Zhang et al., 2003, 2006), to detect fracture from log data on water saturation (Tokhmechi et al., 2009), and to remove the effect of thickness of cap rock on intensities of anomalies (Zhang et al., 2012). Shahi et al. (2015) used Haar as mother wavelet to recognize mineralization related geochemical anomalies of the Dali porphyry system, central Iran. Lately, a wavelet-based multi-scale decomposition (WMD) method was proposed by Chen and Cheng (2016) for consideration of the multi-scale nature of geochemical patterns.

This paper is concerned with recognition of anomalies in element concentration data from stream sediment samples collected from the Dashui ore-concentration district (OCD), China. As recommended by Carranza (2011a), the stream sediment geochemical data were subjected to isometric log-ratio (ILR) transformation (Egozcue et al., 2003). Then, the ILR-transformed data were subjected to PCA and to SWPCA (spatially-weighted principal components analysis) to extract mineralization-related multivariate anomalies. However, because ILR-transformed data do not have full rank or one-to-one relationship with original data (Egozcue et al., 2003), results of either PCA or SWPCA do not have straightforward interpretations. Accordingly, the principal component scores and loadings derived by either PCA or SWPCA were back-transformed to centered log-ratio (CLR) space (Filzmoser et al., 2009) to facilitate interpretation of mineralization multivariate anomalies. Then, WT (2-D wavelet transformation) was performed on the image of multivariate anomalies to study the spatial patterns of multi-element distribution and to enhance weak anomalies. Finally, a combined SWPCA–WT geochemical anomaly model was proposed and compared with PCA–WT model and SWPCA using the Prediction-Area (P-A) test and the Receiver Operating Characteristics (ROC) test.

## 2. Study area

The Dashui OCD is located east of Maqu county between the Songpan–Ganzi Basin and the West Qinling Orogen (Chen et al., 2004) (Fig. 1). It contains > 46 t indicated gold reserve with ~9.9 g/t, and > 100 t possible gold resource (Mao et al., 2002). In the last 20 years or so, four gold deposits have been discovered in this OCD, namely: Geerke, Gongbei, Zhongqu and Xinqu (Fig. 1c). These deposits, classified as epithermal Au deposits (Zeng et al., 2013), are still being mined up to now. The average elevation of this OCD is ~3900 m.

The strata in the area comprise bioclastic limestone, micritic limestone, and marine platform facies of clastic, sandy or muddy limestone of the Triassic Maresongduo Group, overlain by a Jurassic sequence of interbedded siltstone, sandstone, coal, and conglomerate. These Triassic and Jurassic strata unconformably overlie Cretaceous molasses comprised of sandstone and carbonate gravel. The Triassic limestone strata are intruded by stocks of intermediate-acid rocks or as small laccoliths, comprising mainly granite, granodiorite, diorite porphyry, (quartz) diorite, monzonite porphyry from the Yanshan orogenic phase, some of which are associated with breccia in the host limestone (Dai et al., 2009). However, most of the area is covered with thick eluvium of weathered bedrock outcrops or Quaternary residuals.

The fracture structures developed in the area mainly have nearly E–W trends (the overall trend is 100–115°) with steep dips of 60–75° to the N or S. Several granodiorite dikes and calcite veins fill fractured zones, forming several bead-like open spaces that apparently were favorable areas for hydrothermal fluid transportation and gold deposition such that, with alternate permutation, ladder shaped structures were formed that control the ore-forming process. Typically, intersections of fractures were the most enriched and concentrated sites of ore deposition. According to Zeng et al. (2013), fractures trending N−S were important controls on ore formation and distribution.

A regional stream sediment geochemical survey has been carried out by the Third Geological Survey of Gansu in the Dashui OCD during 2015–2016. The average sampling density was five samples per km$^2$, which resulted in an irregular sampling grid with sample spacing of ~400 m. In total, 1479 stream sediments samples, each weighing ~150–400 g, were collected from the 264 km$^2$ survey area. Each sample has been placed in plastic receptacles and has been covered with fabric to avoid contamination. After thorough drying at room temperature, the samples have been sieved to 100 mesh fraction. Then the −100 mesh samples, each of ~50 g, have been stored in plastic containers, and the chemical analysis was performed using ICP-MS to measure concentration of 18 elements: Ag, As, Au, Bi, Cd, Co, Cr, Cu, Hg, Mo, Ni, Pb, Rb, Sb, Sn, Ti, W, Zn. The accuracy and precision of the analysis were within acceptable limits set by the Geological Survey of Gansu. The geochemical dataset was interpolated to a 50 m × 50 m grid using inverse distance weighting method and then stored as raster files in the GIS database.

## 3. Methodology

### 3.1. SWPCA (spatially-weighted principal components analysis)

Typically, principal components analysis (PCA) has been applied to investigate relationships among multiple geochemical variables for recognition and mapping of geological bodies (e.g., Reimann et al., 2002; Cheng et al., 2006; Cheng, 2007; Grunsky et al., 2009, 2014; Zuo, 2011b; Deutsch et al., 2016; Parsa et al., 2017). Prior to PCA, the geochemical data were subjected to isometric logratio (ILR) transformation using R package (R Development Core Team, 2017) to avoid the closure problem in compositional data analysis (Filzmoser et al., 2009, 2010; Carranza, 2011b, 2016). The PCA then results in several uncorrelated variables termed as principal components (PCs), each of which is a weighted combination of individual input variables, and loadings on individual input variables are interpreted in a geological context for recognition of anomalous geochemical processes.

Cheng et al. (2011) introduced a spatial weighting to ordinary PCA to derive SWPCA by computing for every pair of variables a spatially-weighted correlation coefficient defined as:

$$R(A,B) = \frac{\frac{1}{mn}\Sigma S_{ij}(A_{ij} - \overline{A})(B_{ij} - \overline{B})}{\sqrt{\frac{1}{mn}\Sigma S_{ij}(A_{ij} - \overline{A})^2}\sqrt{\frac{1}{mn}\Sigma S_{ij}(B_{ij} - \overline{B})^2}} \quad (1)$$

where $A_{ij}$ and $B_{ij}$ are values of geochemical elements $A$ and $B$ at location $(i, j)$, respectively; $\overline{A}$ and $\overline{B}$ are average values of $A$ and $B$, respectively; $m$ and $n$ are numbers of rows and columns, respectively, of $A$ and $B$ in the data matrix, and $S_{ij}$ is the spatial weighting factor.

A spatial weighting factor, which was used to highlight Au mineralization in the area, is defined as follows. Firstly, known orebodies are outlined and the Au grade level of each ore body is measured. Secondly, Euclidean distances between every location (i.e., pixel in image data) and the nearest orebody were measured to derive a function for weighting, which is then multiplied by the Au grade of the closest ore body. This means that the closer a location is to an orebody with high Au grade, the higher the weight is assigned to that location. Thirdly, locations outside the known mining areas, which are not known yet to possess Au mineralization, were weighted as 0. In practice, values of spatial weighting factors are assigned to locations according to their importance relative to the target of interest (i.e., Au mineralization), and the value ranges from 0 to 1 (Cheng et al., 2011): $S_{ij} = 0$ means that a location with $(i,j)$ coordinates is not associated with Au mineralization and its influence is ignored the correlation coefficient computation whereas $S_{ij} = 1$ means that a location with $(i,j)$ coordinates is strongly associated with Au mineralization and its influence is enhanced in the

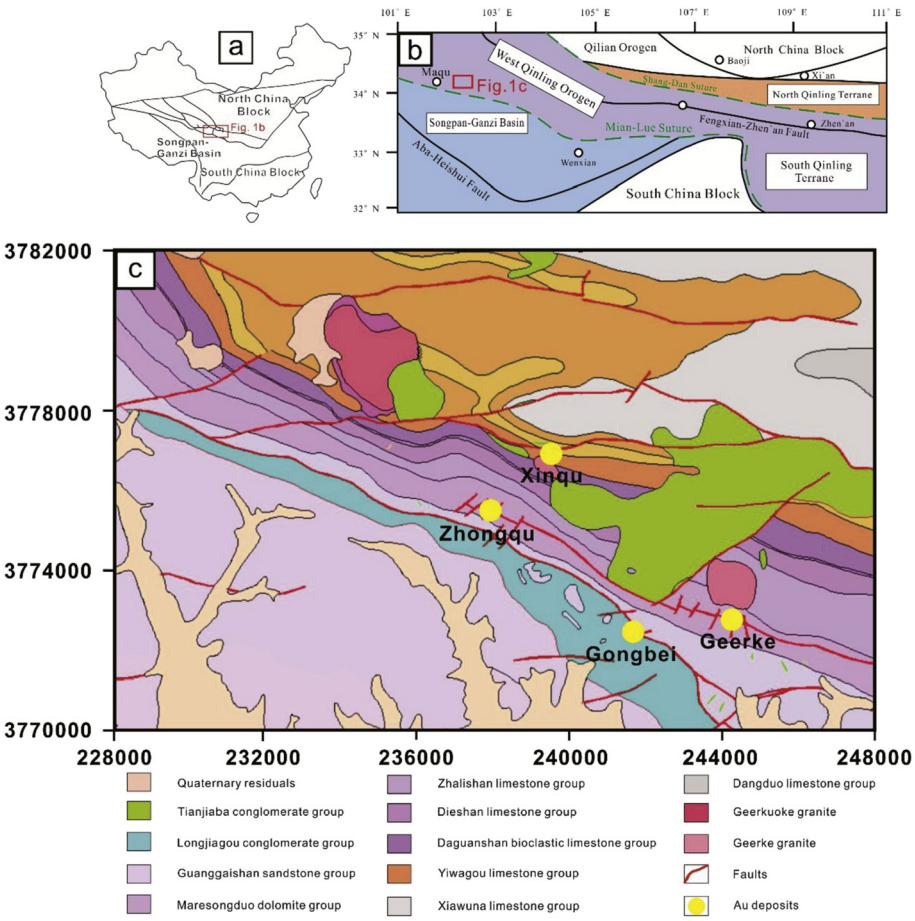


**Fig. 1.** Locations of (a) the West Qinling Orogen in China and (b) the study area in West Qinling Orogen (modified from Chen et al., 2004). (c) Simplified geological map of the Dashui ore concentration district, with locations of known Au deposits annotated.

computation of the correlation coefficient. Due to the spatial weighting, the correlation coefficients of $k$ elemental concentrations form a symmetrical $k \times k$ correlation matrix $R$ (Cheng et al., 2011). After applying a spatial weighting factor to PCA, new spatially-weighted principal components (SWPCs) are produced as:

$$SWPC_n = b_{1n}X_1 + b_{2n}X_2 + \ldots + b_{pn}X_p \tag{2}$$

where $SWPC_n$ refers to the $n$th spatially-weighted PC of the ILR-transformed data, $[b_{1n}, b_{2n}, \ldots, b_{pn}, n = 1, 2, \ldots, p]$ are the new eigenvectors of SWPCA, and $X_p$ refers to the values derived from the original PCA (Cheng et al., 2011).

However, the results of either PCA or SWPCA do not have straightforward interpretations because ILR-transformed data do not have full rank or one-to-one relationship with original data (Egozcue et al., 2003). Accordingly, the principal component scores and loadings derived by either PCA or SWPCA were back-transformed to centered log-ratio (CLR) space (Filzmoser et al., 2009) to facilitate interpretation of mineralization multivariate anomalies.

### 3.2. WT (wavelet transformation)

The WT is robust for noise removal, image compression, and analysis of signals (Grossmann and Morlet, 1984; Mallat, 1989). It has been shown that WT, compared to Fourier transformation (FT), is more flexible and performs better in the analysis of non-stationary signals (Cartas et al., 2009; Zhang et al., 2012). The WT is able to obtain information on scale, frequency, and spatial relation of a signal simultaneously (Partal, 2009). The WT computes the similarity between frequency content of signals and mother wavelets (Daubechies, 1992). For a signal $x(t)$, the continuous wavelet transformation (CWT) is (Partal, 2009):

$$w(a,s) = \frac{1}{s^n}\int_{-\infty}^{+\infty} x(t)\phi * \left(\frac{t-a}{s}\right)\mathrm{d}t \quad (a,s) \in R^n \tag{3}$$

where $\phi$ denotes the mother wavelet, $t$ is time (location) parameter, $a$ is translation parameter, * is conjugate complex function, and $s$ is the scale of a wavelet. The $a$ is the time/location step for iteration of the window function. The $w(a, s)$ is a $n$-dimensional signal/map of wavelet power transformed to another scale, resulting from the signal being compressed or expanded by $s$. The translation results in the WT being localized in a particular location or time. The scale parameter have a reverse connection, and the latter results in either compression or stretching. The value of $a$ controls the degree of movement of the window. However, computing the wavelet coefficient for all scales using CWT generate enormous amounts of data and consumes so much time (Nakhaei and Nasr, 2012). In DWT (discrete wavelet transformation), position and scale are chosen based on powers of two. The transformed signal is divided into *Details*, which is of low scale and high frequency, and *Approximations*, which is of low frequency and high scale. In this way, the DWT of a signal is done with less time and more precision (Nakhaei and Nasr, 2012). The DWT is accomplished by adjusting wavelet representation to (Grossmann and Morlet, 1984):

$$\phi_{i,j}(t) = s_0^{-i/2}\phi\left(\frac{t - ja_0 s_0^i}{s_0^i}\right) \tag{4}$$

where $s$ is a fixed dilation step > 1, $a_0$ is translation parameter that must be > 0, and $i$ and $j$ are integers that correspondingly determine

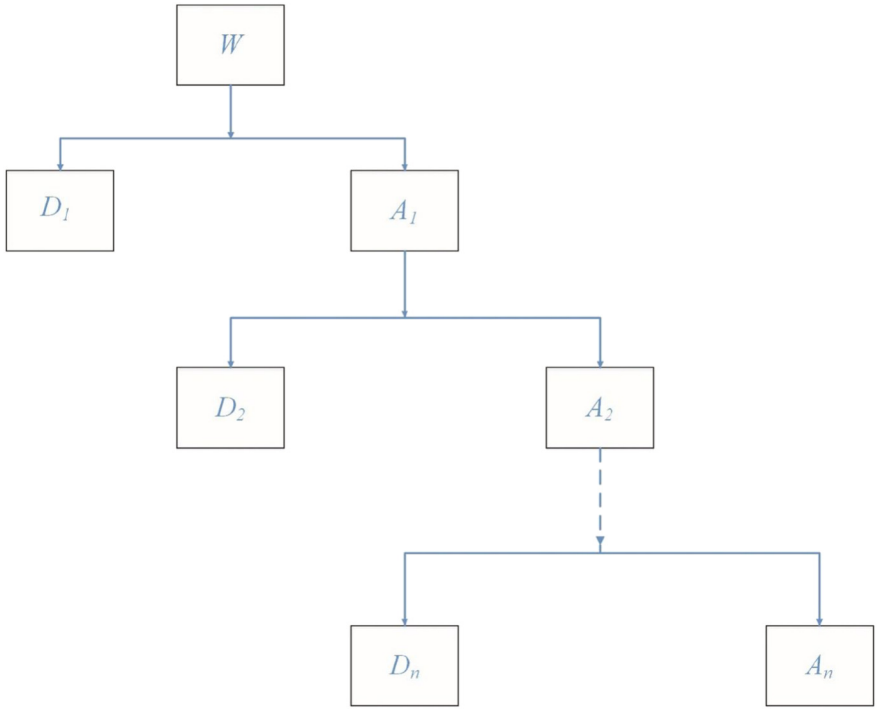


**Fig. 2.** Simplified architecture of wavelet transformation with *Details* (*D*) and *Approximations* (*A*).

scale and time. The coefficients of *Details* and *Approximations* are derived, respectively, through a wavelet algorithm that uses high- and low-pass filters (Zhang and Li, 2001). Iterative applications of these filters result in certain frequencies in the signal being removed, and thereby obtaining the *Details* and *Approximations* of a signal (Fig. 2) (Adamowski and Chan, 2011). Typically, the information related to mineralization is mainly contained in the *Details* part of a wavelet transformed signal (Chen and Cheng, 2016).

### 3.3. ROC (receiver operating characteristics) validation

The ROC technique has been increasingly used in data mining research and mineral prospectivity mapping to test the performance of predictive models (Nykänen et al., 2015, 2017). In the confusion matrix in Fig. 3, TN (True Negatives) and TP (True Positives) are numbers of correctly classified negative and positive instances, respectively; FN (False Negatives) is number of negative instances classified wrongly as positive, and FP (False Positives) is number of positive instances classified wrongly as negative. The ROC curve represents the optimum decision boundary for relative costs of FP and TP. The *x*-axis of a ROC curve represents True Positive rate [TP/(FN + TP)] and the *y*-axis False Positive rate [FP/(TN + FP)]. The AUC (area under a ROC curve), which may vary from 0 to 1, is a commonly used metric for evaluating a predictive model's performance; an AUC of 1 indicates perfectly accurate results whereas an AUC of 0.5 indicates a totally random model and the curve would follow the diagonal (Nykänen et al., 2015).

| | Predicted Negative | Predicted Positive |
|---|---|---|
| Actual Positive | TN | FP |
| Actual Negative | FN | TP |

**Fig. 3.** Diagram of a confusion matrix.

## 4. Results

The ILR-transformed geochemical data were subjected to PCA and SWPCA to extract multi-element anomalies linked to gold mineralization in the Dashui OCD. Roughly 95% of the total variance of the dataset was explained by the first five components extracted by either PCA

**Table 1**
Result of PCA: loadings of ILR-transformed element data on principal components back-transformed to CLR-space.

| Component loading | $PC_1$ | $PC_2$ | $PC_3$ | $PC_4$ | $PC_5$ |
|---|---|---|---|---|---|
| lc(Ag) | −0.174 | 0.285 | −0.086 | −0.037 | −0.219 |
| lc(As) | −0.209 | 0.349 | 0.183 | 0.105 | −0.065 |
| lc(Au) | −0.162 | 0.197 | −0.419 | 0.665 | −0.18 |
| lc(Bi) | 0.252 | 0.26 | 0.136 | 0.129 | 0.244 |
| lc(Cd) | −0.171 | −0.092 | 0.37 | −0.396 | −0.409 |
| lc(Co) | 0.263 | 0.161 | 0.152 | −0.151 | −0.357 |
| lc(Cr) | 0.314 | 0.143 | −0.064 | −0.069 | −0.205 |
| lc(Cu) | 0.329 | 0.057 | −0.127 | −0.081 | −0.081 |
| lc(Hg) | −0.079 | 0.111 | −0.201 | −0.532 | 0.433 |
| lc(Mo) | −0.122 | 0.22 | 0.508 | −0.098 | −0.088 |
| lc(Ni) | 0.323 | 0.123 | 0.05 | −0.144 | −0.206 |
| lc(Pb) | −0.026 | 0.337 | 0.405 | 0.092 | 0.337 |
| lc(Rb) | 0.278 | 0.207 | 0.03 | 0.156 | 0.311 |
| lc(Sb) | −0.241 | 0.294 | −0.198 | 0.073 | −0.129 |
| lc(Sn) | −0.221 | 0.289 | −0.187 | 0.096 | −0.139 |
| lc(Ti) | 0.321 | 0.151 | −0.093 | −0.042 | −0.107 |
| lc(W) | −0.091 | 0.428 | −0.13 | −0.073 | 0.088 |
| lc(Zn) | 0.318 | 0.146 | −0.124 | −0.052 | −0.11 |
| % of total variance | 46.1 | 21.4 | 14.7 | 8.5 | 3.9 |
| Cumulative % of total variance | 46.1 | 67.5 | 82.2 | 90.7 | 94.6 |

**Table 2**
Result of SWPCA: loadings of ILR-transformed element data on spatially-weighted principal components back-transformed to CLR-space.

| Component loading | $SWPC_1$ | $SWPC_2$ | $SWPC_3$ | $SWPC_4$ | $SWPC_5$ |
|---|---|---|---|---|---|
| lc(Ag) | −0.231 | −0.267 | −0.142 | −0.179 | 0.263 |
| lc(As) | −0.245 | −0.303 | 0.013 | −0.112 | −0.244 |
| lc(Au) | −0.206 | −0.311 | −0.031 | 0.411 | 0.066 |
| lc(Bi) | 0.236 | −0.291 | 0.085 | −0.031 | 0.188 |
| lc(Cd) | −0.004 | 0.303 | −0.473 | 0.285 | −0.199 |
| lc(Co) | 0.3 | −0.09 | −0.096 | 0.175 | −0.321 |
| lc(Cr) | 0.265 | −0.263 | −0.142 | 0.227 | −0.066 |
| lc(Cu) | 0.323 | −0.09 | −0.109 | −0.188 | −0.122 |
| lc(Hg) | −0.009 | 0.076 | −0.299 | 0.603 | 0.376 |
| lc(Mo) | −0.068 | −0.016 | −0.203 | 0.275 | −0.12 |
| lc(Ni) | 0.329 | −0.051 | −0.201 | 0.11 | −0.187 |
| lc(Pb) | 0.067 | −0.104 | 0.165 | −0.139 | 0.379 |
| lc(Rb) | 0.262 | −0.231 | 0.193 | 0.002 | 0.242 |
| lc(Sb) | −0.246 | −0.296 | −0.11 | 0.342 | −0.285 |
| lc(Sn) | −0.206 | 0.374 | −0.144 | −0.264 | −0.127 |
| lc(Ti) | 0.296 | −0.239 | −0.104 | −0.242 | −0.034 |
| lc(W) | −0.213 | 0.344 | −0.163 | 0.076 | 0.166 |
| lc(Zn) | 0.291 | −0.235 | −0.101 | −0.104 | −0.01 |
| % of total variance | 38.1 | 28.5 | 12.5 | 9.6 | 6.7 |
| Cumulative % of total variance | 38.1 | 66.6 | 79.1 | 88.7 | 95.4 |

(Table 1) or SWPCA (Table 2); the remaining components extracted by either PCA or SWPCA were considered to represent noise. Element loadings on each of the first three components extracted by either PCA or SWPCA mainly reflect intrusive rocks, alteration zones and surface residual materials. The fifth component extracted by either PCA or SWPCA, which is dominated by Pb and Hg, represents polluted areas due to mining activities. These components are not of interest for mineral exploration, and so they are not discussed further. However, the fourth component extracted by either PCA or SWPCA, which explains barely 10% of the total variance, depicts a Hg-Au-Sb multi-element association that is typical of epithermal Au mineralization in the area (Yan, 1998). Thus, both the $PC_4$ (Table 1) and the $SWPC_4$ (Table 2) in this study were considered indicative of Au mineralization in the Dashui OCD.

Since a great part of the Dashui OCD is covered with grasslands and/or weathered residual soil, and some areas are contaminated with mining pollutions, geochemical anomalies in these areas are considered to be weakened by deeply-weathered surface covers or interfered by high background. This adds to the difficulty of identifying geochemical anomalies related to Au mineralization. As preliminary analysis, WT was trialed using the PCA and SWPCA transformed geochemical data to enhance weak or complex geochemical anomalies. For instance, the outlined area in Fig. 4 is covered with grass and Quaternary residual soil based on field observation, and, in the geochemical map of $PC_4$ (Fig. 4a, c), the geochemical anomalies are dull and difficult to identify; whereas, in the $PC_4$–WT map three obvious anomalies can be recognized (Fig. 4b, d). Additionally, in areas that have experienced heavy metal contamination or pollution, such as the mining areas, geochemical anomalies are often interfered by high background, making it difficult to recognize the 'concentration center' of geochemical anomalies. For example, the outlined area in Fig. 5 contains two operating Au deposits (Geerke and Gongbei) and this area has suffered severe contamination due to mining activities. In the $SWPC_4$ map (Fig. 5a, c), the patterns of ore-related anomalies are obscured by high Au background; whereas, in the $SWPC_4$–WT map (Fig. 5b, d), two significant geochemical 'concentration centers' were revealed coincide with the locations of the two deposits. Therefore, WT is apparently a powerful tool for enhancing weak or complex anomalies in this study, with a major improvement to the recognition of mineral potential in covered or/and contaminated areas.

We subjected the images of $PC_4$ and $SWPC_4$ to WT to obtain $PC_4$–WT and $SWPC_4$–WT models, respectively. The locations of known ore deposits coincide with high favorability areas in both the $PC_4$–WT and $SWPC_4$–WT models, but the concentration centers in the $SWPC_4$–WT model have stronger spatial association with the approximate outlines of orebody of the four known Au deposits in the Dashui OCD (Fig. 6). The classification of anomalies based on the $SWPC_4$–WT model (Fig. 7) shows several spots as favorable targets in the Dashui OCD, some of which have already been planned for further detailed assessments.

The performance of the prospectivity models generated in this study were evaluated using prediction-area (P-A) plots to check the predictive ability of each model for epithermal Au deposits in the Dashui OCD (Fig. 8). According to Yousefi and Carranza (2015a, 2015b), if there are some known mineral occurrences (KMOs) in a study area, the P-A plot can be used as an effective tool to compare and evaluate the ability to predict exploration targets with respect to the size of the corresponding class. In the P-A plots, the prediction ability of each model and its ability to delimit exploration targets is evaluated in a scheme that shows the relation between the percentage of occupied areas of each class and the corresponding percentage of KMOs contained in each class versus the class thresholds. The intersection point of the curve of prediction rate and the curve of occupied area in a P-A plot is a proper criterion of evaluating the predictive models; the higher the Y value of an intersection point represents a smaller area containing large number of exploration targets, and thus it is easier to find potential mineral occurrences in these areas (Yousefi and Nykanen, 2016).

Inspection of the intersection points in the P-A plots of the four models generated in this study shows that 63.12% of the known epithermal Au occurrences were predicted in 36.88% of the total area based on the result of $PC_4$ model (Fig. 8a), 69.87% of the known epithermal Au occurrences were predicted in 30.13% of the total area based on the result of $SWPC_4$ model (Fig. 8b), 71.89% of the known epithermal Au occurrences were predicted in 28.11% of the total area based on the result of $PC_4$ –WT model (Fig. 8c), and 81.36% of the known epithermal Au occurrences were predicted in 18.64% of the total area based on the result of $SWPC_4$ model (Fig. 8d). This comparison shows the main advantage of the $SWPC_4$–WT model because in this study predicted the highest percentage of known mineralization occurrences in the least occupied area, which means that we have the highest probability of finding undiscovered epithermal Au deposits in relatively smallest area based on the $SWPC_4$–WT model in comparison with the other models generated in this study.

The performances of the PCA, SWPCA, PCA–WT and $SWPC_4$–WT model were further evaluated for comparison using the ROC test (Fig. 9). Validation of different models is critical in mineral potential mapping, and the main difference between conventional validation techniques and the ROC method is that the latter not only takes into consideration the known deposit sites (i.e., TP sites) but also the known not-deposit sites (i.e., TN sites) (Nykänen et al., 2017). In this study, the known Au deposits and other mineralization occurrences (newly discovered Au outcrops) were used to validate the performance of the geochemical models. According to Nykänen et al. (2015, 2017), if there is not much sites of other type of deposits, random sites could be used as valid TN sites, since there are no deposits of other type in this study area. Therefore, we used random sites as TN sites in this study. Because the total number of known deposits ($N = 4$) and mineralization occurrences ($N = 7$) is relatively low in the area ($N = 11$), we generated more instances (of known mineralization occurrences) by introducing synthetic points on the decision boundary between the minority instances and their *k*-nearest neighbors using SMOTE (synthetic minority over-sampling technique) (Chawla et al., 2002). We subsequently generated 275 TP sites for ROC validation using the SMOTE technique. The TP sites generated by SMOTE are mainly based on the Geerke and Zhongqu gold mines, which are two major operating gold mines owned by the Gansu Gesaer Mining Company. The performances of each geochemical model are presented as ROC curves in Fig. 8 and listed in Table 3.

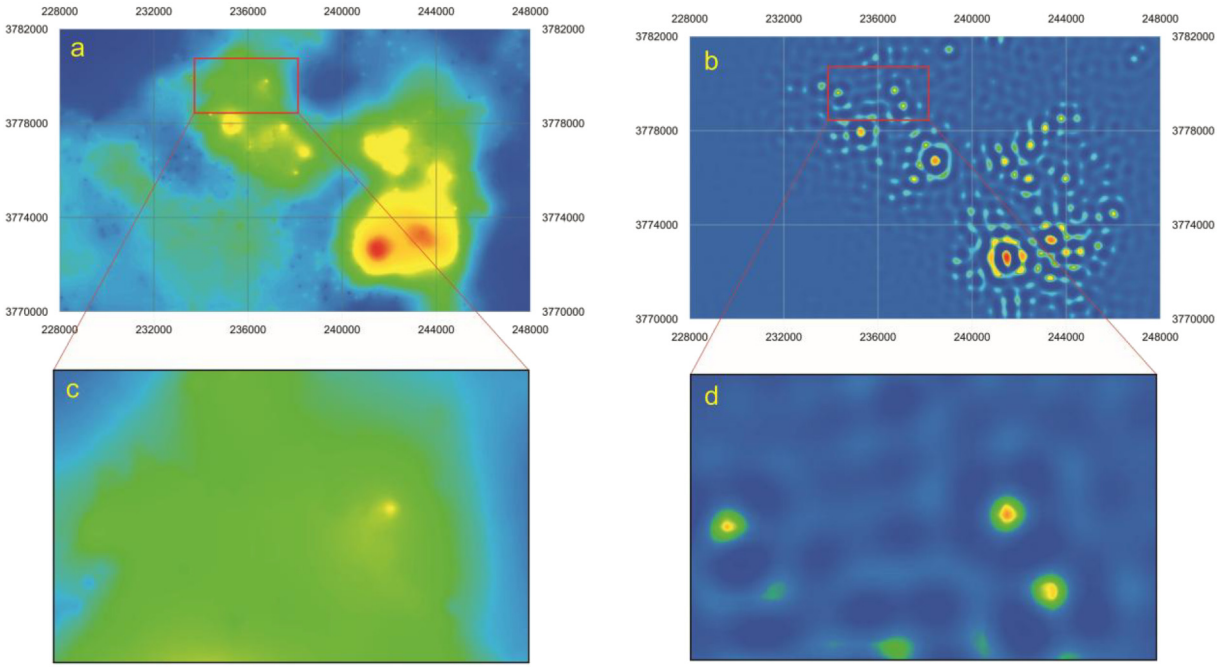


**Fig. 4.** (a) Geochemical map of $PC_4$ model and (b) corresponding wavelet transformed geochemical map of WT-$PC_4$. (c) Anomalies of $PC_4$ in a selected area (grass covered) and (d) corresponding wavelet transformed anomalies of WT-$PC_4$ in the selected area (grass covered).

## 5. Discussion

Anomalies in element concentrations in stream sediments have been conventionally exploited to guide mineral exploration of different types of mineral deposits at different scales, but prospecting for mineral deposits in covered and/or contaminated areas brings further challenges to stream sediment geochemical exploration. In covered areas, stream sediment geochemical anomalies can be quite weak due to the dilution of element concentrations by strongly weathered and transported surface materials whereas in polluted areas the background is elevated due to human activities, such that element concentrations derived from mineralization are not easily recognized by ordinary techniques (Cheng, 2012).

As shown in this study, the combined application of SWPCA and WT provides a new way for identifying weak and complex geochemical anomalies in covered and/or contaminated areas. The SWPCA serves as a technique for integrating multi-element geochemical data and known mineral occurrences to enhance multi-element signature of

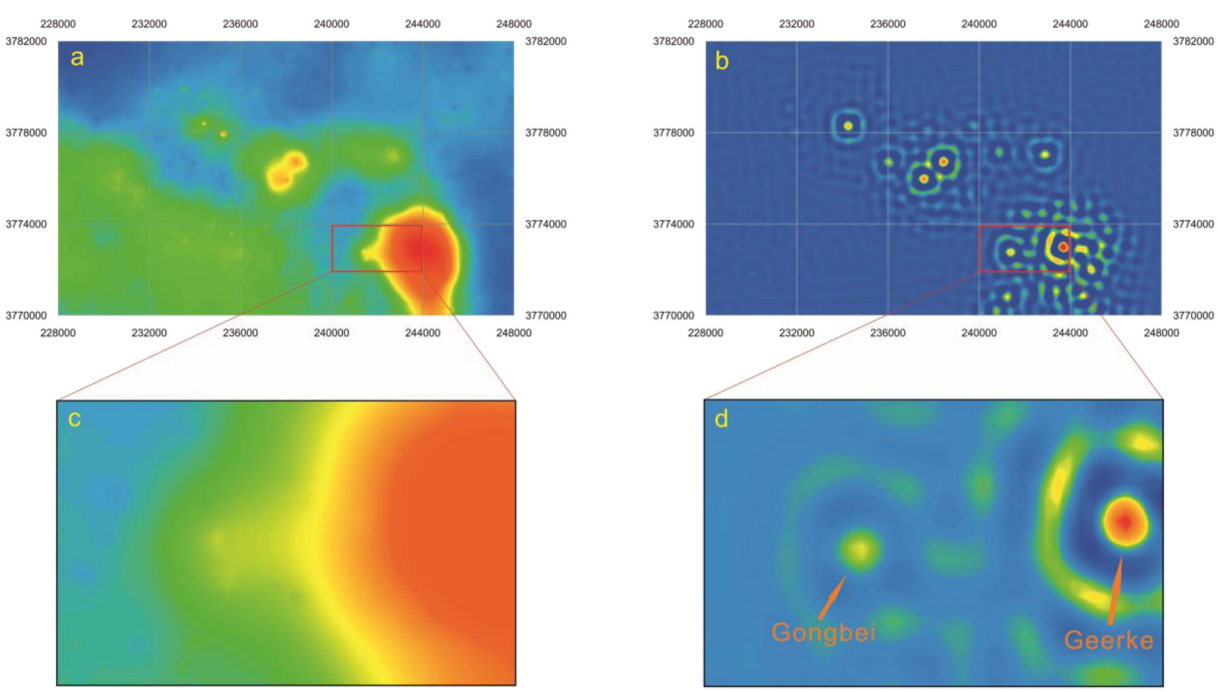


**Fig. 5.** (a) Geochemical map of the $SWPC_4$ model and (b) corresponding wavelet transformed geochemical map of the $SWPC_4$-WT model. (c) Geochemical anomalies of $SWPC_4$ in selected area (contaminated) and (d) corresponding wavelet transformed geochemical anomalies of $SWPC_4$-WT in selected area (contaminated) with location of deposits under operation annotated.

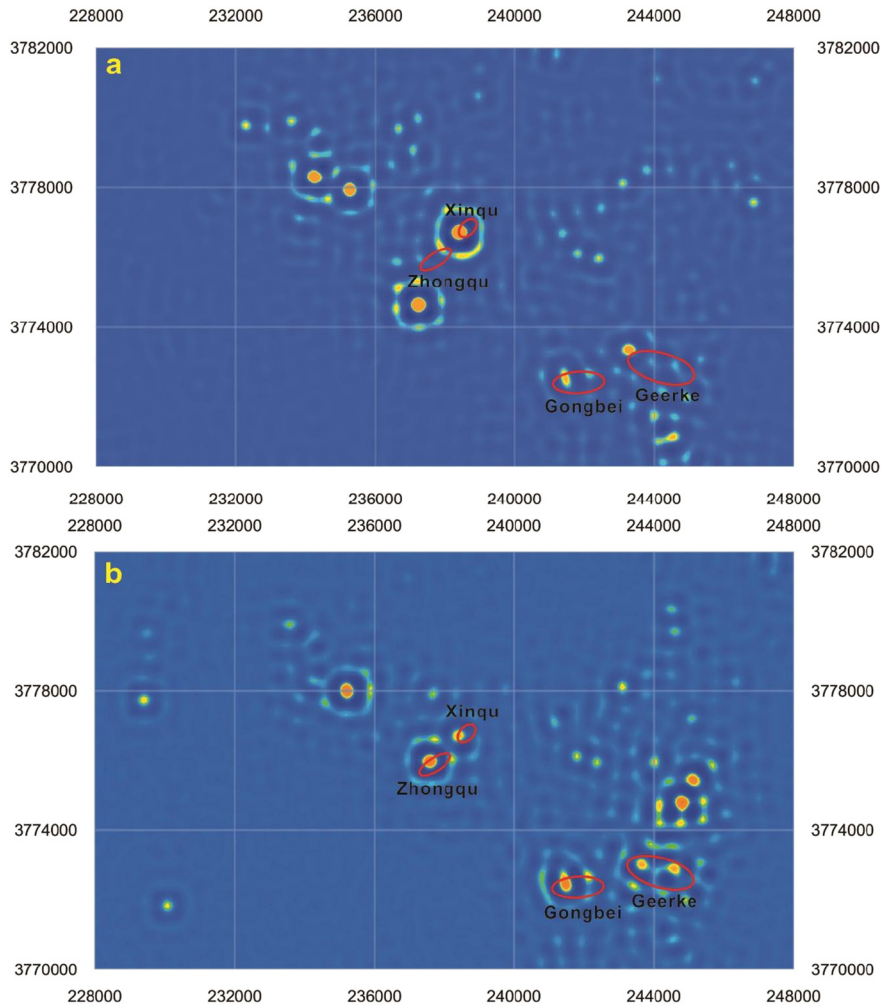


**Fig. 6.** Geochemical models of the Dashui OCD based on (a) $PC_4$–WT and (b) $SWPC_4$–WT. Ellipses in red are approximate outlines of ore in existing Au deposits in the area (see Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

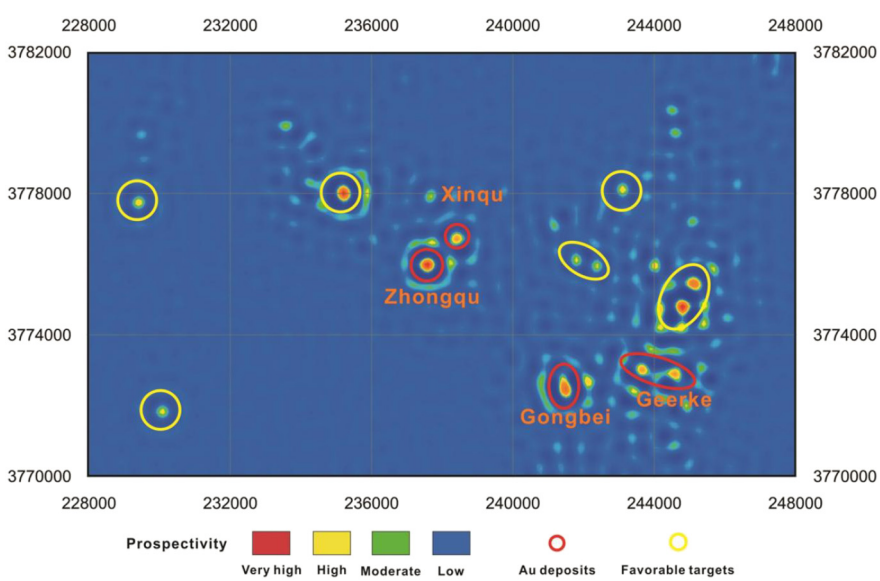


**Fig. 7.** Classified Au prospectivity map of the area based on the $SWPC_4$–WT geochemical model. Ellipses in red are approximate outlines of ore in existing Au deposits in the area (see Fig. 1). Polygons in yellow are new favorable targets from this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mineralization in covered and/or contaminated areas. Therefore, we selected $SWPC_4$ as a proxy of epithermal Au mineralization in the Dashui OCD. The WT has shown great potential in identifying weak and complex geochemical anomalies in covered and/or contaminated areas, the concentration centers of mineralization related geochemical distributions were significantly enhanced, making the wavelet transformed geochemical anomalies easier to recognize. In this study, based on the comparison of the P-A plots of the four predictive models, the

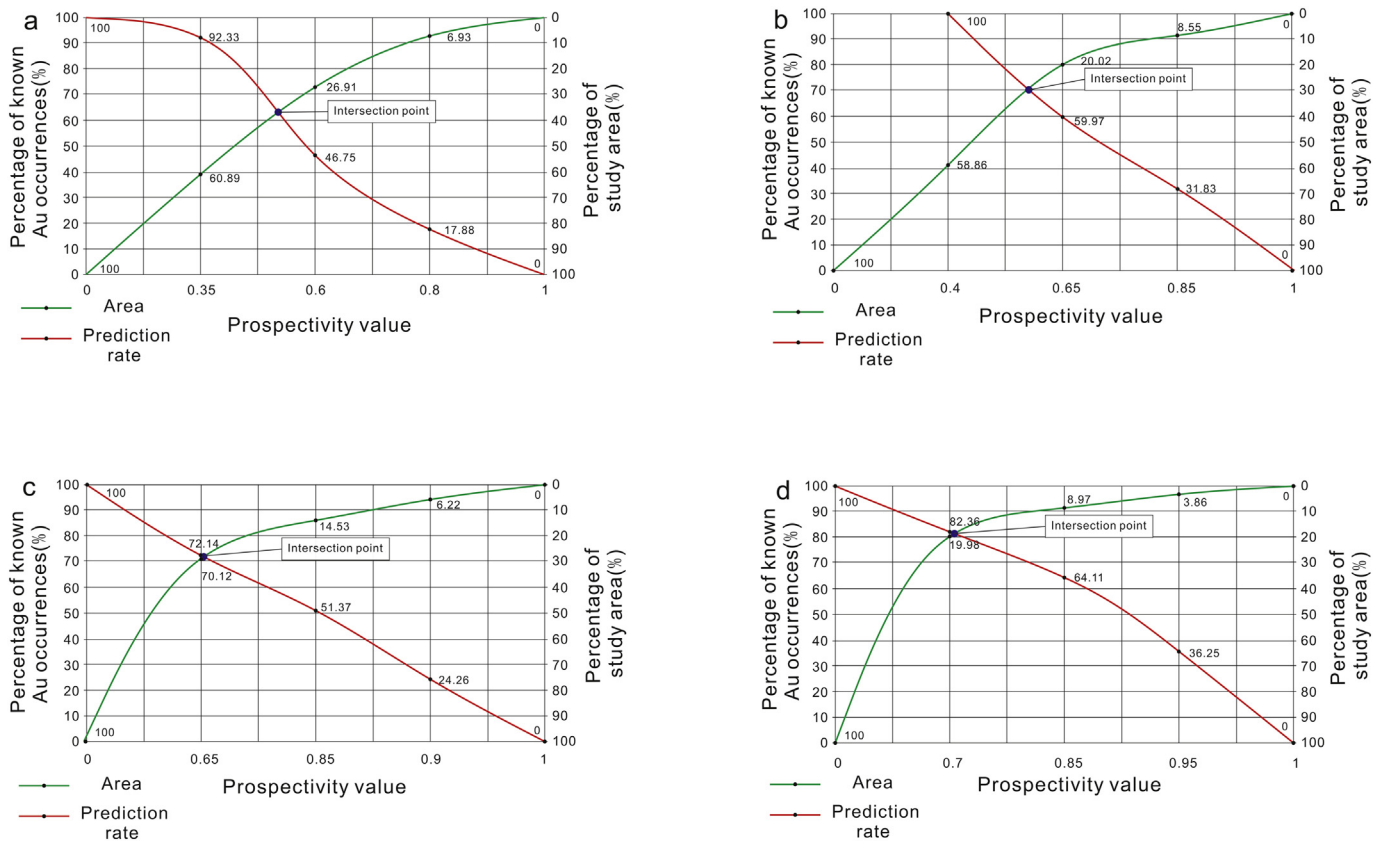


**Fig. 8.** (a) P-A plot for $PC_4$ model and (P-A) plot for $SWPC_4$ model and (c) P-A plot for $PC_4$-WT model and (d) P-A plot for $SWPC_4$-WT model (intersection point labeled blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

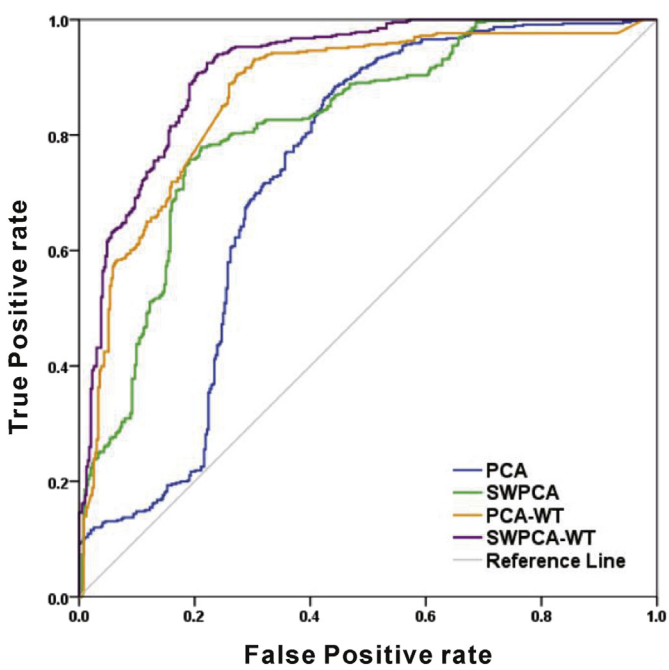


**Fig. 9.** ROC test results for geochemical anomaly models.

$SWPC_4$–WT model has shown the highest Y value in the intersection point of the curves, which signifies that the $SWPC_4$–WT model has the best ability of predicting the highest proportion of mineralization occurrences in the smallest prospective area, and the field observations also indicate that the locations of known mineralization occurrences match well with $SWPC_4$–WT anomalies, the high favorability areas in the $SWPC_4$ -WT model have better spatial consistence than those in the $PC_4$–WT model. Meanwhile, the result of ROC validation has shown that the $SWPC_4$–WT model, having the highest AUC value (Table 3), out-performed the predictive ability of the $SWPC_4$ and $PC_4$–WT models. We recommend, therefore, that the $SWPC_4$–WT geochemical model generated in this study can be used as a targeting model for further exploration of the Dashui OCD.

**Table 3**
Performance of geochemical anomaly models.

| Model | Area under curve | Std. error |
|---|---|---|
| $PC_4$ | 0.732 | 0.015 |
| $SWPC_4$ | 0.781 | 0.013 |
| $PC_4$–WT | 0.848 | 0.013 |
| $SWPC_4$–WT | 0.872 | 0.012 |

## 6. Conclusions

This paper demonstrated a geochemical anomaly model by means of spatially-weighted principal components analysis (SWPCA) and wavelet transformation (WT) to detect anomalies linked with Au mineralization and, thus, favorable exploration target areas in the Dashui ore-concentration district, Central China. The SWPCA works as a data integration method to extract spatial information of interest. In this regard, the $SWPC_4$ was selected as a multivariate signature of Au mineralization in the Dashui OCD. The application of WT to the $SWPC_4$ image as a filter technique has enhanced concentration centers of anomalies, demonstrating the effectiveness of WT for recognition of mineralization-related anomalies in covered and/or contaminated areas. The empirical proof of this is that the results of the prediction-area plots and ROC (Receiver Operator Characteristics) and AUC (area

under the curve) tests showed that the SW $PC_4$–WT model out-performed either the $PC_4$–WT or $SWPC_4$ model, resulting in anomalies with stronger spatial correspondence with the existing ore deposits and, therefore, robust targets for exploration of undiscovered deposits. The hybrid $SWPC_4$–WT model has effectively enhanced weak and/or complex geochemical anomalies for recognition of mineralization-related anomalies in the Dashui OCD. However, this method need further testing in other areas or on other types of ore deposits to further determine its usability and generalizability.

## Acknowledgements

This project received funding from National Natural Science Foundation of China, China (Grant #41730426 and 41573036), Major Basic Research Program of People's Republic of China, China (Grant #2014CB440903), China Geological Survey, China (Grant #12120115036401), by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Beijing), China (Grant No. 2652017264), and 111 Project under the Ministry of Education and the State Administration of Foreign Experts Affairs, China, China (Grant #B07011). We thank Senior Geologist Chunxian Liu and Zhirui Li at No. 3 Geology and Mineral Exploration Team, Gansu Provincial Bureau of Geology and Mineral Exploration and Development for their enthusiastic help during fieldwork.

## References

Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407, 28–40.
Afzal, P., Khakzad, A., Moarefvand, P., Omran, N.R., Esfandiari, B., Alghalandis, F.Y., 2010. Geochemical anomaly separation by multifractal modeling in Kahang (Gor Gor) porphyry system, Central Iran. J. Geochem. Explor. 104, 34–46.
Agterberg, F.P., Cheng, Q., 1999. Introduction to special issue on "Fractals Multifractals". Comput. Geosci. 25, 947–948.
Archibald, N., Gow, P., Boschetti, F., 1999. Multiscale edge analysis of potential field data. Explor. Geophys. 30, 38–44.
Arias, M., Gumiel, P., Martín-Izard, A., 2012. Multifractal analysis of geochemical anomalies: a tool for assessing prospectivity at the SE border of the Ossa Morena Zone, Variscan Massif (Spain). J. Geochem. Explor. 122, 101–112.
Bolton, E.W., Maasch, K.A., Lilly, J.M., 1995. A wavelet analysis of Plio-Pleistocene climate indicators: a new view of periodicity evolution. Geophys. Res. Lett. 22, 2753–2756.
Carranza, E.J.M., 2008. Geochemical Anomaly and Mineral Prospectivity Mapping in GIS. Handbook of Exploration and Environmental Geochemistry. vol. 11. Elsevier, Amsterdam, pp. 351.
Carranza, E.J.M., 2009. Controls on mineral deposit occurrence inferred from analysis of their spatial pattern and spatial association with geological features. Ore Geol. Rev. 35, 383–400.
Carranza, E.J.M., 2010. Mapping of anomalies in continuous and discrete fields of stream sediment geochemical landscapes. Geochem.: Explor., Environ., Anal. 10, 171–187.
Carranza, E.J.M., 2011a. Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. J. Geochem. Explor. 110, 167–185.
Carranza, E.J.M., 2011b. Geo-computation of mineral exploration targets. Comput. Geosci. 37, 1907–1916.
Carranza, E.J.M., 2016. Geochemical mineral exploration: should we use enrichment factors of log-ratios? Nat. Resour. Res. 26, 411–428.
Cartas, R., Moreno-Barón, L., Merkoçi, A., Alegret, S., Del Valle, M., Gutiérrez, J.M., Muñoz, R., 2009. Multivariate calibration model for a voltammetric electronic tongue based on a multiple output wavelet neural network. In: Biologically Inspired Signal Processing for Chemical Sensing. Springer, Berlin Heidelberg, pp. 137–167.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.
Chen, G., Cheng, Q., 2016. Singularity analysis based on wavelet transform of fractal measures for identifying geochemical anomaly in mineral exploration. Comput. Geosci. 87, 56–66.
Chen, Y., Zhang, J., Zhang, F., Pirajno, F., Li, C., 2004. Carlin and Carlin-like gold deposits in Western Qinling Mountains and their metallogenic time, tectonic setting and model. Geol. Rev. 50, 134–152.
Cheng, Q., 1999. Spatial and scaling modelling for geochemical anomaly separation. J. Geochem. Explor. 65, 175–194.
Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geol. Rev. 32, 314–324.
Cheng, Q., 2012. Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. J. Geochem. Explor. 122, 55–70.
Cheng, Q., Zhao, P., 2011. Singularity theories and methods for characterizing mineralization processes and mapping geo-anomalies for mineral deposit prediction. Geosci. Front. 2, 67–79.
Cheng, Q., Jing, L., Panahi, A., 2006. Principal component analysis with optimum order sample correlation coefficient for image enhancement. Int. J. Remote Sens. 27, 3387–3401.
Cheng, Q., Bonham-Carter, G., Wang, W., Zhang, S., Li, W., Xia, Q., 2011. A spatially weighted principal component analysis for multi-element geochemical data for mapping locations of felsic intrusions in the Gejiu mineral district of Yunnan, China. Comput. Geosci. 37, 662–669.
Dai, W., Chen, G., Ma, X., 2009. Features and origin of metallogenetic fluid of the Dashui gold deposit in Gansu Province. Gansu Geol. 18, 21–27 (in Chinese with English abstract).
Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM, Philadelphia, pp. 235.
Deutsch, J.L., Palmer, K., Deutsch, C.V., Szymanski, J., Etsell, T.H., 2016. Spatial modeling of geometallurgical properties: techniques and a case study. Nat. Resour. Res. 25, 161–181.
Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Math. Geol. 35, 279–300.
Filzmoser, P., Hron, K., Reimann, C., 2009. Principal components analysis for compositional data with outliers. Environmetrics 20, 621–632.
Filzmoser, P., Hron, K., Reimann, C., 2010. The bivariate statistical analysis of environmental (compositional) data. Sci. Total Environ. (19), 4230–4238.
Goswami, J.C., Chan, A.K., 2011. Fundamentals of Wavelets: Theory, Algorithms, and Applications. vol. 233 John Wiley & Sons.
Grossmann, A., Morlet, J., 1984. Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J. Math. Anal. 15, 723–736.
Grunsky, E., Drew, L.J., Sutphin, D.M., 2009. Process recognition in multi-element soil and stream-sediment geochemical data. Appl. Geochem. 24, 1602–1616.
Grunsky, E., Mueller, U.A., Corrigana, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: applications for predictive geological mapping. J. Geochem. Explor. 141, 15–41.
Howarth, R.J. (Ed.), 1983. Statistics and Data Analysis in Geochemical Prospecting, Handbook of Exploration Geochemistry. vol. 2 Elsevier, Amsterdam.
Levinson, A.A., 1974. Introduction to Exploration Geochemistry. Applied Publishing Ltd, Calgary.
Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693.
Mandelbrot, B.B., 1983. The Fractal Geometry of Nature. WH Freeman, San Francisco, pp. 468.
Mao, J., Qiu, Y., Goldfarb, R., Zhang, Z., Garwin, S., Ren, S., 2002. Geology, distribution, and classification of gold deposits in the western Qinling belt, central China. Mineral. Deposita 37, 352–377.
Nakhaei, M., Nasr, A.S., 2012. A combined wavelet-artificial neural network model and its application to the prediction of groundwater level fluctuations. Geopersia 2, 77–91.
Nykänen, V., Lahti, I., Niiranen, T., Korhonen, K., 2015. Receiver Operating Characteristics (ROC) as validation tool for prospectivity models – a magmatic Ni-Cu case study from the Central Lapland greenstone belt, Northern Finland. Ore Geol. Rev. 71, 853–860.
Nykänen, V., Niiranen, T., Molnar, F., Lahti, I., Korhonen, K., Cook, N., Skytta, P., 2017. Optimizing a knowledge-driven prospectivity model for gold deposits within Perapohja Belt, northern Finland. Nat. Resour. Res. 26, 571–584.
Parsa, M., Maghsoudi, A., Carranza, E.J.M., Yousefi, M., 2017. Enhancement and mapping of weak multivariate stream sediment geochemical anomalies in Ahar area, NW Iran. Nat. Resour. Res. 26, 443–455.
Partal, T., 2009. River flow forecasting using different artificial neural network algorithms and wavelet transform. Can. J. Civ. Eng. 36, 26–38.
Pei, T., 1998. Spatial Structure Analysis of Geochemical Field of Oil and Gas and Its Application to Exploration in North Tarim Basin (PhD thesis No. Bo 9817). China University of Geo-sciences, Wuhan, PR China, pp. 1–65 (in Chinese).
R Development Core Team, 2017. R: A Language and Environment for Statistical Computing. Vienna. http://www.r-project.org.
Reimann, C., Filzmoser, P., Garrett, R.G., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. Appl. Geochem. 17, 185–206.
Rose, A.W., Hawkes, H.E., Webb, J.S., 1979. Geochemistry in Mineral Exploration, 2nd edn. Academic Press, London.
Sadeghi, B., Madani, N., Carranza, E.J.M., 2015. Combination of geostatistical simulation and fractal modeling for mineral resource classification. J. Geochem. Explor. 149, 59–73.
Shahi, H., Ghavami, R., Rouhani, A.K., Kahoo, A.R., Haroni, H.A., 2015. Application of Fourier and wavelet approaches for identification of geochemical anomalies. J. Afr. Earth Sci. 106, 118–128.
Tokhmechi, B., Memarian, H., Rasouli, V., Noubari, H.A., Moshiri, B., 2009. Fracture detection from water saturation log data using a Fourier–wavelet approach. J. Pet. Sci. Eng. 69, 129–138.
Wang, J., Zuo, R., 2016. An extended local gap statistic for identifying geochemical anomalies. J. Geochem. Explor. 164, 86–93.
Yan, S., 1998. Studies on the Genesis of the Dashui Super-large Hematite Rich Silicalite-type Gold Ore Deposit in Gansu Province (PhD dissertation). pp. 1–95 (in Chinese).
Yousefi, M., Carranza, E.J.M., 2015a. Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. Comput. Geosci. 74, 97–109.
Yousefi, M., Carranza, E.J.M., 2015b. Prediction-area (P-A) plot and C-A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. Comput. Geosci. 79, 69–81.

Yousefi, M., Nykanen, V., 2016. Data-driven logistic-based weighting of geochemical and geological evidence layers in mineral prospectivity mapping. J. Geochem. Explor. 164, 94–106.
Yousefi, M., Kamkar-Rouhani, A., Carranza, E.J.M., 2012. Geochemical mineralization probability index (GMPI): a new approach to generate enhanced stream sediment geochemical evidential map for increasing probability of success in mineral potential mapping. J. Geochem. Explor. 115, 24–35.
Yousefi, M., Carranza, E.J.M., Abolghasem Kamkar-Rouhani, A., 2013. Weighted drainage catchment basin mapping of geochemical anomalies using stream sediment data for mineral potential modeling. J. Geochem. Explor. 128, 88–96.
Zeng, Q., Evans, J., McInnes, B., Batt, G., McCuaig, T., Bagas, L., Tohver, E., 2013. Geological and thermochronological studies of the Dashui gold deposit, West Qinling Orogen, Central China. Mineral. Deposita 48, 397–412.
Zhang, X., Li, D., 2001. A trous decomposition applied to image edge detection. Cartogr. Geogr. Inf. Sci. 7, 119–123.
Zhang, L., Bai, G., Xu, Y., 2003. A wavelet-analysis-based new approach for interference elimination in geochemical hydrocarbon exploration. Math. Geol. 35, 939–952.
Zhang, L., Bai, G., Zhao, K., Sun, C., 2006. Restudy of acid-extractable hydrocarbon data from surface geochemical survey in the Yimeng uplift of the ordos Basin, China: improvement of geochemical prospecting for hydrocarbons. Mar. Pet. Geol. 23, 529–542.
Zhang, L., Bai, G., Zhao, Y., 2012. A method for eliminating caprock thickness influence on anomaly intensities in geochemical surface survey for hydrocarbons. Math. Geosci. 44, 929–944.
Zuo, R., 2011a. Decomposing of mixed pattern of arsenic using fractal model in Gangdese belt, Tibet, China. Appl. Geochem. 26, 271–273.
Zuo, R., 2011b. Identifying geochemical anomalies associated with Cu and Pb–Zn skarn mineralization using principal component analysis and spectrum–area fractal the Gangdese Belt, Tibet (China). J. Geochem. Explor. 111, 13–22.
Zuo, R., 2014. Identification of geochemical anomalies associated with mineralization in the Fanshan district, Fujian, China. J. Geochem. Explor. 139, 170–176.
Zuo, R., Wang, J., 2016. Fractal/multifractal modeling of geochemical data: a review. J. Geochem. Explor. 164, 33–41.
Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multifractal perspective. J. Geochem. Explor. 148, 12–24.